Toward a Categorical Conception of Cognitive Normativity

Andrew Paul Moser
Sewanee, Tennessee

B.A., University of the South, 1993
M.A., University of Virginia, 1998

A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Philosophy

University of Virginia
January 2007

© Copyright by
Andrew Paul Moser
All Rights Reserved
January 2007

Abstract: Recent discussion of the nature of cognitive normativity has tended to be somewhat one-sided, favoring a hypothetical conception according to which the authority of cognitive norms is contingent upon our desires. Largely absent in this discussion is any consideration of the possibility that some cognitive norms might be categorical by virtue of being constitutive of thought. This strategy for explaining the categorical nature of certain cognitive norms, which I will call the constitutive strategy, holds that there are at least some cognitive norms conformity to which is constitutive of one's status as a thinker; such norms would therefore have authority over an individual qua thinker, not qua thinker-desirous-of-x. Although this strategy holds out the prospect of an account of how certain cognitive norms could be categorical, it seems at first glance to face two serious problems. First, if we construe these constitutive norms as basic principles of logic or rationality, as seems most plausible, the constitutive strategy seems to involve an overintellectualized picture of thought that distorts the reality of human cognition: if conformity to basic principles of logic or rationality is constitutive of thought, then much of what passes for thought among human beings doesn't merit the description, since it is clear that human thought often fails to conform to such principles. Second, it is unclear whether the same principles could play both normative and constitutive roles in thought: if a principle is constitutive of thought, then it would seem that one could not fail to think in accordance with it; but if that is the case, it is hard to see how such a principle could be normative. I argue that the constitutive strategy can be adapted to avoid these problems, and that, having adapted it in this way, we have a promising explanation of how certain cognitive norms could be categorical.

Table of Contents

Chapter 1: Values, Reasons, and Naturalism..1

Chapter 2: The Hypothetical Conception...11

2.1 Stich and the failure of analytic epistemology..14
2.2 Papineau and the mysteriousness of the categorical......................................23
2.3 The categorical conception: mysterious or mistaken?...................................31
2.4 Conclusion...36

Chapter 3: The Constitutive, Normative, and Dual Role Theses.........................38

3.1 Frege..40
3.1.1 Psychologism and the autonomy of logic..40
3.1.2 The normative thesis: logical laws as categorical.................................47
3.1.3 The constitutive thesis...50
3.2 Davidson...61
3.2.1 Psychophysical laws and reductionism...62
3.2.2 Conceptual relativism and the methodology of interpretation.............68
3.3 Conclusion...71

Chapter 4: The Coherence of the Dual Role Thesis..75

4.1 A strong version of the constitutive thesis...76
4.1.1 Is human thought really logical?...77
4.1.2 Are the constitutive and normative theses consistent?.........................78
4.2 A moderate version of the constitutive thesis...81
4.2.1 Normative concepts...82
4.3 A normative concept of belief...98
4.4 Does the dual role thesis yield categorical normativity?...............................100
4.5 Conclusion...106

Chapter 5: In Defense of the Dual Role Thesis..107

5.1 Gary's objection...107
5.1.1 The presuppositions behind Gary's objection.....................................108
5.1.2 A hollow victory?...110
5.1.3 Conceptual claims and Stichean objections: a Stichean Gary..............112
5.2 Disambiguating the Stichean Gary..116
5.2.1 *Quid juris?*..117
5.2.2 An argument from relativity..119
5.2.3 Internalism and the motivational impotence of conceptual claims........121

5.3 Responding to the Stichean objections..123
5.3.1 Objections from substantive indispensability......................................124
5.3.1.1 A "wide range of possible alternative notions"?......................127
5.3.1.2 Aberrant belief, perceptual belief, and action..........................137
5.3.2 Objection from motivational impotence...145

Chapter 6: Conclusion...154

Bibliography..165

Chapter 1: Values, Reasons, and Naturalism

> The Enlightenment starts out from a deep commitment to achieving universal peace and justice. The source of misery and injustice, it holds, is ignorance and confusion. . . . It is a historical fact, however, that this Enlightenment project of achieving human well-being through greater knowledge tended to undermine itself. The initial motivation for scientific objectivity was the dream of bringing about universal peace, justice, and happiness by extinguishing the narrow, parochial attachments that sustain illusion and prejudice. Yet, as this quest for knowledge unfolded, the discoveries made by modern science tended toward a mechanistic and materialistic picture of reality as a vast aggregate of brute, meaningless material objects in causal interactions. Given this mechanized and objectified worldview, however, it becomes increasingly difficult to see why one should be committed to the values that initially motivated the Enlightenment. For if nothing exists except inherently valueless material objects in push-pull causal interactions, it becomes plausible to suppose that values are not part of the furniture of the world, but are instead merely subjective projections of human wishes and longings onto things. [Guignon, xxxvi]

> It is notoriously difficult to see how ethical features such as intrinsic goodness and moral obligation can be part of the sort of reality that science describes. Like such ethical evaluations, some epistemological categories seem to resist assimilation to a scientific worldview. Several of the main epistemological characterizations, such as 'justified', 'rational' and 'reasonable', seem to make comparably evaluative appraisals of the mental states or processes to which they are applied. It is comparably unclear how such epistemic attributes of belief and reasoning fit into the world of natural science. Naturalistic epistemologists characteristically attempt to clarify this fit or eliminate the recalcitrant epistemic features. [Conee, §3]

> Where do we get these [evaluative or normative] ideas that outstrip the world we experience and seem to call it into question, to render judgment on it, to say that it does not measure up, that it is not what it ought to be? Clearly we do not get them from experience, at least not by any simple route. [Korsgaard 1996, 1]

Naturalism could be characterized as the view that supernatural entities or explanations have no legitimate place in our understanding of things. Of course, this characterization doesn't get us very far, since it leaves unexplained the notion of the

'natural' at work here. So let me propose the following statement as a way of remedying this problem: "Science is the measure of all things, of what is that it is and of what is not that it is not" (Sellars, 83). It may not be readily apparent *how* this statement remedies the problem, so let me explain. Naturalism, as I have described it, asks us to reject the supernatural in favor of the natural, which raises the question of what we are to count as natural. I am suggesting that we look at the Sellars passage as supplying the answer: the deliverances of science tell us what counts as natural. As our most reliable guide to what the world is like, science is what we should rely on in trying to understand things; the characterization of naturalism with which this paragraph began is simply the obverse of this claim: we should rely on nothing supernatural in our understanding of things, nothing other than science. Although there's plenty of room for improvement of this characterization[1], and no doubt there are some views we might be inclined to identify as forms of naturalism that it fails to capture, I believe that it does capture a wide range of the views typically described as naturalistic.

As the epigraphs preceding this chapter indicate, it's often said that there is no place for value in the world described by science, or that our normative and evaluative categories cannot be accommodated within a naturalistic worldview. From a metaphysical point of view, the thought seems to be that values do not comport well with the entities with which science populates the world: as Mackie puts it in *Ethics: Inventing Right and Wrong*, "if there were objective values, then they would be entities or

[1] In particular, it ignores the usual distinction between metaphysical or ontological and methodological or epistemological versions of naturalism.

qualities or relations of a very strange sort, utterly different from anything else in the universe" (Mackie, 38); they would have "to-be-pursuedness somehow built into" them, and it is hard to imagine how any of the brute features of the world described by science could exhibit such a characteristic (ibid., 40). Values are seen as epistemologically puzzling as well: as Korsgaard puts it, they "outstrip the world we experience", which suggests that "we do not get them from experience, at least not by any simple route." And if we hold, with most naturalists, that "virtually nothing is knowable a priori" (Kitcher 1992, 76), it can begin to look somewhat mysterious how we might have any knowledge of values.

Whatever we make of these claims, it is clear that many self-confessed naturalists themselves regard certain notions of value as inconsistent with their views. What is typically supposed to be problematic here, from a naturalistic perspective, is a very specific notion of value: a notion of value as 'intrinsic' or 'objective'. Mackie offers the following explanation: "The objective values which I am denying would be action-directing absolutely, not contingently. . . upon the agent's desires and inclinations" (Mackie, 29). In the background of this passage is the assumption that it is characteristic of values that they are 'action-directing' — part of what it is for something to be a value is for it to be capable of motivating us to action — and Mackie points to two ways in which we might imagine them to be so: absolutely, in which case their motivating power is not dependent upon the "agent's desires and inclinations", or contingently, in which case their motivating power is dependent in this way. We are mistaken, he claims, if we believe that any values are absolutely action-directing; the only genuine values are those whose motivating power is dependent upon our desires

and inclinations. To put the point another way, all values are subjective or extrinsic, in the sense that they have motivating power, and therefore status as values, only in relation to our desires or inclinations.

There's nothing essential about the language of values here. Talk of values is shorthand for talk about things being valuable and, as an initial rough approximation, we could say that for something to be valuable is, among other things, for us to have reason to preserve, respect, or cultivate it. Instead of speaking of values, we could therefore speak instead of reasons: a consideration counts as a reason for me to do something only if it stands in an appropriate relation to my desires or inclinations. There are no objective or intrinsic reasons. Or, to borrow Kant's terminology, all reasons are hypothetical, contingent upon one's desires and inclinations. The notion of a categorical reason — a consideration that constitutes a reason for one to do something independently of one's desires or inclinations — is an illusion.

So there is a particular conception of normativity (to choose a term broad enough to encompass both values and reasons) that is said to resist accommodation within a naturalistic worldview. A conception of normativity as a product of our desires and inclinations — of values as "subjective projections of human wishes and longings onto things", as Guignon puts it — seems perfectly congenial to such an outlook, but a conception of normativity according to which there are sources of normative authority that have some hold over us independently of such non-cognitive states is presumed to be inconsistent with it. I'll discuss some of the reasons offered for this presumption in Chapter Two. But at this point it's important to recognize that although these points about normativity are often made in the context of discussions of

ethics, they apply across the board, irrespective of the sort of values or reasons we consider. If there's reason to think that normativity is, by its very nature, hypothetical, then there's reason to think not just that all ethical normativity is hypothetical, but that cognitive or epistemic normativity is hypothetical as well.

In what follows I'll refer to the view that all cognitive norms are hypothetical as the *hypothetical conception*, and the denial of that view — the claim, in other words, that some cognitive norms are categorical — as the *categorical conception*. Given the widespread presumption that naturalism is inconsistent with the categorical conception, and what seems to be a prevailingly naturalistic mood in contemporary philosophy, at least in this country, it is perhaps no surprise that most accounts that treat explicitly of the nature of cognitive normativity advocate a hypothetical conception; defenses of a categorical conception are scarce. It is my aim in this dissertation to begin addressing this scarcity by revisiting the prospects for the categorical conception. There is a strand of recent philosophical thought, exhibited most clearly in Frege and Davidson, that I believe is congenial to the categorical conception. Moreover, the bearing of their views on the debate between these two conceptions of normativity has gone largely unnoticed. Illuminating this strand therefore affords us an opportunity to reexamine this debate from a fresh perspective, and perhaps to articulate a view of cognitive normativity as categorical that withstands naturalist objections.

Let me elaborate a bit on these views that I am claiming we can find in Frege and Davidson. Both philosophers are committed to a pair of views that, in combination, seem puzzling. They share a commitment to what I will call the *normative*

thesis: the view that there are norms governing thought — principles dictating, for example, how we ought to go about forming beliefs, making judgments, or reasoning. For Frege, these are principles of logic, while Davidson identifies them as more general principles of rationality. But these principles are not just normative for thought, but also constitutive of it: in some sense, thinking in accordance with these principles is a precondition for thinking at all. So they also accept what I will call the *constitutive thesis*.

There's nothing individually puzzling about these two theses, and, indeed, their combination would be unproblematic if each thesis were taken as applying to different principles. It is the application of the two theses to the *same* set of principles — the claim that certain principles play both normative *and* constitutive roles in thought, or what I will call the *dual role thesis* — that seems problematic. For if the constitutive thesis is true, and we cannot fail to think in accordance with these principles without at the same time simply failing to think, it's far from clear how these principles might play a normative role in thought. The idea of a normative demand on our cognition seems to be premised on the idea that we might fail to meet the demand while still engaging in thought; it is because we might fail to meet the demand in our thought that a normative principle is necessary to guide us. If by failing to conform to a principle we simply fail to engage in the conduct that would be governed by that very principle, were it normative, it's unclear how we can reasonably regard it as a normative principle governing cognition.

It may not be immediately evident how the dual role thesis is relevant to the debate over the nature of cognitive normativity, particularly given the aforementioned difficulty. But if this difficulty can be overcome — if it can be shown that a principle

constitutive of thought can also be normative — then the dual role thesis holds out the prospect of an account of how certain cognitive norms might possess categorical authority. Any norm deriving its authority from its constitutive role in thought would possess that authority not by virtue of any desires or inclinations on the part of a thinker, but simply by virtue of an individual's status as a thinker.

My general strategy in what follows will therefore be to argue for the following two points: first, that the dual role thesis is coherent — that a principle can play both normative and constitutive roles in thought — and second, that that thesis does in fact allow us to make sense of the categorical conception in a way that withstands naturalist objections. But in working out this strategy, there are several subsidiary points that require development, so let me briefly outline the plan of the dissertation.

In Chapter Two I will introduce the hypothetical conception and two of its recent proponents: David Papineau and Stephen Stich. The importance of Papineau and Stich for my project, and for the larger debate between the hypothetical and categorical conceptions, is that they are among the few proponents of the former conception who make a self-conscious effort to argue for that conception as against its rival.[2] In examining their views, therefore, I hope to shed some light on what motivates the hypothetical conception and the opposition among its proponents to the categorical conception: what is at work here, I will argue, is a sense that proponents of the categorical conception have no easy way of accounting for how categorical reasons might motivate us. We can make sense of this charge in terms of the passages I quoted

[2] I should also mention, in this connection, Kornblith 1993, but his arguments against the categorical conception are for the most part rehearsals of Stich's arguments in *The Fragmentation of Reason*.

from Mackie earlier. I explained that, for Mackie, reasons (or values) must be 'action-directing', or capable of motivating us to action. Given this presumption, to make sense of cognitive reasons as categorical we must be able to show how such reasons are capable of motivating us. What Papineau and Stich suggest is that there are no promising accounts of how categorical reasons might motivate us, and therefore that the very notion of a categorical reason is an illusion.

Chapter Three is primarily interpretive, arguing that we can find in Frege and Davidson the three theses I've mentioned: the normative, constitutive, and dual role theses. The third of these theses, I argue, is incompatible with the hypothetical conception, and holds out the prospect of a way of making sense of the categorical conception. However, as I've already suggested, there is a prima facie case to be made for the incoherence of the dual role thesis, so in Chapter Four I explain the difficulties with that thesis and argue that, properly understood, it is not only coherent, but also provides valuable materials for making sense of the categorical authority of certain cognitive norms. The key to recognizing its coherence — that is, the key to understanding how the same principle can play both normative and constitutive roles in thought — is a notion that I borrow from Davidson — the notion of a normative concept. If we regard certain cognitive concepts — the concept of belief, for example — as normative, in the sense I specify in Chapter Four, then we can render coherent the dual role thesis. I should emphasize that while my aims in Chapter Three are interpretive, the attempt in Chapter Four to resolve the difficulties with the dual role thesis requires going beyond interpretive claims about Frege and Davidson; although there is a strand of thought in their works that is congenial to the categorical

conception, it is incomplete as it stands.

Finally, in Chapter Five, I flesh out the account of cognitive normativity introduced in the preceding chapter and argue that it is not susceptible to the standard objections to the categorical conception that I describe in Chapter Two. In particular, I focus on the objection mentioned earlier — what we might call the motivational objection to the categorical conception: there are no categorical reasons, cognitive or otherwise, because reasons by their very nature are capable of motivating us to action, and there is no plausible account of how categorical reasons could do so. My response to this objection is, at one level, straightforward — I construct what I take to be a plausible account, or at least the outline of such an account, of how cognitive reasons can be both categorical and have motivational, and therefore normative, force for thought. But at another level the details of the response are complicated by the strategy I've adopted in the rest of the dissertation. By arguing that we can make sense of the dual role thesis, and therefore the categorical conception, on the assumption that certain concepts in terms of which we understand cognition are normative concepts, I am essentially providing a conceptual basis for our understanding of cognitive normativity. But, as I point out in Chapter Two and elsewhere, one important thread of the naturalist argument against the categorical conception is the view that conceptual claims lack any motivational force. Now if worries about the categorical conception tend to converge on the idea that categorical reasons cannot motivate us, and are therefore not really reasons at all, and those who share such worries also tend to hold that conceptual claims lack motivational force, then the prospects for an argument that seeks to rehabilitate the categorical conception by basing it on

conceptual claims begin to look somewhat dim. So part of the burden of my discussion in Chapter Five is to defend the idea that conceptual claims *can* have motivational force.

I cannot claim to have exhaustively developed the categorical conception of cognitive normativity that I defend here — as I point out in Chapter Six, there are various ways in which the account I have begun here requires further elaboration. Nevertheless, I hope to have developed and defended it in sufficient detail to restore some plausibility to the idea that certain cognitive norms might be categorical.

Chapter 2: The Hypothetical Conception

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me normative epistemology is a branch of engineering. It is the technology of truth-seeking, or, in a more cautious epistemological term, prediction. . . . There is no question here of ultimate value, as in morals; it is a matter of efficacy for an ulterior end, truth or prediction. The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed. [Quine 1986, 664-5]

The view that Quine expresses here is a version of what I am calling the hypothetical conception. To say that normative epistemology is a branch of engineering is to suggest that whatever advice the normative epistemologist gives us is contingent upon our specification of a 'terminal parameter', a goal that we seek to attain in our cognition. This goal is typically truth, and so epistemology is more often than not" the technology of truth seeking", but Quine's insistence that we must specify a terminal parameter presupposes that cognition need not be directed toward truth. Whatever the goal we choose, the normative authority of the epistemic or cognitive principles that the epistemologist recommends is not 'ultimate', either in the sense that these principles depend on nothing else for their authority, or in the sense that any rational thinker must accept them as authoritative. Whatever authority such principles possess is derived from, and wholly contingent upon, our desire for the end toward which they are conducive, and we cannot assume that this end must be truth. Hence, if truth is not one of our cognitive goals, we need not take seriously the normative principles that the "technology of truth-seeking" yields. Quine's view, then, seems to be this: all normative principles governing cognition derive their authority from our desire for ends to which

adherence to those principles is conducive. This is simply a statement of the hypothetical conception.

I have a number of goals in this chapter. My most basic goal is simply to illustrate the hypothetical conception by examining some of its proponents: I have done that in a limited fashion in the discussion of Quine above, but I will look at Stephen Stich's development of the view in more detail below. I also want to bring to light some of the motivations fueling the hypothetical conception: I will suggest that the hypothetical conception is motivated primarily by a perception that the categorical conception is incapable of explaining the authority of cognitive norms, that it lacks any compelling explanation for how these norms might motivate us. In developing an account of this motivation, I will look not just at Stich's views, but also at those of David Papineau, who has argued recently against the notion that cognitive norms should be understood as categorical.

Stich and Papineau are also critical of what I have termed the dual role thesis, albeit in somewhat different ways in each case. Recall that the dual role thesis comprises both the constitutive and normative theses, where these are seen as applying to the same set of principles. All parties to the debate I am concerned with accept the normative thesis, so that leaves two routes one might follow in criticizing the dual role thesis: one can reject the constitutive thesis, thereby knocking one of the legs out from under the dual role thesis, or argue directly against the dual role thesis. Stich takes the former route while Papineau takes the latter, and I will sketch their objections to the dual role thesis alongside their critique of the categorical conception. Although the critique of the dual role thesis may seem peripheral to my discussion of the categorical

conception, there is an important reason for connecting the two issues. What unites the two seemingly disparate targets is that they both conflict, in different ways, with the hypothetical conception. The conflict between the hypothetical and categorical conceptions is clear, but to understand the tension between the dual role thesis and the hypothetical conception it is important to recognize that the latter view entails the idea that all normative principles are extrinsic to the nature of cognition. In other words, on the hypothetical view, there are no cognitive norms observance of which is essential to the nature of cognition: the nature of cognition is one thing; the cognitive goals one adopts, and therefore the norms one follows, another thing entirely.[1] This view is inconsistent with the dual role thesis, construed as the view that certain cognitive norms are constitutive of thought, for that is essentially the view that some normative principles are not extrinsic to cognition, but intrinsic or essential.

One final aim that I have in mind in presenting the criticisms advanced by Papineau and Stich against both the categorical conception and the dual role thesis is to lay out some of the criticisms to which I must respond in trying, later on, to articulate a plausible version of the categorical conception of cognitive normativity.

[1] Consider, for example, Quine's insistence that the advice given to us by the normative epistemologist is contingent upon our specification of a 'terminal parameter', or Papineau's claim that "From the naturalist point of view, sensitivity to norms of judgement is an addendum to the possession of beliefs itself. Such sensitivity can arise in beings with beliefs, but it does not have to" (NJ, 32). I discuss this claim at the end of Section 2.2.

2.1 Stich and the failure of analytic epistemology

Stephen Stich's antipathy toward the categorical conception, as well as the constitutive thesis, is tied to a suspicion of the aims and methods of analytic epistemology, which "proposes to ground an account of the justification of cognitive norms in an *analysis* or *explication* of the commonsense notion of justified inference" (FR, 75). Stich takes issue with the idea that the conceptual analysis characteristic of analytic epistemology can yield much insight into the nature of cognition or the norms governing it, and he sees attachment to both the constitutive thesis and the categorical conception as involving a misguided commitment to this mistaken idea. In place of the categorical conception, he argues in *The Fragmentation of Reason* for the view that all cognitive and epistemic normativity is 'extrinsic': cognitive and epistemic norms have authority over us only to the extent that they are conducive to our attaining whatever it is that we find "intrinsically valuable" (FR, 132).

Because cognitive normativity is extrinsic, and people vary in what they value, Stich advocates what he terms "normative cognitive pluralism", according to which "there is no unique system of cognitive processes that people should use, because various systems of cognitive processes that are very different from each other may all be equally good" (FR, 13). This view is opposed to normative cognitive monism, which asserts that there is a unique set of norms telling us how we ought to think. Normative cognitive pluralism, Stich argues, entails a form of relativism concerning the norms of cognition — a form of relativism that, as he puts it, maintains that "assessments of cognitive systems. . . are sensitive to facts about the person or group using the system" (FR, 136). Foremost among the facts to which such assessments are sensitive are the

agent's cognitive abilities and limitations as well as what she finds intrinsically valuable.

Much of *The Fragmentation of Reason* is devoted to dismantling positions that represent obstacles to the acceptance of normative cognitive pluralism. One such position Stich terms 'descriptive cognitive monism'. This view holds that there are strict constraints on how people can think, on the cognitive processes available to us for the formation and revision of beliefs. The relevance of this view to Stich's pluralism should be clear: any constraints on how people can think entail corresponding constraints on how they ought to think. Such constraints threaten the idea behind pluralism, that there are a variety of ways in which people ought to think. What I have termed the constitutive thesis is a form of descriptive cognitive monism that Stich is particularly concerned to refute. As he characterizes the view, it argues for the existence of conceptual constraints on thought. The basic idea is a familiar one: there are certain rational principles to which mental states must conform to some extent if they are to be intentionally characterizable, and the requirement that they must so conform is simply part of our concept of intentional content. In the absence of any possibility of intentional characterization we have no reason for regarding a person as thinking at all, since intentional states are essential to thought. So anything that is to count as thought must conform to a specific set of rational norms.

Stich's response to this form of cognitive 'conservatism' starts with an attempt to explain the link between content ascription and rationality, or conformity to rational norms. Once we grasp the proper explanation of this link, Stich argues, we'll see that no substantive conclusions concerning the potential diversity of forms of thought

follow from the conceptual constraints proposed by advocates of the constitutive thesis. The key to this explanation is an understanding of what we are doing in making ascriptions of the form 'S believes that p'. First, we are attributing a cognitive state of a particular sort to S. Second,

> we are using the content sentence, 'p', to identify the particular belief we are attributing. . . by first picking out a hypothetical belief state that we ourselves might have — the one which in this setting we would express by uttering 'p' — and then by attributing to S a belief state which is similar to this one. To say 'S believes that p', then, is to say S has a belief state similar to the one which would underlie my own assertion of 'p' were I (just now) to have uttered 'p' in earnest. [FR, 49]

On the assumption that we, as interpreters, are for the most part rational in our beliefs, any speakers we interpret will turn out to be mostly rational as well. The link between content ascription and rationality, then, is a result of the fact that in ascribing beliefs or other intentional content to others we use ourselves as models. But, Stich argues, this argument for the necessity of conceptual constraints does not, by itself, yield any substantive metaphysical conclusions concerning the nature of thought or its limits. This link between content ascription and rationality represents "an observer-relative, situation-sensitive constraint that marks no natural or theoretically significant boundary. Rather, it follows the contours of a commonsense concept" (FR, 51-2; see also 12).

Stich's disparagement of these conceptual constraints for their dependence on a "commonsense concept" is an important claim, for something like it informs his rejection of both the constitutive thesis and the categorical conception. It is related to his description of such constraints as 'observer-relative' and 'parochial' (FR, 12). All of

these characterizations are meant to draw attention to the origins of our concepts in an effort to emphasize their contingency. He argues, for example, that the potential for cognitive diversity, and the absence of any way of demonstrating the superiority of one's own cognitive processes, suggest that "it is ultimately no more than a historical accident that we use the cognitive processes we do or that we hold the beliefs that those processes generate, just as it is an historical accident that we speak English rather than Spanish, and wear trousers rather than togas" (FR, 91). Earlier he claimed that "if descriptive cognitive pluralism is true. . . then most of this [cognitive] divergence is likely to be traceable to cultural differences, though genetic factors and idiosyncratic differences in individual experience may also play a role" (FR, 20). Our own thought, as well as the concepts in terms of which we characterize it, are contingent products of cultural and biological factors, as well as "idiosyncratic differences in individual experience", and cannot, for this reason, be regarded as providing any sort of definitive criteria for what counts as thought.

Stich argues, then, that the problem with the constitutive thesis and other arguments for the necessity of conceptual constraints on what can count as thought is that they rely on 'parochial', 'observer-relative', and 'commonsense' concepts. In an earlier treatment of this same topic Stich had claimed that as a result of the link between content ascription and rationality "there will be no comfortable intentional characterization for the cognitive states of subjects whose inference patterns or whose stock of beliefs are radically different from our own" (Stich 1984, 227). This claim captures the thrust of Stich's argument against the constitutive thesis and related views rather well: what it suggests is that the only conclusions we can reasonably draw from

the constraints on content ascription are epistemic ones about the limits of our understanding of others who are sufficiently different from us cognitively, not a metaphysical conclusion about what sort of cognitive 'others' there could be.

Stich also rejects the categorical conception, insisting that "there are no intrinsic epistemic virtues" (FR, 24) — in other words, no epistemic or cognitive norms have authority independently of the ends we hope to achieve by adhering to them. The strategy he pursues against the categorical conception is similar to the one he follows in his arguments against the constitutive thesis. He focuses on the use of 'commonsense' normative concepts — those that are "embedded in everyday thought and language" (FR, 90) — in the evaluation of cognition and argues that their contingent and parochial nature belies their pretensions to unconditional authority:

> . . . being sanctioned by those concepts and practices [embedded in everyday thought and language] is of no particular value to most of us. For even if it should turn out that our own evaluative notions are reasonably coherent, systematic, and stable, they mark only one spot in a rich and varied space of possible (and probably actual) alternatives. If the principal reason that our evaluative epistemic concepts, concepts like rationality and justification, stand out from the crowd is that they happen to be the evaluative notions passed on by our language and culture, it's hard to see why anyone but an epistemic chauvinist would much care whether his cognitive processes were sanctioned by those notions. [FR, 130]

Stich offers what is, in essence, an error theory explaining philosophical commitment to the categorical conception. We mistakenly invest certain normative concepts with intrinsic or categorical authority only because we fail to recognize that these concepts represent an arbitrary selection from a wide range of competing normative concepts: ". . .once the arbitrariness and idiosyncrasy of our own concept of

justification is clearly understood — once it is seen that the notion . . . is but one among many possible alternative notions — most people are not much inclined to say that they find having justified beliefs to be intrinsically valuable" (FR, 95). What Stich suggests here is that the most sense we can make of the notion of categorical authority is to see it as an illusion engendered by the historical dominance of a narrow range of normative concepts. But historical dominance — the mere fact that these are the concepts and practices that history has bequeathed to us — signifies nothing with respect to the legitimate authority of these concepts.

While few of us, according to Stich, will care whether our thinking exhibits allegedly 'intrinsic epistemic virtues' or conforms to putative categorical norms, "most of us do care whether the cognitive processes we invoke are 'best suited to the attainment of our ends'" (FR, 100). The true source of the legitimate authority of cognitive norms lies not in anything intrinsic to those norms, but in something extrinsic: their instrumental efficacy with respect to our ends.[2] The proper way of

[2] As Stich's views here make clear, proponents of the hypothetical conception seek to ground cognitive and epistemic normativity in our desires: we have reason to ϕ, where ϕ-ing is some cognitive action, because we desire something to which ϕ-ing is a means. But having a particular end, and recognizing some means to it that lies within one's power, yields a normative conclusion — the conclusion that one ought to ϕ — only on the assumption that one ought to take whatever means are necessary and available to the ends we have adopted. In other words, there is arguably a principle underlying the normative authority of particular hypothetical imperatives: what Hill calls *the* Hypothetical Imperative: "If a person wills an end and certain means are necessary to achieve that end and are within his power, then he ought to will those means" (Hill, 17-18). And it is because one ought to take whatever means are necessary and available to one's ends that we are justified in

assessing cognitive norms and determining their authority, then, is through an investigation of their pragmatic or instrumental value:

> In evaluating systems of cognitive processes, the system to be preferred is the one that would be most likely to achieve those things that are intrinsically valued by the person whose interests are relevant to the purposes of the evaluation. In most cases, the relevant person will be the one who is or might be using the system. So, for example, if the issue at hand is the evaluation of Smith's system of cognitive processes in comparison with some actual or hypothetical alternative, the system that comes out higher on the pragmatist account of cognitive evaluation is the one that is most likely to lead to the things that Smith finds intrinsically valuable. [FR, 131-2]

Stich is clearly committed to a hypothetical conception of cognitive normativity: all cognitive norms presuppose for their authority certain ends or desires — those things, as Stich puts it, to which we attach intrinsic value — and we can therefore think of them as having the form, "if you desire or value x, you ought to do y", where 'y' specifies certain cognitive processes. Proponents of the categorical conception, on this view, are thrown in with other 'epistemic chauvinists' who perversely persist in their adherence to idiosyncratic and parochial norms that have not been sanctioned by any investigation of their instrumental value.

criticizing someone who fails to take the means to their end as irrational. But it's not clear that we can explain the authority of the Hypothetical Imperative itself in hypothetical terms — in terms of our desires — since it is only in terms of that principle that we can account for the way in which our desires or ends can have normative consequences. Although this is not a point for which I intend to argue, beyond what I've said here, it's not unreasonable to think, therefore, that lurking in the background of the hypothetical conception are categorical norms, among them the Hypothetical Imperative.

I claimed at the beginning of this section that Stich's opposition to the categorical conception and constitutive thesis is related to his opposition to analytic epistemology generally. Let me explain what I have in mind. Among the traditional preoccupations of epistemologists is a concern with epistemic reform: an interest, as Stich puts it, in determining "which ways of going about the quest for knowledge — which ways of building and rebuilding one's doxastic house — are the good ones, which are the bad ones, and why" (FR, 1). Stich's project falls squarely in this epistemological tradition: he advocates what he describes as a "pragmatic theory of cognitive evaluation", according to which we should evaluate cognitive processes in term of their conduciveness to whatever it is that we find intrinsically valuable. Stich contrasts his pragmatic theory with analytic epistemology, the characteristic technique of which, as he sees it, is conceptual analysis; the analytic epistemologist seeks to arrive at an account of the nature of knowledge or cognition, for example, as well as the norms governing thinkers, by analyzing what we mean by 'knowledge', by investigating our concept of knowledge. This analytic procedure presupposes that the concepts in terms of which we think about something like knowledge, and the intuitions and practices associated with those concepts, are our best guide to understanding and improving cognition: "I propose to use the term *analytic epistemology* to denote any epistemological project that takes the choice between competing justificational rules or competing criteria of rightness to turn on conceptual or linguistic analysis" (FR, 91). It is precisely this claim that Stich challenges in his characterization of such concepts as 'parochial', 'observer-relative', and 'commonsense'. His complaint about analytic epistemology is that its obsession with

analysis of the parochial epistemic intuitions and concepts we already have "reduces the normative evaluation of inquiry to a rather bloodless, scholastic preoccupation" that "can hardly be infused with the reformer's zeal" (FR, 16). We must assume, he argues, that our cognitive concepts and practices — the ones we happen to have — are the products of our historical development. Had that development taken a different course, or taken place under somewhat different conditions, we might have ended up with any of a variety of other concepts and practices. Similarly, there could be actual cognitive agents who share none of our cognitive concepts or associated normative commitments. Stich puts the point in the following way:

> [I]magine that we have located some exotic culture that does in fact exploit cognitive processes very different from our own and that the notions of epistemic evaluation embedded in their language also differ from ours. Suppose further that the cognitive processes prevailing in that culture accord quite well with their evaluative notions, while the cognitive processes prevailing in our culture accord quite well with ours. Would any of this [i.e., the effort on the part of the analytic epistemologist to analyze, for example, our commonsense concepts of justification or rationality, and thereby to arrive at a standard for evaluating cognitive processes] be of any help at all in deciding which cognitive processes we should use? Without some reason to think that one set of evaluative notions [or one set of cognitive concepts] was preferable to the other, it seems clear that for most of us it would be of no help at all. [FR, 92-3]

The possibility of alternative systems of cognitive processes, intuitions, and concepts confronts us with the question of the basis on which we are to choose between competing systems. If analytic epistemology is effectively a mere documentation of what is familiar to us, cognitively speaking — a form of "domestic cognitive anthropology" (Stich 1991, 209), as Stich puts it elsewhere — we cannot count on it to provide us with meaningful advice in such matters. To rely, with the analytic

epistemologist, on mere explication of our own normative cognitive intuitions as a way of settling this question — to take for granted, for example, that there is something intrinsically valuable about a belief's conforming to our own concept of justification, as opposed to being sanctioned by some alternative normative concept — is essentially to trumpet our own cognitive system simply on the basis of its being ours. Analytic epistemology, and the categorical conception and constitutive thesis with it, represent a kind of cognitive complacency, taking as authoritative the parochial concepts, intuitions, and practices with which we are familiar. It is only by assigning far too much importance to our own, idiosyncratic ways of thinking about cognition that we arrive at the constitutive thesis and the categorical conception; these views are really driven by the same misguided "epistemic chauvinism" that drives analytic epistemology. Stich's proposal is that we take control of our cognitive lives and, rather than blindly following principles possessing merely de facto influence over our thought, determine which principles *ought* to influence us. We do this not by taking for granted the value of our own epistemic intuitions, but by determining which cognitive processes are most conducive to our ends.

2.2 Papineau and the mysteriousness of the categorical

Papineau defines his position by contrast to what he identifies as 'non-naturalist' views of cognitive normativity and intentional content. Although he does not describe the views he opposes in precisely these terms, his targets are essentially the categorical conception and the dual role thesis. He sees the connection between these two views in the following way. The distinguishing feature of non-naturalist theories of content is

that they "place normativity inside the analysis of content" (NJ, 20); on the non-naturalist view, in other words, any plausible account of the truth-conditional or intentional content of such things as beliefs must appeal to a "peculiar species of content-constituting norms" (NJ, 22). These norms are sui generis in the sense that they are distinct from, and cannot be assimilated to, ethical or hypothetical norms. To place them 'inside' the theory of content is to regard them as an essential element of any such theory — to maintain, for example, that we cannot explain the content of particular beliefs without appealing to the normative relations between beliefs. Papineau cites Davidson as a proponent of this view, because he "take[s] content to depend inter alia on facts about when it is reasonable to form a belief" (NJ, 20). Papineau's opposition to non-naturalism is, at least in part, a rejection of the dual role thesis.[3] His rejection of that thesis depends on two ideas: first, that we need not, and ought not, explain truth-conditional content by appeal to a "peculiar species of content-constituting norms" — we can "place normativity outside the theory of content" (NJ, 22) — and second, that the view of normativity presupposed by non-naturalism — essentially, the categorical conception — is singularly implausible and unmotivated.

[3] This point requires some clarification. Papineau's 'non-naturalism' is the view that there are norms constitutive of *thoughts* — that is, norms that play a special 'content-constituting' role with respect to units of truth-conditional content such as thoughts and beliefs. I have described the dual role thesis as the view that there are norms constitutive of *thought* — norms, in other words, constitutive of one's thinking, or being a thinker. The connection between the two theses lies in the idea that these norms are constitutive of thought *because* they are constitutive of one's having thoughts, beliefs, or other contentful propositional attitudes.

Papineau's defense of the first idea depends largely, but not entirely, on his development of the second idea. Not entirely, because he does cite "the availability of a number of accounts of truth-conditional content which do not assume normativity in explaining truth-conditional content" (NJ, 21).[4] So we need not buy into non-naturalist theories of content simply for lack of naturalist alternatives. But pointing to the availability of naturalist theories of content — theories that "offer to explain truth-conditional content without any commitment to prior norms governing judgement" (NJ, 21) — does not settle which sort of theory we ought to prefer. It is Papineau's treatment of the conception of normativity presupposed by non-naturalism that is meant to shoulder this argumentative burden.

His treatment of this issue points to "a number of awkward problems about normativity facing non-naturalist theories of content" (NJ, 22) — problems that a naturalist account avoids. One of these problems is particularly relevant to the debate over the categorical conception. Papineau describes it as a problem concerning "the status of judgemental norms" (NJ, 27):

> If these norms are quite distinct from moral or personal 'oughts', then where do they come from? What kind of fact is it that we categorically 'ought' to reason in certain ways? And whence does the guiding force of these 'oughts' derive — why should we reason in these ways? I know that I should do what is morally required. And there is an obvious sense in which I should do what will get me what I want. But I find myself in difficulty understanding why I should be moved by the non-naturalists' putative sui generis norms of judgement. . . . Even if we can't reduce judgemental norms to other kinds of facts, it is surely desirable that we should have some kind of understanding of the peculiar force that judgemental norms are supposed to exert on us. [NJ, 27-8]

[4] Among the alternatives are "such theories as indicator semantics. . ., success semantics. . ., teleosemantics. . ., and Fodor's 'asymmetric dependence' theory of content" (NJ, 21).

It is on the characterization of the non-naturalists' norms as 'sui generis' that I want to focus first. The issue here is how we ought to understand the authority of such norms; this is clear in Papineau's insistence that "we should have some kind of understanding of the peculiar force that judgemental norms are supposed to exert on us." In particular, the issue is whether cognitive norms are 'sui generis': whether they possess a form of authority "quite distinct from moral or personal 'oughts'." If cognitive norms are sui generis, then they will be non-hypothetical, or categorical, norms whose authority is distinct from that of ethical norms.

One point of clarification: I have described the position opposed by Papineau as the view that there are categorical norms governing thought, but it should be clear now that this characterization is slightly misleading. The question is not whether there are categorical norms governing thought, but whether there are sui generis categorical cognitive norms. It is important to clarify this point, because Papineau does distinguish between moral and personal 'ought's, and might therefore be willing to acknowledge the possibility of categorical norms governing thought, provided we understand that the source of their authority lies with ethical considerations rather than anything distinctively cognitive. In other words, Papineau seems open to the possibility that categorical ethical norms might sometimes have consequences for how we ought to think. This is a possibility proponents of the categorical conception could acknowledge as well — it is not part of that view that ethical and hypothetical considerations are absent in cognition — but Papineau's openness to it should not be seen as a concession to what I am describing as the categorical conception. As we will see, what Papineau is

not open to is the possibility of categorical cognitive norms whose authority is not derived from other sources, such as hypothetical or ethical forms of normativity. It is commitment to the existence of such sui generis cognitive norms that I take as characteristic of the categorical conception.[5]

I have claimed that Papineau is opposed to the categorical conception, but after the clarifications I have made in the preceding two paragraphs, it may be worthwhile to consider briefly the question of whether he should be regarded as a proponent of the hypothetical conception. After all, his acknowledgment that cognitive norms may be derived moral 'ought's seems to leave room for him to argue that some cognitive norms, while not sui generis in their authority, may nevertheless be categorical rather than hypothetical.[6] But because his opposition to the categorical conception seems to be driven, in large part, by opposition to the very notion of the categorical, it is hard to

[5] In characterizing the categorical conception as I have in this paragraph, I have assumed that there is a clear distinction to be made between cognitive and ethical norms. There are, however, conceptions of the ethical on which this distinction will be much more difficult to make. If, for example, we adopt a virtue ethical approach, and take the scope of ethics to be human flourishing generally, cognitive norms may very well fall within the scope of the ethical. Or one might argue that from a Kantian perspective, where ethical norms are ultimately justified as principles of reason, the ethical falls within the scope of the cognitive. So I should emphasize that the distinction I am making here between cognitive and ethical norms depends on a narrow conception of what counts as ethical, similar to one articulated by Scanlon in *What We Owe to Each Other*: his concern there, he explains, lies with a "narrower domain of morality having to do with our duties to other people, including such things as requirements to aid them, and prohibitions against harming, killing, coercion, and deception" (Scanlon, 6). I would add, though, that some of the duties that Scanlon describes here under the heading of morality are also things we owe ourselves.

see how his suspicion of that notion in the case of cognitive normativity ought not to carry over to ethics. When he asks "[w]hat kind of fact is it that we categorically 'ought' to reason in certain ways?" (ibid.), it is natural to ask what difference there is between the notion of the categorical in cognition and the same notion in morality that renders the former so baffling and the latter clear enough that Papineau can confidently say, "I know that I should do what is morally required" (ibid.). Particularly in view of the fact that traditional accounts of categorical norms (Kant's, for example) trace the authority of such norms back to the cognitive sphere, identifying it with that of principles of rationality, there is good reason to wonder whether commitment to any form of categorical normativity is consistent with the position Papineau defends.[7] In any case, it is clear that Papineau conceives of cognitive norms as hypothetical, which is enough to identify him as a proponent of the hypothetical conception, as I have defined it.

Having clarified some of the views at issue here, I now want to look more closely at Papineau's objection to the categorical conception as it appears in the long passage quoted above. The importance of this passage lies in its clear articulation of the dissatisfaction that many naturalists feel at the prospect of categorical cognitive norms. At the end of that passage Papineau explains that "[e]ven if we can't reduce judgemental norms to other kinds of facts, it is surely desirable that we should have

[6] What this shows is that one need not oppose the notion of the categorical in all normative arenas in order to oppose what I am calling the categorical conception. One need only oppose the idea that there are *sui generis* categorical cognitive norms.

[7] I discuss this traditional account at greater length below, in Section 2.3.

some kind of understanding of the peculiar force that judgemental norms are supposed to exert on us." What the rest of the passage suggests is that the problem with the categorical conception is that it provides no such understanding. While we can make good sense of cognitive norms grounded in human desires and the means for pursuing them, or perhaps even in ethical imperatives, the idea of sui generis, categorical 'ought's that any thinker must heed is completely mysterious. And in the absence of any convincing explanation of the authority of such 'ought's, we do best to conclude that their alleged sui generis authority is illusory.

If non-naturalist theories of content are committed to the existence of such mysterious norms, Papineau concludes, this is all the more reason to prefer naturalist theories of content that avoid such problematic commitments. Although naturalist theories reject the idea that we must appeal to some form of normativity in explaining intentional content, they are not for this reason incapable of explaining the role of normativity in judgment: according to Papineau, "if you have a naturalist [theory] of content, then you can explain norms of judgement as derived prescriptions orientated to the end of truth" (NJ, 22). Norms of judgment, or cognitive norms, have authority over our thought only insofar as adherence to them is conducive to attaining ends we value. Papineau describes this view in more detail in the following passage:

> [T]he most significant norms of judgement can be viewed as prescriptions to the effect that, in order to achieve the truth, you ought to judge in such-and-such ways. In my view, there is nothing constitutively normative about the end of truth itself. So I take the force of the prescriptions to derive from independent moral or personal reasons for attaching value to truth. [NJ, 18]

Cognitive norms are "derived prescriptions": their normative force is a product of "independent moral or personal reasons for attaching value to truth." Obviously what Papineau has in mind here are truth-conducive norms of judgment, and his claim is that, absent some moral demand or personal desire for the truth, there is no reason to follow these norms. In certain circumstances failure to follow cognitive norms may very well be immoral or imprudent, but there is no further sense in which such failure is genuinely mistaken or wrong.[8]

Non-naturalism, according to Papineau, is mistaken on two counts, corresponding to the two senses in which it takes cognitive norms to be necessary. First, cognitive norms lack normative necessity: they impose no sui generis categorical demands on our thought. One who places little value in truth or consistency is neither bound in any way to respect the norms associated with their pursuit nor blameworthy, in any uniquely cognitive sense, in the event that she fails to do so. Second, cognitive norms lack what we might call descriptive or nomological necessity: the character of thought and judgment is in no way constrained by these norms, so there is no need to keep these norms 'inside' the theory of content. In fact, the possession of beliefs and other contentful psychological states does not require that one follow any norms at all: "From the naturalist point of view, sensitivity to norms of judgement is an addendum to the possession of beliefs itself. Such sensitivity can arise in beings with beliefs, but it

[8] Unless, of course, we take 'wrong' or 'mistaken' in this context as having no normative connotations and as merely pointing to the absence of the property of truth-conduciveness. In other words, there is no further sense, on Papineau's view, in which failure to follow cognitive norms is wrong or mistaken if we take those words to imply that one ought to have done otherwise.

does not have to" (NJ, 32). Sensitivity to norms is, of course, distinct from conformity: the former suggests self-conscious efforts to follow norms, while the latter does not. But Papineau is committed to the idea that neither sensitivity nor conformity is essential to the possession of contentful psychological states. We need not be sensitive to them, because we may have no interest in the goal or goals to which they lead. But to say that norms of judgment fall outside the theory of content is to maintain that there is no need to appeal to such norms at all in explaining what constitutes content; there is no essential or necessary connection between these norms and thinking, having beliefs, or making judgments. This view of the role of normativity in the theory of content encompasses not just the idea that we need not be sensitive to such norms in our thought, but also that we need not conform to them.

2.3 The categorical conception: mysterious or mistaken?

Both Stich and Papineau seem to think there is something mysterious about the categorical conception. Papineau gives voice to this charge in his insistence that "it is surely desirable that we should have some kind of understanding of the peculiar force that judgemental norms are supposed to exert on us" (NJ, 28). And one might argue that Stich's failure to consider anything but a debunking explanation of such norms — an explanation to the effect that the only reason we regard certain processes as intrinsically valuable, or certain norms as categorical, is because of the historical blinkers that prevent us from seeing alternatives to them — is itself evidence of pessimism concerning the prospects for meeting something like Papineau's demand. They share the view that any account of the authority of cognitive norms must be able

to explain why we should care about following such norms, and the charge of mysteriousness reflects a sense that the categorical conception fails this test. Papineau has difficulty understanding why he "should be moved by the non-naturalists' putative sui generis norms of judgement." On Stich's view, recognition of the "arbitrariness and idiosyncrasy" of our own normative concepts will tend to disabuse us of the idea that we have any reason to take seriously ascriptions of "intrinsic value", or categorical authority, to those concepts. But where the categorical conception fails, the hypothetical conception seems a clear success: unlike the former, the latter is supposed to explain the authority of cognitive norms in terms that are clear and indisputable. As Papineau puts it, "there is an obvious sense in which I should do what will get me what I want." And, according to Stich, while few of us will care whether our thinking exhibits allegedly 'intrinsic epistemic virtues' or conforms to putative categorical norms, "most of us do care whether the cognitive processes we invoke are 'best suited to the attainment of our ends'".

The disparity that Stich and Papineau note between the categorical and hypothetical conceptions seems genuine, at least if we stick to typical characterizations of what it means for a norm to be categorical or hypothetical. To say of a norm that it is hypothetical is not simply to characterize its authority, but to explain the ground for that authority: it is grounded in the desires of the agent. In contrast, to describe a norm as categorical is merely to characterize its authority — categorical norms are those whose authority is not grounded in desires — and not to explain the basis for that authority. Moreover, given these definitions, the notion of the categorical seems parasitic on the notion of the hypothetical, in the sense that categorical norms are

defined as norms lacking a certain feature of hypothetical norms — but lacking precisely that feature that serves to explain how hypothetical norms fit into our psychological and motivational economy, without the specification of any alternative explanation. So, given the typical characterization of what it means for a norm to be categorical, there is some justification for regarding the notion as mysterious.

Now one might protest, of course, that there is a particular account of normative authority associated with the notion of the categorical. In the *Groundwork of the Metaphysics of Morals*, for example, Kant describes categorical imperatives as representing an action "as *in itself* good, hence as necessary in a will in itself conforming to reason, as its principle" (Kant 1997, 4:414). Categorical ethical norms, at least on the Kantian view, appeal to us not as beings with this or that desire, but as rational beings: their authority is that of principles of rationality. With a ready answer to the question of whence categorical norms derive their authority, the argument continues, there is no basis for charging the categorical conception with mysteriousness. The problem with this response, however, is that our concern here is not morality, but cognition and the principles governing it, *including* principles of rationality. Stich and Papineau are plausibly read as challenging the idea that there is anything intrinsically valuable about the processes or states that these principles recommend, that there is anything categorically normative about principles of rationality themselves. Stich, for example, whose concern lies with "normative standards of cognition" generally (FR, 75), maintains that "the fact that certain inferences are rational and certain beliefs justified. . . is not a state of affairs that many of us find intrinsically valuable" (FR, 99). Because it is precisely the sort of principles

that Kant identifies as the source of categorical normativity whose authority is in question here, appeals to those principles cannot address Papineau's demand that we have "some kind of understanding of the peculiar [sui generis, categorical] force that judgemental norms are supposed to exert on us." A fundamental aspect of the categorical conception — its explanation of the authority of categorical norms — remains mysterious.

Merely to charge the categorical conception with mysteriousness, however, is to leave open the possibility that some account might be available that would dispel the mystery, that we might be able to give an account of how categorical norms get a grip on us independently of our desires. But let me briefly sketch one more reason for skepticism on the part of naturalists concerning the notion of the categorical — a reason for thinking not simply that the categorical conception is mysterious, but mistaken. What I have in mind is often described as the Humean theory of motivation.

According to the categorical conception, at least some cognitive principles have categorical authority over thought. To speak of authority here is simply to indicate that we have reasons for following these principles. To characterize their authority as categorical is to describe what kinds of reasons they give us, or why we accept their demands: we have reason to follow them — we accept their demands — regardless of whatever desires we happen to have. The Humean theory, as I am characterizing it here, draws its inspiration from passages in Hume like the following: first, that "morals excite passions, and produce or prevent actions" (Hume, 457), and second, that "Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them" (ibid., 415). Hume's concern,

particularly in the first passage, lies with moral motivation, but his points can be extended to other normative contexts as well. The first element of the Humean theory, generalizing from the former passage, is that it is characteristic of reasons (moral or otherwise) that they are capable of motivating us to action; that is, on at least some occasions our reasons serve to explain why we act the way we do. The second view, summarized in the latter passage, is that our cognitive faculties do not "produce or prevent actions" — they do not serve as a source of motivation — and therefore that motivation always involves non-cognitive elements, such as desires. The role of cognition in action is limited to the formation of beliefs concerning the means to the satisfaction of our desires; reason is "the slave of the passions". The combination of these views yields the conclusion that if one is to have a reason to ϕ, one must have both a desire (or some similar non-cognitive state) and a belief that ϕ-ing will satisfy or contribute to the satisfaction of that desire. But given this conclusion, there are no categorical reasons. If a categorical reason is one that is capable of motivating us regardless of our desires, there can be no such reasons on the Humean view, for all reasons derive their normative and motivational force from antecedently existing desires on the part of the agent.

It's not clear whether Stich or Papineau subscribes to the Humean theory, but it is consistent with many of their complaints against the categorical conception, as well as with their view of the hypothetical conception as a providing a paradigmatically clear account of cognitive normativity. I mention it here, however, not because I intend to make a case for their commitment to it, but simply because it is an indispensable part of the explanation for naturalist opposition to the categorical conception.

2.4 Conclusion

In this chapter I have sketched some of the criticisms leveled by naturalists against the categorical conception and the dual role thesis. No doubt there are criticisms I've overlooked, but one of the things I have tried to emphasize in the case of the categorical conception is the claim, evident in different ways in Stich, Papineau, and the Humean theory, that it does not provide an adequate account of the authority of cognitive norms.

One of the issues I addressed in the preceding section was the question whether there is any clear account of normative authority available to proponents of the categorical conception. What I want to explore in the rest of the dissertation is whether what Papineau terms 'non-naturalism' is itself a serviceable account of the authority of categorical cognitive norms. Recall that non-naturalism, as he defines it, holds that there is a "peculiar species of content-constituting norms" — norms, in other words, that are constitutive of the intentional content of such things as beliefs. This view is a version of the dual role thesis, of the view that there are principles playing both normative and constitutive roles in thought: these content-constituting norms would be constitutive of thought inasmuch as they are constitutive of intentional content.[9] Papineau dismisses non-naturalism because he sees it as offering no account of normative authority, but he ignores the possibility that the authority of the non-naturalist's norms lies precisely in the fact that they are "content-constituting". So what I will be exploring are the prospects for the idea that the dual role thesis might

[9] For a brief discussion of this point in the context of Papineau's views, see Chapter Two, n. 3.

provide the basis for an account of cognitive normativity consistent with the categorical conception.

Chapter 3: The Constitutive, Normative, and Dual Role Theses

> For there cannot be any doubt: We cannot think or use our understanding otherwise than according to certain rules. . . . All rules according to which the understanding proceeds are either *necessary* or *contingent*. The former are those without which no use of the understanding would be possible at all. . . . Now this science of the necessary laws of the understanding and reason in general, or — which is the same — of the mere form of thinking, we call *logic*. [Kant 1974, 14-15]

> The propositions of logic are 'laws of thought', 'because they bring out the essence of human thinking' — to put it more correctly; because they bring out, or show, the essence, the technique, of thinking. They show what thinking is and also show kinds of thinking.
>
> Logic, it may be said, shows us what we understand by 'proposition' and by 'language'. [Wittgenstein 1978, I, §§133-4]

On the view expressed in these passages, certain logical principles are inextricably linked to the nature of thought. For Kant they are rules that reveal the form of thought and "without which no use of the understanding would be possible at all", while for Wittgenstein they "show what thinking is," or reveal its 'essence'. Both passages embody a version of what I have called the constitutive thesis: the view that there are certain principles, typically logical, observance of which is constitutive of thought. What these passages suggest is that failure to observe these principles is a failure to think, a failure to meet the conditions that would justify the characterization of a person as a thinker. The constitutive thesis is common to Kant, Wittgenstein, Frege, and Davidson, but in the latter two we find a second thesis — the normative thesis — conjoined with it in a particular way. This is the claim that there are certain principles that are normative for thought, and both Frege and Davidson seem to regard the

normative thesis as holding true of the same principles that they identify as constitutive. In other words, they are committed to what I will call the dual role thesis, according to which certain principles are both constitutive of thought and normative for it.

In this chapter I want to focus on these three theses as they appear in Frege and Davidson. The bulk of the chapter will be taken up by discussion of the constitutive thesis: the position that this thesis occupies in their thought is fairly complex, and its presence there is not always immediately evident, so it requires more extensive discussion than the other two principles. The role of the normative thesis in their work, and its implication in the dual role thesis along with the constitutive thesis, should become clear in the course of this discussion. I will have more to say about the dual role thesis at the end of this chapter, and throughout the next, but at this point let me indicate what I take its importance to be.

In the preceding chapter I looked at various critiques of the categorical conception — critiques that also serve to motivate the hypothetical conception. The dual role thesis, I will argue, is incompatible with the hypothetical conception. Assuming that it is compatible with any conception of cognitive normativity (a point I will clarify and address in the following chapter), commitment to the thesis therefore pushes one in the direction of a categorical conception of cognitive normativity. What I ultimately intend to argue is that the dual role thesis offers a way of understanding the categorical conception that is not susceptible to the objections directed against it by philosophers like Papineau and Stich, but my discussion in this chapter merely lays the groundwork for that argument.

3.1 Frege

I want to begin by looking at Frege's conception of the nature of logic and its authority over thought, as a way of getting clear on his commitment to the three theses I have mentioned. I will argue that he subscribes not only to these three theses, but also to the categorical conception.

3.1.1 Psychologism and the autonomy of logic

Much of Frege's writing on logic shows a preoccupation with naturalistic and, in particular, psychologistic accounts of the nature and authority of logical laws. There is considerable ambiguity, however, in his understanding of psychologism. Often he seems to regard it as a thesis concerning the content of logical laws; in these contexts he offers at least two distinct models for understanding the thesis. On one model the laws of logic are descriptive laws bearing a relation to human psychology similar to the relation borne by the laws of physics to nature: they "govern thinking in the same way as laws of nature govern events in the external world" (BLA, 12). On this view logical laws are empirical generalizations describing regularities in human psychological processes. The principle of identity would accordingly be read as elliptical for the claim that "It is impossible for people in the year 1893 to acknowledge an object as being different from itself" (BLA, 14). Alternatively, the psychologistic logician might conceive of such laws as akin to the laws of grammar: they are to be "regarded as guiding principles in the sense that they give an average, like statements about 'how it is that good digestion occurs in man', or 'how one speaks grammatically', or 'how one dresses fashionably'" (BLA, 13). Frege seems to regard these two models as

equivalent, but the second model arguably presents a more complex picture than the first of the descriptive content of logical laws: such laws might, on this view, describe the thought of a competent thinker, where the notion of competence is derived from the actual practices of thinkers in a way similar to the derivation of grammatical rules from the practices of certain speakers. But these two models are equivalent in the sense that, on either conception, the content of the laws of logic is tied to the actual practices of thinkers in such a way that, were these practices to change, the laws themselves would change as well.

Frege rejects all such descriptive conceptions, maintaining that the laws of logic "do not make explicit the nature of our human thinking and change as it changes" (ibid.). Rather, such laws are "boundary stones set in an eternal foundation, which our thought can overflow, but never displace" (ibid.). The laws of logic are unchanging, and therefore cannot describe the vagaries of human thought. Rather than describing how we do think, they prescribe how we ought to think. So the laws of logic are prescriptive or normative, but it is important to recognize that this is not because they take the form of explicit prescriptions. They are normative for the same reason that any truth is normative: "any law asserting what is, can be conceived as prescribing that one ought to think in conformity with it, and is thus in that sense a law of thought. This holds for laws of geometry and physics no less than for laws of logic" (BLA, 12). Like the laws of physics, the laws of logic express truths; the difference between the two sorts of law is that while the former express truths about physical reality, the latter are "the most general laws", expressing truths about everything whatsoever. And just as a law of physics is normative in the sense that it tells us how we ought to think

about physical reality, so a law of logic is normative in the sense that it tells us what we ought to believe, regardless of subject matter. So it is clear why Frege thinks that psychologism misconstrues the content of logical laws: the laws of logic are no more about the course of human thinking than the laws of physics.[1]

In the passages I have discussed thus far, Frege regards psychologism as a thesis concerning the content of logical laws. But at times he seems more concerned with a thesis concerning the authority of logical laws, according to which that authority is grounded only in psychological facts concerning thinkers. One passage in particular, which I will quote at length, reveals Frege's disposition toward this view. I have numbered parts of the passage to facilitate my exposition below.

> [1] The question why and with what right we acknowledge a law of logic to be true [2] logic can answer only by reducing it to another law of logic. Where that is not possible, logic can give no answer. [3] If we step away from logic, we may say: we are compelled to make judgments by our own nature and by external circumstances; and if we do so, we cannot reject this law — of Identity, for example; we must acknowledge it unless we wish to reduce our thought to confusion and finally renounce all judgment whatever. [4] I shall neither dispute nor support this view; I shall merely remark that what we have here is not a logical consequence. What is given is not a reason for something's being true, but for our taking it to be true. [5] Not only that: this impossibility of our rejecting

[1] When Frege speaks of logical laws he seems to have in mind the laws of noncontradiction, excluded middle, and identity. He does distinguish, however, between laws of logic and rules of inference, and it is unclear what status he assigns the latter. In particular, it is not clear whether he regards rules of inference as expressions of truths. The following passage, while not conclusive, can be read as an endorsement of that view: "Following Aristotle, we can enumerate quite a few modes of inference in logic; I employ only this one, at least in all cases in which a new judgment is derived from more than a single one. For, the truth contained in some other kind of inference can be stated in one judgment, of the form: if M holds and if N holds, then L holds also" (B, §6).

> the law in question hinders us not at all in supposing beings who do reject it; where it hinders us is in having doubts whether we or they are right. [6] At least this is true of myself. If other persons presume to acknowledge and doubt a law in the same breath, it seems to me an attempt to jump out of one's own skin against which I can do no more than urgently warn them. [7] Anyone who has once acknowledged a law of truth has by the same token acknowledged a law that prescribes the way in which one ought to judge, no matter where, or when, or by whom the judgment is made. [BLA, 15]

[1] is a statement of the question that Frege spends the rest of the paragraph addressing. This is the question of how we are to explain the authority logic has over our thought, and it is a question to which Frege and the psychologistic logician take very different approaches. Frege sees this question primarily as a request for justification, and explains in [2] that the best we can do in terms of justifying a law of logic is to show that it follows from other logical laws to which we are committed. It is as important to note what Frege does not say here as it is to note what he does. For someone like Mill, logic does not exhaust the possibilities for the justification of logical laws: we can resort to inductive reasoning to justify a logical truth like the principle of non-contradiction.[2] For Frege, in contrast, a move like Mill's is not a viable option: nothing external to logic (e.g., special sciences or experience) can play a justificatory

[2] "I consider it [the principle of contradiction] to be, like other axioms, one of our first and most familiar generalisations from experience. The original foundation of it I take to be that Belief and Disbelief are two different mental states, excluding one another. This we know by the simplest observation of our own minds. And if we carry our observation outwards, we also find that light and darkness, sound and silence, motion and quiescence. . . any positive phenomenon whatever and its negative, are distinct phenomena, pointedly contrasted, and the one always absent where the other is present. I consider the maxim in question to be a generalisation from all these facts" (Mill, 183).

role with regard to logic. Logic is in this way autonomous for Frege: it borrows nothing from other sciences, and depends in no way upon them. So where we exhaust our appeals to other logical principles in attempting to justify a truth of logic, there the justificatory task comes to an end. If, at this point, someone still presses the question of the authority of logic, what they may be after is a psychological account; Frege contemplates such an account in [3]. He does not reject it, but argues in [4] that its legitimate ambitions must be carefully circumscribed: it may well be that it is part of our nature to find certain forms of reasoning compelling, but psychological accounts do not address our initial question, if we take it as a question concerning the justification or legitimate authority of logical laws. If we understand the initial question in this way, then what we are concerned with is truth, and although we can imagine creatures who seem to deny a law of logic, we cannot regard the thoughts they express as true [5].[3]

One problem with psychologism, from Frege's perspective, is that it tries to maintain what he sees as an untenable stance toward logic; this much is clear, in [6], from Frege's characterization of psychologism as "an attempt to jump out of one's own skin." What is less clear is how to make explicit the criticism that this image suggests. Perhaps part of the point of [6] and [7] is that psychologism demonstrates a sort of intellectual bad faith: the psychologistic logician is committed to the laws of logic in his own intellectual practice, but offers a theory that does not exactly support that

[3] Assuming they succeed in expressing any thoughts at all. As we will see, Frege's commitment to the constitutive thesis could be taken to entail that creatures who deny logical laws in the sense that they consistently fail to conform to them are not capable of expressing thoughts at all.

commitment. But we can also read Frege's criticism in another way. Psychologism takes seriously the appearance of necessity that logical laws have, and tries to explain that appearance in terms of psychological necessity, but explains it in such a way that the necessity it sought to explain in the first place has disappeared. Even if logical laws have only psychological necessity for us, then, as Frege says, they must hinder us in thinking that beings who reject these laws are correct in doing so. But the nature of the psychologistic logician's project suggests that these laws do not hinder us in this way, for the psychologistic logician must maintain that in some broader sense, beings who accept logical laws at odds with our own are not thinking incorrectly, but merely differently. If the stance ostensibly taken by the psychologistic logician toward logical aliens is a possible stance for us to take, then logical laws lack even psychological necessity for our thought. So the psychologistic logician presumes "to acknowledge and doubt a law in the same breath": he regards the necessity of logic as a datum to be explained psychologically and then, in the course of his explanation, undermines that datum.

Psychologism, then, is committed to denying that the laws of logic have any authority over our thought, psychological or otherwise. But this denial runs up against what seems to be an incontrovertible fact (a fact, it is worth emphasizing, that the psychologistic logician initially acknowledged and sought to explain): we do regard the laws of logic as having authority over our thought. But if this appearance of authority is a fact, and psychologism denies it, then we might see Frege as suggesting in [7] that the result of the failure of psychologism to account for this fact is that we are thrown

back upon a traditional conception of logic as providing laws both eternal and universal in their application.

Concerning attempts to ground logical laws in psychology, Frege says "I take it as a sure sign of a mistake if logic has need of metaphysics and psychology — sciences that require their own logical first principles. In this case then, where is the ultimate basis upon which everything rests?" (BLA, 18). Whether it concerns the content or the authority of logical laws, psychologism is essentially an attempt to subordinate or assimilate logic to psychology. Truths of logic depend on truths of psychology, and there are no normative questions of the sort that might traditionally have been within the purview of logic that are not answered with reference to psychology: there is no question of how one ought to think over and above how one's psychological constitution disposes one to think (or how human psychology disposes people in general to think). Frege's point in this passage is that attempts like that of psychologism to ground logical laws in other disciplines, or to reduce logical truths to other sorts of truths, are a mistake. Such attempts misconstrue the relationship between logic and other sciences by failing to appreciate the authority and autonomy of logic. Not only does logic depend on no other sciences, but all sciences depend on it.

By characterizing logic as both autonomous and as providing a foundation upon which all other sciences rest, Frege takes issue with the idea that a naturalization of fundamental cognitive norms, like the laws of logic, is possible. Any attempt at naturalization or explanation that construes the authority of logic as hostage to extra-logical (e.g., empirical or metaphysical) fact misconstrues logic. Its truths are more basic than, and moreover, presupposed by, any other body of truths in terms of which

we might try to explain them; this is the point of Frege's characterization of logic as the 'ultimate basis', or as providing the first principles, for other sciences.[4]

3.1.2 The normative thesis: logical laws as categorical

The psychologistic logician fails, on Frege's view, to account for the authority of logic; psychologism is irrelevant to the question of authority, considered as a question concerning justification, and fails to do justice to the idea that logical laws have any sort of necessity for us. But what of Frege's own account of the authority of logic? He regards the account offered by psychologism as fundamentally mistaken, but does he have anything to offer in its stead? On the view he explicitly states in various places, the normative authority of logical laws is tied to their truth. He explains that "any law asserting what is, can be conceived as prescribing that one ought to think in conformity with it, and is thus in that sense a law of thought. This holds for laws of geometry and physics no less than for laws of logic" (BLA, 12). What gives a law normative authority — what makes it that case that "one ought to think in conformity with it" — is the fact that it asserts what is. The suggestion here, then, is that normative authority is a product of truth; any truth can be regarded as prescribing

[4] For a lucid explanation of Frege's claim that the special sciences presuppose or depend upon logic see Goldfarb 2001, particularly pp. 34-5. Goldfarb illustrates how, on Frege's view, logical laws are essential to the justification of any inferences in the special sciences. For example, to infer the conclusion that all whales are vertebrates from the premises, "All whales are mammals" and "All mammals are vertebrates", we need, in addition to the premises, a logical law stating that $(\forall F)(\forall G)(\forall H)[(\forall x)(Fx \rightarrow Gx) \rightarrow ((\forall x)(Gx \rightarrow Hx) \rightarrow (\forall x)(Fx \rightarrow Hx))]$.

how one ought to think. Logical laws, like laws of geometry or physics, express truths; they differ from the latter two types of law only in their generality: logical laws are "the most general laws" (ibid.), expressing truths concerning not just the physical or geometrical, but everything whatsoever. So logical laws are normative for the same reason that any truth is potentially normative: they "have authority for our thought if it would attain to truth" (BLA, 13).

But the idea that logical laws are authoritative in virtue of their truth still leaves open the question of the nature of their authority. One possibility would be that their authority is conditional upon our desire for truth. Logical laws "have authority for our thought if it would attain to truth", but if it would attain to something else — if truth is not something we find intrinsically valuable, as Stich puts it — perhaps we need not heed such laws. If we accept this reading, Frege might seem to be a proponent of the hypothetical conception. But merely to claim that logical laws "have authority for our thought if it would attain to truth" is consistent with maintaining that our thought categorically must strive for the truth, so that claim provides no unambiguous support for the hypothetical reading. Moreover, it is hard to square that reading with Frege's insistence that logical laws possess "unconditional and eternal validity" (BLA, 14), and "prescribe universally the way in which one ought to think if one is to think at all" (BLA, 12). The former passage, on the face of things at least, seems like a straightforward denial of the thesis that the authority of laws of logic is conditional upon anything. And the latter passage seems to underscore this denial, suggesting that it is merely by virtue of one's status as a thinker that logical laws gain their authority over one's thought: "if one is to think at all" — that is, if one is even to engage in

thought, regardless of one's ends — these laws prescribe how one ought to do it. So although Frege offers little detailed explanation of his views concerning the nature of the authority of logical laws, what evidence is available suggests that he is committed to the categorical conception. If logical laws are categorical, and their normativity is derived from their truth, then the source of their categorical authority must lie with the status of truth for thought: what Frege's views seem to imply is that the goal of truth has an authority for thought that is not conditional upon one's desires or aims.

It is important not to misunderstand Frege's view here. Any truth *can* be regarded as prescribing how one ought to think, but that is not to say that one's obligation as a thinker is to seek out truths indiscriminately simply as a way of maximizing one's stock of true beliefs. In a passage elaborating on the analogy between laws of logic and laws of geometry or physics, Frege explains that, just as we should regard logical laws as prescriptive for thought rather than descriptive of it, "[w]e could, with equal justice, think of the laws of geometry and the laws of physics as laws of thought or laws of judgement, namely as prescriptions to which our judgements must conform in a different domain if they are to remain in agreement with the truth" (PW, 145-6). The reference to 'domains' here can be read as suggesting the following picture: within each domain of thought (e.g., physics, geometry, etc.) there are laws to which one must conform — truths one must accept — if one's inquiries are to yield true conclusions. There will be certain geometrical laws that represent laws of thought for the geometer, and certain physical laws that do the same for the physicist, but, with an important exception, what counts as a law of thought within one domain won't

necessarily have the same status within other domains.[5] The laws of logic are the exception: because they state truths concerning everything whatsoever, they are authoritative within all domains of thought.

3.1.3 The constitutive thesis

To explain the authority of logical laws in terms of their truth is to emphasize the similarity between logical laws and other truths. This is a similarity that Frege himself emphasizes, as I pointed out above:

> Any law asserting what is, can be conceived as prescribing that one ought to think in conformity with it, and is thus in that sense a law of thought. This holds for laws of geometry and physics no less than for laws of logic. The latter have a special title to the name 'laws of thought' only if we mean to assert that they are the most general laws, which prescribe universally the way in which one ought to think if one is to think at all. [BLA, 12]

Like laws of physics or geometry, laws of logic prescribe how we ought to think simply because they state truths that we ought to believe. On this view, the only difference in authority between logical laws and other truths lies in their scope: the former are completely general, applicable in all disciplines and to all objects of thought. As Frege points out, "we could, with equal justice, think of the laws of geometry and the laws of physics as laws of thought or laws of judgement" (PW, 145), provided we understand that these laws exert authority over our thought only on the condition that we are

[5] So the geometer needn't make it his aim to master the body of truths known by the physicist, although presumably part of what it means for the geometer to conform his thought to the truth would be to refrain from judging in ways that are inconsistent with the truths of physics.

seeking truth within the domains to which they apply; logic, on the other hand, prescribes principles to which we ought to conform no matter what domain our inquiries belong to.

At various points, however, Frege seems committed to a different conception of the role of logic in thought, according to which logical laws represent truths recognition of which is essential to the very possibility of thought. In the final sentence of the passage quoted above, he explains that logical laws "prescribe universally the way in which one ought to think if one is to think at all." There are various ways of reading this claim, but let me suggest just one: if what one is doing is to count as thought at all, it must conform to, or take as authoritative, the laws of logic. I have described the constitutive thesis as the view that there are certain principles constitutive of thought, and, if we accept the reading I've just suggested, the view to which Frege adverts in this passage is clearly a version of that thesis. As we will see below, this is a view that surfaces throughout Frege's work. Logical laws, for Frege, serve as principles that we must observe in our thought in the following sense: they are truths the recognition of which must be manifest in our thinking, at least at a tacit level. There are several pieces of evidence to support the ascription of this view to Frege.

(1) The constitutive thesis is arguably discernible in Frege's campaign against psychologism. As we have seen, one consequence of the psychologistic claim that logical laws depend on contingent facts concerning human psychology is that these laws are "susceptible to alteration with the constitution of [our] minds" (BLA, 15). On

one reading, part of Frege's anti-psychologistic strategy is to question this consequence of psychologism.

> But what if beings were found whose laws of thought flatly contradicted ours and therefore frequently led to contrary results even in practice? The psychological logician could only acknowledge the fact and say simply: those laws hold for them, these laws hold for us. I should say: we have here a hitherto unknown type of madness. [BLA, 14]

We can certainly imagine beings who often seem to say things of the form 'p and not-p', or who consistently seem to infer in ways we would regard as mistaken, but do we confront in such cases thought that takes as authoritative logical laws different from our own? Frege's invocation of the notion of madness can be taken as a negative answer to this question: what we confront in such cases is not logically alien thought, but merely a semblance of thought. This answer would be in keeping with the constitutive thesis: if logical laws are constitutive of thought, then consistent failures to conform to these laws indicate an absence of thought.

(2) The following passage from the *Foundations of Arithmetic* provides further support for the ascription of the constitutive thesis to Frege:

> For purposes of conceptual thought we can always assume the contrary of some one or other of the geometrical axioms, without involving ourselves in any self-contradictions when we proceed to our deductions, despite the conflict between our assumptions and our intuition. The fact that this is possible shows that the axioms of geometry are independent of one another and of the primitive laws of logic, and consequently are synthetic. Can the same be said of the fundamental propositions of the science of number? Here, we have only to try denying any one of them, and complete confusion ensues. Even to think at all seems no longer possible. The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry. The truths of arithmetic

> govern all that is numerable. This is the widest domain of all; for to it belongs not only the actual, not only the intuitable, but everything thinkable. Should not the laws of number, then, be connected very intimately with the laws of thought? [FA, §14][6]

In §14 Frege has been contrasting empirical and geometrical truths, on the one hand, with laws of arithmetic and logic, on the other, with the aim of showing that the laws

[6] It is worth noting that Austin translates as declarative statements a number of sentences that appear in the original German as rhetorical questions. The following is a more literal translation of the passage:

> In conceptual thought one can still assume the contrary of one or other of the geometrical axioms without becoming involved in any self-contradictions when one proceeds to one's deductions, despite the conflict between these assumptions and intuition. This possibility shows that the geometrical axioms are independent of one another and of the basic laws of logic, and therefore that they are synthetic. Can one say the same of the basic propositions of the science of number? Is it not the case that everything falls into confusion if one tries to deny one of them? Would thinking in that situation even be possible? Is it not the case that the ground of arithmetic lies deeper than that of all empirical sciences, deeper even than that of geometry? Arithmetical truths govern the realm of the countable. This is the most comprehensive realm of all; for not only the actual or the intuitable belong to it, but everything thinkable. Should not the laws of number, then, stand with the laws of thought in the most intimate relation?

Nevertheless Austin's translation does not seem to distort the content of this passage. In any case, even a more literal rendition of the passage preserves the two points central to my reading of §14: first, that logical laws stand to the thinkable as geometrical axioms to the intuitable, and second, that Frege expresses doubt concerning the possibility of thought that fails to observe logical laws.

of arithmetic are neither empirical nor synthetic, but connected "very intimately with the laws of thought." Concerning geometry, he explains that "the truths of geometry govern all that is spatially intuitable, whether actual or product of our fancy." This is not to say that we cannot conceive of what things might be like were these truths different, but to do this we must, he says, "leave the ground of intuition entirely behind" and make use only of "conceptual thought." We cannot intuit spaces that do not conform to the axioms of Euclidean geometry, but this is no impediment to our conceiving of such spaces. So the laws of geometry are constitutive of intuition in the sense that anything we are capable of intuiting, anything we can think of by means of intuition, is governed by these laws. To "shake off this yoke" imposed by the laws of geometry, we must temporarily give up intuitive thought, but we can do this by prescinding from the issue of what we are capable of intuiting and considering alternative geometrical axioms from the perspective of conceptual thought. What Frege seems to suggest in §14 is that the laws of logic stand to thought as Euclidean laws of geometry stand to intuition; just as we cannot intuit anything except in accordance with the laws of geometry, we cannot think while denying the laws of logic. So the laws of logic play a constitutive role in conceptual thought similar to the constitutive role played by geometrical laws in intuition. It is, in part, for this reason that Frege argues that laws of arithmetic must "be connected very intimately with the laws of thought": "we have only to try denying any one of them [i.e., any one of the laws of arithmetic], and complete confusion ensues. Even to think at all seems no longer possible." Part of what leads Frege to conclude that arithmetical laws must be "connected very intimately" with the laws of thought is that a denial of the laws of

arithmetic leads to confusion and even threatens to undermine the possibility of thought. While the laws of geometry govern everything intuitable, the laws of arithmetic and logic govern everything thinkable. The difference between the two sorts of laws is that in the former case we can contemplate geometrical laws being other than what they are, because we can transcend the restrictions imposed on us in intuition and consider these laws solely from a conceptual perspective. In the latter case, however, there is no alternative perspective from which we can consider alternatives to these laws: hence, as Frege says, to contemplate their falsity is to risk giving up thought altogether.

(3) If this interpretation is plausible, §14 of *The Foundations of Arithmetic* lends support to the idea that Frege endorsed, at least at times, some version of the constitutive thesis. But there is also reason to think that certain other views to which he was explicitly committed entail the constitutive thesis. This seems to be the case, for example, in his account of judgment. Frege distinguishes between the act of judgment and a "content of possible judgement". To make a judgment, he explains, is "inwardly to recognize something as true" (PW, 7). But before we are in a position to make a judgment, we must already have grasped the 'something' which we may or may not recognize as true. That which we grasp, he refers to as a "content of possible judgement", or a thought. So it is only after we grasp something concerning which the question of truth arises — a thought — that we are in a position to make a judgment. But what is involved in grasping a thought? Frege compares grasping a thought to

being faced with a propositional question, a question to which we may answer either 'yes' or 'no'.

> A judgement is often preceded by questions. A mathematician will formulate a theorem to himself before he can prove it. A physicist will accept a law as an hypothesis in order to test it by experience. We grasp the content of a truth before we recognize it as true, but we grasp not only this; we grasp the opposite as well. In asking a question we are poised between opposite sentences. Although it is usually only one side that is expressed when we speak, the other is still always implied; for the sense of the question remains the same if we add 'or not?'. . . . Now whatever can thus be posed in a question, we wish to call a content of possible judgement. [PW, 7-8]

So in grasping a thought we face the question whether it or its negation is true. Facing this question is not merely ancillary to grasping the thought; it is an essential element of what it is to grasp a thought, on Frege's view. A thought just is something for which the question of truth arises[7], and one cannot grasp a thought or understand its content without recognizing that the thought confronts one with this question. Frege's description of what is involved in grasping a thought suggests one particular way in which a recognition of the laws of logic might be constitutive of thought. The ability to grasp thoughts is essential to our capacity for thought: "in thinking we do not produce thoughts, we grasp them" (CP, 368). But, on Frege's view, one cannot even so much as grasp a thought without recognizing how logical laws apply to it. The claim that grasping a thought involves facing the question whether it or its negation is true is an expression of the idea that one cannot understand an item with propositional content without also recognizing, at least tacitly, that this is an item subject to such logical laws

[7] CP, 353.

as the principles of noncontradiction and excluded middle. On this view, these laws are constitutive of thought in the sense that the possibility of thought depends on a recognition of them.

(4) The idea that a recognition of logical laws is essential to thought also receives support from Frege's conception of the role of logic in "the conflict of opinions."[8] The following passage is from the Introduction to *The Basic Laws of Arithmetic*, and contains an important part of Frege's response to what he describes as "the corrupting incursion of psychology into logic" (BLA, 12):

> If we could not grasp anything but what was within our own selves, then a conflict of opinions [based on] a mutual understanding would be impossible, because a common ground would be lacking, and no idea in the psychological sense can afford such a ground. There would be no logic to be appointed arbiter in the conflict of opinions. [BLA, 17]

Frege's general concern in this passage is with a psychologistic conception of thought, but there are two aspects to this conception at issue here. Thoughts themselves are identified with subjective psychological entities such as ideas, and laws of thought, or logical laws, are identified with laws describing regularities in the behavior of, and relations between, these entities. If we are to make sense of the possibilities of agreement and disagreement essential to any communication or science, we must reject both aspects: thoughts or meanings must be objective and shareable, and there must be a logic that does more than describe psychological regularities: logic must be capable of

8 The interpretation I offer in this paragraph is by no means original. See, for example, Conant 1992, Cerbone 2000, and Ricketts 1986.

serving as "arbiter in the conflict of opinions." There are at least two senses in which logic must play this role. First, logic aids us in identifying the locus of an agreement or disagreement through our efforts to map the logical structure and presuppositions of the positions at issue. Second, and more fundamentally, logic must be in place for there even to be agreement or disagreement. The possibility of agreement and disagreement presupposes notions of logical equivalence and incompatibility: it is within the context of a recognition, for example, that one cannot consistently hold both p and not-p that disagreements arise and have significance. Logical laws, on this view, provide the backdrop against which agreement and disagreement are possible. To 'disagree' with these laws, then, would be to give up all possibility of agreement or disagreement. If we assume that an ability to discern at least obvious agreements and disagreements is internal to thought, in the sense that one could not be credited with a grasp of thoughts without this ability, then logical laws are constitutive of thought to whatever extent one must follow these laws in order to have this ability.

(5) Further evidence for Frege's commitment to the constitutive thesis can be found in the *Begriffsschrift*. The subtitle of the work is "a formula language, modeled upon that of arithmetic, for pure thought", and it indicates that one purpose of the *Begriffsschrift* is to introduce a script or notation designed for the perspicuous expression of thoughts. What this means, among other things, is that this notation, the begriffsschrift, is not designed to capture everything that we would ordinarily regard as part of an expression's meaning. The relation between the begriffsschrift and ordinary language, he explains, is analogous to the relation between a microscope and the eye: the

microscope is not as versatile as the eye, and it proves useful only in a comparatively narrow range of circumstances, but in those circumstances it allows us a resolution far surpassing anything possible without it. Similarly, the begriffsschrift grants us a precision of expression unattainable in ordinary language, but at the cost of ignoring certain aspects of ordinary language, such as "all those peculiarities. . . that result only from the interaction of speaker and listener" (B, §3). In ordinary language, for example, we often use the subject position in a sentence to direct the listener's attention to a particular element of the thought we express. To take Frege's example, if we wish to focus on the Persians, we may say (a) "The Persians were defeated by the Greeks at Plataea" rather than (b) "The Greeks defeated the Persians at Plataea". These nuances common in ordinary language are not, however, part of the thoughts expressed by our sentences: (a) and (b), for example, express the same thought.

Frege identifies the aspect of meaning relevant to the expression of thoughts as 'conceptual content', and offers the following explanation of this notion:

> . . .the contents of two judgments may differ in two ways: either the consequences derivable from the first, when it is combined with certain other judgments, always follow also from the second, when it is combined with these same judgments, or this is not the case. The two propositions 'The Greeks defeated the Persians at Plataea' and 'The Persians were defeated by the Greeks at Plataea' differ in the first way. Even if one can detect a slight difference in meaning, the agreement outweighs it. Now I call that part of the content that is the same in both the conceptual content. [ibid.]

What is important about this explanation is that it characterizes the conceptual content of a judgment, that aspect of its meaning relevant to the thought expressed, in terms of the logical relations between that judgment and others. Two judgments have the same

conceptual content if they yield the same logical consequences when combined with the same set of judgments. Thoughts are individuated, and conceptual contents identified, by logical relations. Hence the begriffsschrift is designed to make these logical relations as explicit as possible and to represent judgments in such a way that their properties relevant to inference are obvious: "in a judgment I consider only that which influences its possible consequences. Everything necessary for a correct inference is expressed in full, but what is not necessary is generally not indicated" (ibid.). What Frege's notion of conceptual content suggests is that there is an aspect of meaning the grasp of which involves at least a basic understanding of correct inference: if the conceptual content of a sentence, or the thought it expresses, cannot be identified independently of a grasp of its logical relations to other sentences, then one cannot grasp a thought or fully understand a sentence without being aware of these relations. Because such awareness presupposes some understanding of the laws of logic, Frege's notion of conceptual content commits him to the view that our ability to grasp thoughts depends on this understanding. Once again, laws of logic, on this view, are constitutive of thought in the sense that we cannot grasp thoughts, or understand conceptual content, independently of a recognition of these laws.[9]

[9] At the end of the last section I briefly considered how, in light of an account of the authority of logical laws in terms of their truth, we might understand Frege's claim that logic serves as the 'ultimate basis' for other sciences. The importance of the passage in which Frege makes this claim is that it is a statement of Frege's opposition to attempts to naturalize logical laws: his rejection of the idea that "logic has need of metaphysics and psychology" is essentially a rejection of the view that we can explain the authority of logic in terms borrowed from any special science. The conception of

3.2 Davidson

In the preceding section I argued that the constitutive thesis occupies a central position in Frege's thought, even though it is a commitment that he fails to identify explicitly. Like Frege, Davidson is not explicit about his commitment to the constitutive thesis, but in his case too, it is nevertheless arguably central to his treatment of a variety of issues. In this section I will show how Davidson employs the constitutive thesis to support both his treatment of reductionism in the philosophy of mind and his rejection of conceptual relativism.

logical laws as constitutive of thought provides another way of making sense of the claim that these laws serve as the ultimate basis for all other sciences. On this view it is not just the truth of the sciences that depends on observance of these laws, but their very possibility: as paradigmatically rational forms of activity, the sciences involve, as part of their nature, thought. Any constraints on the possibility of thought, then, are also constraints on the possibility of science. If there are laws constitutive of thought, these laws are also constitutive of science. Frege rejects the idea that "logic has need of metaphysics and psychology", and explains that these sciences "require their own logical first principles." This explanation seems to be meant as a justification for the rejection, and suggests something like the following principle (where the notion of 'theory' is to be taken very loosely): if a theory t2 presupposes another theory t1, in the sense that the possibility of t2 depends on the prior acceptance of t1, then t1 cannot be explained in terms of t2. To apply this principle to the case Frege considers, because psychology presupposes logic ("requires its own logical first principles") logic has no need of, or cannot be grounded in, psychological truths.

3.2.1 Psychophysical laws and reductionism[10]

> The mental is nomologically irreducible [to the physical]: there may be true general statements relating the mental and the physical, statements that have the logical form of a law; but they are not lawlike. [EAE, 216]

What Davidson rejects here is the possibility of psychophysical laws: although there may be true generalizations connecting mental and physical phenomena, laws — that is, true generalizations "that support counterfactual and subjunctive claims, and are supported by their instances" (EAE, 217) — are not possible. The mental, in this case, is essentially the intentional: paradigmatic mental states, for Davidson, include beliefs, intentions, and desires.[11]

Much of the work of "Mental Events" goes into combatting reductionism concerning the mental, and Davidson's rejection of psychophysical laws is part of this effort. As we will see, Davidson argues that nomological reduction of the mental to the physical would result in the subsumption of mental concepts under the category of physical concepts; mental concepts would be, in essence, physical concepts, and would share the conditions of attribution of the latter. It is precisely this prospect that leads him to reject the possibility of a nomological reduction of the mental to the physical. Such attempts at reduction, he argues, succeed only in changing the subject, with the result that we are no longer talking about mental concepts, but about physical ones:

[10] In this section I am relying, in part, on discussions of Davidson's rejection of psychophysical laws in McLaughlin and Kim.

[11] EAE, 210, 211.

> [T]o allow the possibility of such laws [linking the mental and the physical] would amount to changing the subject. By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of the propositional attitudes. [EAE, 216]
>
> There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. [EAE, 222]
>
> [T]here cannot be tight connections between the realms if each is to retain its allegiance to its proper source of evidence. [EAE, 222]

Three phrases from these passages stand out and require further explanation: 'changing the subject', 'disparate commitments', and 'proper source of evidence'. I will begin with the latter two phrases; understanding those will shed light on Davidson's charge that nomological reduction succeeds only in changing the subject.

Above I mentioned that nomological reduction of the mental to the physical would result in mental concepts sharing the conditions of attribution of physical concepts.[12] 'Condition of attribution', in this context, is simply short for 'condition for the proper attribution of a concept'. Davidson's claim that mental and physical concepts have 'disparate commitments' is the claim that these two conceptual realms are characterized by different conditions of attribution. Of course, this by itself does not distinguish the difference between mental and physical concepts from any other conceptual difference: it seems likely that any two genuinely distinct concepts will have different conditions of attribution. So Davidson's claim must be stronger than the mere assertion that mental and physical concepts differ in their conditions of attribution: they must differ not merely in specific conditions, but in the type of conditions

[12] Like much of my discussion in this section, the notion of a condition of attribution I am borrowing from Kim's "Psychophysical Laws".

governing their attribution, and they must differ in such a way that to mistake the conditions of attribution of one for those of the other is to misconstrue the very nature of the two realms.

Consider the following passage:

> [T]he whole set of axioms, laws, or postulates for the measurement of length is partly constitutive of the idea of a system of macroscopic, rigid, physical objects. I suggest that the existence of lawlike statements in physical science depends upon the existence of constitutive (or synthetic a priori) laws like those of the measurement of length within the same conceptual domain. [EAE, 221]

One condition essential to the proper use of our concepts involving length is a law of transitivity: if x is longer than y, and y is longer than z, then x is longer than z. As Davidson says, without such a principle "we cannot easily make sense of the concept of length" (EAE, 220). In the first two sentences of the passage above Davidson seems to be making one or both of two points concerning the relevance of this reflection regarding the concept of length to our physical concepts as a whole. The first sentence can be read as suggesting that our concepts related to length and physical measurement are essential components of our physical conceptual scheme as a whole[13], and therefore that the conditions governing such concepts are essential to the physical (or at least to parts of it) as well. The second sentence may suggest a somewhat weaker

[13] I realize it sounds odd describing Davidson's views in terms of conceptual schemes, but there is some reason to think that this is not inconsistent with his attack on the notion of a conceptual scheme in "On the Very Idea of a Conceptual Scheme". For Davidson's attempt to reconcile his commitment to incommensurable physical and mental conceptual schemes with his rejection of conceptual relativism, see his "Reply to Solomon" (EAE, 243-4).

point: just as there are conditions and principles essential to our concept of length, so there are also analogous conditions and principles essential to our physical concepts generally. But however we read the first two sentences of this passage, we arrive at the same destination: the view that there are synthetic a priori principles concerning the physical, and that these principles represent conditions that are constitutive of the physical conceptual scheme in the sense that they play an essential role in determining what counts as physical. Where these conditions do not obtain, there we are not concerned with the physical, and concepts that do not presuppose these conditions are not physical concepts. These conditions or principles are conditions of attribution of physical concepts in the sense that the proper attribution of such concepts, and indeed, the very idea that a concept we are using is genuinely physical, presuppose these conditions.

But in the sentence immediately following the passage quoted above, Davidson implies that there must also be laws constitutive of the mental: "Just as we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort, we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions" (EAE, 221). The physical theory embodied in our practices of measurement includes descriptive laws like the law of transitivity concerning length. But the theory comprising our mental concepts must be very different. First, its 'laws' cannot support counterfactual and subjunctive claims, if Davidson is to preserve his thesis of the anomalism of the mental.[14] Second, the laws characteristic of the mental are not

descriptive, but normative: they do not tell us what beliefs or intentions a person will have, but what beliefs and intentions they ought to have.[15]

The constitutive principles governing mental concepts are norms of rationality — "in inferring this system [of an agent's intentional states] from the evidence, we necessarily impose conditions of coherence, rationality, and consistency" — and it is their normativity that prevents the nomological reduction of mental concepts to physical: "These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena" (EAE, 231). Davidson makes this point concerning the importance of the normativity of mental concepts more clearly in later writing:

> Perhaps it is obvious that definitional reduction [of mental concepts to physical] is out of the question; but why can't there be laws — strict laws — that connect each mental event or state with events or states described in the vocabulary of an advanced physics? When writing about this twenty years ago I said, in effect, that one can hope for strict connecting laws only when the concepts connected by the laws are based on criteria of the same sort, and so a strict law could not combine normative with non-normative concepts. This answer still seems to me right as far as it goes. . . . [SIO, 170][16]

[14] This is the thesis that "there are no strict deterministic laws on the basis of which mental events can be predicted and explained" (EAE, 208). It is worth noting that the synthetic a priori principles or 'laws' to which Davidson refers as constitutive of the physical and mental realms do not seem to be laws in the same sense as the deterministic psychophysical laws whose possibility he rejects.

[15] So for Davidson, as for Frege, the principles constitutive of thought are also prescriptive. Here we see Davidson's commitment to both the normative and dual role theses.

[16] In the remainder of this sentence, not quoted here, Davidson acknowledges that although he stands by his earlier explanation of the irreducibility of mental to physical concepts, that explanation was not sufficient to convince his critics, and requires some elaboration and supplementation.

Intentional or mental concepts are normative, while physical concepts are non-normative.[17] What this claim amounts to is the idea that the conditions of attribution or constitutive principles of mental concepts include normative principles, while the conditions of attribution of physical concepts do not. This is the difference in conditions of attribution between the mental and the physical that is supposed to block the reduction of the former to the latter.

Returning to the three phrases I set out to explain, how do Davidson's claims concerning the constitutive principles of mental and physical concepts contribute to the conclusion that a nomological reduction of the mental to the physical would simply 'change the subject'? The existence of psychophysical laws would entail a reduction of the mental to the physical that would result in the possibility of using criteria appropriate to physical concepts in the application of mental concepts. If, for example, it is a matter of law that some mental state M is coextensive with some neurophysiological state N, such that M occurs if and only if N occurs, then whatever conditions of attribution the concept of N has will also be conditions of attribution of the concept of M. As Kim puts it, laws of the sort that Davidson rejects would "underwrite certain inferences that mere de facto generalizations cannot sanction"

[17] Whether these traits are necessary features of the respective sorts of concepts is arguable. In "Praise the Lord! Ye heavens adore Him", a hymn appearing in the late 18th century and based on Psalm 148, the heavens are described as governed by "Laws which never shall be broken, for their guidance He hath made." The suggestion here, perhaps, is that the laws of nature that God has fashioned for heavenly bodies are in fact normative, but such bodies would never will to disobey them. See The Church Hymnal Corporation 1985, Hymn 373.

(Kim, 374)[18], allowing the transmission of conditions of attribution from one conceptual 'scheme' to the other. The existence of psychophysical laws, then, would entail the possibility of applying mental concepts using criteria at odds with their "proper source of evidence." Our evidence for the application of mental concepts, according to Davidson, includes considerations about what an agent ought to believe, desire, or intend. There is, however, no such normative element in our application of physical concepts, so the existence of psychophysical laws would mean the possibility of applying mental concepts completely independently of normative considerations. But the normative criteria pertaining to mental concepts are not, on Davidson's view, optional, so long as we want to treat of the mental; if, as he maintains, these criteria are constitutive of mental concepts, then we cannot employ physical criteria in the application of mental concepts without 'changing the subject', without ceasing to speak of the mental.

3.2.2 Conceptual relativism and the methodology of interpretation

Davidson's claims concerning the 'disparate commitments' of the mental and physical schemes and the sources of evidence proper to each point to the idea that there are normative principles constitutive of the mental. But the constitutive thesis is an essential component not just of Davidson's rejection of psychophysical laws, but also of his arguments concerning conceptual relativism. These arguments seek to

[18] This, on Kim's view, explains why Davidson rejects the possibility of psychophysical laws but accepts the possibility of true generalizations between the mental and physical.

undermine the idea that there could be languages with concepts so radically different from our own as to be wholly untranslatable into our language.[19]

At the heart of Davidson's treatment of conceptual relativism is the following claim: "Given the underlying methodology of interpretation, we could not be in a position to judge that others had concepts or beliefs radically different from our own" (ITI, 197). This claim concerning the methodology of interpretation is best understood in terms of the central role Davidson assigns to the principle of charity in interpretation. This principle counsels us to interpret others in such a way that they turn out to be largely rational and to share our beliefs concerning at least the obvious features of our common environment. Davidson identifies two particular principles that represent different aspects of the principle of charity: the principle of coherence, which requires "a degree of logical consistency in the thought of the speaker" (SIO, 164), and the principle of correspondence, which "prompts the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances" (ibid.). These "two key principles. . . must be applicable if a speaker is interpretable" (ibid.).

This last quotation reflects Davidson's claim that "charity is not an option, but a condition of having a workable [interpretive] theory" (ITI, 197). The necessity of

[19] For passages supporting the idea that Davidson's arguments against conceptual relativism are aimed at the notion of languages untranslatable as a whole, see his "Reply to Solomon" (EAE, 243-4): "It makes sense to speak of irreducible or semi-autonomous systems of concepts, or schemes of description and explanation, but only as these are less than the whole of what is available for understanding and communication."

charity in interpretation is due in part to the starting point Davidson chooses in trying to make sense of linguistic communication: the situation of a radical interpreter, who knows nothing in advance about the beliefs or language of those he is to interpret.[20] Any interpreter in such a situation must make certain assumptions about the beliefs, interests, and other intentional states of his subjects in order get his interpretive project off the ground, and Davidson's claim is that the principle of charity embodies many of these assumptions. But the necessity of charity is not solely a matter of the demands of radical interpretation, and this is particularly clear in the case of what Davidson describes as the principle of coherence. The requirement that we find "a degree of logical consistency in the thought of the speaker" reflects the very nature of the concepts we use in interpretation: concepts of belief, intention, desire, and so on, the application of which is governed by normative criteria. On Davidson's view, beliefs cannot be ascribed singly solely on the basis of descriptive criteria; rather, "beliefs are identified and described only within a dense pattern of beliefs" (ITI, 200).[21] And not

[20] I do not mean to suggest that this starting point is arbitrary. Davidson's view seems to be that the situation of the radical interpreter is in some way characteristic of all attempts at interpretation: "[a]ll understanding of the speech of another involves radical interpretation" (ITI, 125).

[21] As an example of what I mean by 'descriptive criteria' here, consider a person's utterances. Often utterances take the form of statements of belief, but Davidson's point, I take it, is that the utterance by itself—the descriptive criterion in this case—is not decisive when it comes to the ascription of beliefs to a speaker based on her utterances. Due to the possibility of misstatements and malapropisms, as well as considerations of coherence with other beliefs held by the speaker, the beliefs we ascribe often diverge from those we would ascribe if we relied solely on the apparent meaning of the speaker's utterances.

just any pattern will do; it must be "the right sort of pattern of beliefs to support that one belief" (ibid.). Spelling out what this notion of the 'right sort of pattern' means for Davidson requires appealing to the constitutive role of norms of rationality in content ascription. Davidson speaks, in this regard, of a "constitutive ideal of rationality" (EAE, 223) that governs interpretation. It is constitutive in the sense that interpretation, or the ascription of intentional content, cannot proceed independently of this ideal; it is an ideal in the sense that it is normative: it specifies how one ought to think. So the 'right sort of pattern' to support a given belief will be a pattern of beliefs one ought to have, logically and epistemically speaking, to support the belief: minimally, and in keeping with the principle of coherence, this will be a pattern that has a sufficient degree of logical consistency with that belief.

Even if my explanation leaves many of the details of Davidson's argument against conceptual relativism murky, it should be clear, at least in broad outline, how the principles of coherence and correspondence are supposed to yield an argument against relativism: given that these principles are constitutive of intentionality and essential to interpretation, the idea that we might encounter beings with radically different concepts or forms of thought is incoherent. What motivates this argument, however, is the same view underlying Davidson's rejection of psychophysical laws — the view that there are logical and epistemic principles constitutive of mental concepts. This view concerning mental concepts is fundamental to Davidson's claims concerning the relevance of the methodology of interpretation to conceptual relativism, as well as to his treatment of psychophysical laws. But it is clear, on Davidson's view, that this view concerns not just our mental concepts and methods of interpretation, but also

mental states themselves: "An aura of rationality, of fitting into a rational pattern, is. . . inseparable from these phenomena [e.g., beliefs, desires, intentions]" (POI, 289). The normative criteria essential to mental concepts and the inseparability of an aura of rationality from psychological states are two aspects of the same phenomenon: it is part of the nature and identity of a psychological state that it belongs to a rational pattern of such states, and mental concepts and the methodology of interpretation reflect this.

3.3 Conclusion

I have argued that Frege and Davidson should both be understood as proponents of what I have called the constitutive thesis. It should also be clear, from what we have seen, that the principles they identify as constitutive of thought they also regard as normative for thought. We can conclude, then, that they are both committed to what I have called the dual role thesis — the view that certain principles are both constitutive of thought and normative for it. I have argued that Frege is, in addition, committed to the categorical conception. It is perhaps no accident that Frege holds the two views that I have ascribed to him: the dual role thesis is incompatible with the hypothetical conception, and so commitment to that thesis pushes one in the direction of a categorical account of cognitive normativity. The hypothetical conception, as I have defined it, is the view that all cognitive normativity is hypothetical. To endorse that conception, then, is to endorse the view that our thought need not observe any particular cognitive norms: because conformity to such norms is extrinsic to the nature of thought — these are principles to which one is responsible not *qua* thinker, but merely *qua* thinker committed, say, to truth — failure to think in accordance with them

does not jeopardize one's status as a thinker; it merely calls into question one's commitment to the goal of truth. As Papineau puts it, "sensitivity to norms of judgement is an addendum to the possession of beliefs itself" (NJ, 32). The view he expresses here is in direct contrast with Frege's claim that logical laws, as cognitive norms, "prescribe universally the way in which one ought to think if one is to think at all" (BLA, 12). If we take that claim as a statement of the dual role thesis, it represents a denial of Papineau's view: it's not the case, according to that thesis, that all norms are extrinsic to the nature of thought. On the contrary, there are some norms that we must heed if we are to think at all. Such norms will be categorical, possessing an authority that is not contingent upon one's desires, but merely on one's status as a thinker.

It is hard to discern in Davidson's writing any clear account of the authority of the "constitutive ideal of rationality" to which he refers, but if my argument in the preceding paragraph is correct, then we can identify the account he *should* embrace: the categorical conception. Assuming the constitutive thesis is compatible with any form of cognitive normativity — assuming, in other words, that one can consistently regard one and the same principle as playing both constitutive and normative roles in thought — then its incorporation into the dual role thesis seems to entail the categorical conception: any norm constitutive of thought will, it seems, gain its authority simply from one's status as a thinker, not from one's having particular desires.

As I stated at the outset of this dissertation, one of my aims here is to explore the idea that the dual role thesis might meet the demand made by naturalist opponents of the categorical conception like Papineau — the demand for some account of the authority of categorical cognitive norms. The connection I've just sketched between

that thesis and the categorical conception is encouraging, for it suggests that there is a natural affinity between these two views. There is also, however, a problem with the dual role thesis that threatens any prospect for using it to illuminate the categorical conception. This problem is embodied in the assumption that I made in the preceding paragraph — the assumption that one can consistently regard one and the same principle as playing both constitutive and normative roles in thought. Only if this assumption is true — that is, only if the dual role thesis is coherent — can we explain the normative authority of a cognitive principle in terms of its constitutive role in thought, but there are reasons for thinking it is false. The question I will address in the next chapter, then, is whether a principle that is constitutive of thought can be, at the same time, normative for thought. In addressing this question I hope also to clarify what it might mean for a principle to be constitutive of thought, and what versions of that view we can regard as plausible.

Chapter 4: The Coherence of the Dual Role Thesis

The last chapter ended with the following question: can a principle that is constitutive of thought play, at the same time, a normative role for thought? This is essentially the question of whether the dual role thesis is coherent, and how we answer this question depends to a large extent on the sense in which the principle in question is supposed to be constitutive. What I hope to show in this chapter is that while there are significant problems with the way Frege and Davidson sometimes articulate the dual role thesis — problems that might initially seem to undermine any plausibility it has — there is nevertheless a sense in which a principle can simultaneously play both normative and constitutive roles in thought. Dispelling the air of incoherence surrounding the dual role thesis clears the way, I will argue, for an account of how certain cognitive norms might be categorical, and I will begin sketching that account in this chapter.

I should emphasize that my intention in this chapter is not interpretive: I do not intend to try to resolve the problems apparent in the dual role thesis through further analysis of the views defended by Frege or Davidson. While there may be ways of resolving these problems from within their texts, it seems more likely to me that there are not: as far as I can tell, there is no evidence within those texts of an awareness of the questions I am raising concerning the dual role thesis, never mind any effort to resolve them. My concern here is to solve a puzzle that is raised by these texts, but not addressed by their authors.

4.1 A strong version of the constitutive thesis and its difficulties

There are actually two potential problems with the views I attributed to Frege and Davidson in the last chapter. Both problems arise out of a particular understanding of what it means for a principle to be constitutive of thought: on this understanding, if a principle is constitutive there can be no thought that does not conform to it; failure of conformity is simply a failure to think. This strong version of the constitutive thesis is evident, to varying degrees, in both Frege and Davidson. Here are two representative passages:

> For purposes of conceptual thought we can always assume the contrary of some one or other of the geometrical axioms, without involving ourselves in any self-contradictions when we proceed to our deductions, despite the conflict between our assumptions and our intuition. The fact that this is possible shows that the axioms of geometry are independent of one another and of the primitive laws of logic, and consequently are synthetic. Can the same be said of the fundamental propositions of the science of number? Here, we have only to try denying any one of them, and complete confusion ensues. Even to think at all seems no longer possible. The basis of arithmetic lies deeper, it seems, than that of any of the empirical sciences, and even than that of geometry. The truths of arithmetic govern all that is numerable. This is the widest domain of all; for to it belongs not only the actual, not only the intuitable, but everything thinkable. Should not the laws of number, then, be connected very intimately with the laws of thought? [FA, §14]

> An aura of rationality, of fitting into a rational pattern, is thus inseparable from these phenomena [e.g., beliefs, desires, and other intentional states]. [POI, 289]

Frege's analogy between the role of geometrical axioms in intuition and the role of logic in conceptual thought might seem to suggest the following: just as we cannot intuit except in accordance with the axioms of Euclidean geometry, we cannot think except in accordance with the laws of logic. Thought that does not reflect these truths is no thought at all; it is merely a semblance of thought. And Davidson's claim that an

"aura of rationality" is inseparable from intentional states — in other words, that it is essential to such states that they form a pattern structured by principles of rationality — can easily be read as suggesting that there are no intentional states that do not conform to such principles; hence, there is no thought that does not so conform either.

4.1.1 Is human thought really logical?

What is immediately striking and most obviously problematic about this strong version of the constitutive thesis is the highly intellectualized picture of human thought that it paints. If thought that does not conform to logical laws or principles of rationality is merely a semblance of thought, then it would seem that much of what passes for thought in human life doesn't really merit the description: we don't need studies demonstrating that human thought is prone to basic logical errors to recognize the obvious fact that our reasoning is, from a logical point of view, fallible. If one were feeling uncharitable toward the human race, one could of course respond that this is just so much the worse for what passes for human thought: perhaps what we often take for thought truly doesn't merit the description. But if our aim is to make sense of a characteristically human activity — an activity that constitutes who we are as supposedly rational animals — it hardly seems plausible to set the bar for thought so high that little of what we do actually counts. In any case, we typically distinguish between thought that is better and worse, more or less careful, and it's not clear how such distinctions can be sustained if thought, by definition, conforms with principles of logic or rationality. It's not clear, for example, how we can reconcile this strong version of the constitutive thesis with Frege's acknowledgment that "[t]hinking, as it actually

takes place, is not always in agreement with the laws of logic any more than men's actual behaviour is in agreement with the moral law" (PW, 145).

4.1.2 Is the constitutive thesis consistent with the normative thesis?

The foregoing problem concerns only the constitutive thesis, but a further problem arises when that thesis is conjoined with the normative thesis to form what I've called the dual role thesis. Recall that the latter view maintains that certain principles play a dual role in thought: they are both constitutive and prescriptive of thought. Now consider the following principle, what I will call the "ought implies can not" principle (hereafter, OICN):

> The normativity of a principle, rule, norm, etc., depends on the possibility of our failing to conform to it: in order for it to be true that one ought to Φ, it must be the case that one could fail to Φ.

This principle articulates a constraint on normativity that has, I think, at least some intuitive appeal. Norms, standards, principles, and rules serve to correct, or bring into line, action that has gone astray, or to forestall one's going astray in the first place. So the possibility of straying, of acting in a way one ought not, seems integral to their character. Kant notes this feature of norms in discussing the notion of an imperative: "All imperatives are expressed by an ought and indicate by this the relation of an objective law of reason to a will that by its subjective constitution is not necessarily determined by it" (Kant 1997, 4:413). It is for this reason that a holy will, according to Kant, has no obligations: its maxims "necessarily harmonize with the laws of

autonomy" (ibid., 4:439).

Nothing I've said constitutes an argument for OICN, and I'm not sure that I could supply such an argument. But it's not essential to what I have to say in this section that I demonstrate the general plausibility of the principle; all that is necessary is that it have some plausibility in the particular case I'm considering. Here, in brief outline, is that case. If we accept OICN, then if a principle is to be normative for thought, it must be the case that one could fail to think in accordance with it. But if the same principle is also constitutive of thought, one cannot fail to think in accordance with it, for a failure to think in accordance with it is simply a failure to think. So it would seem that no principle could, at the same time, be both constitutive of thought and normative for thought. The dual role thesis is therefore incoherent.

Davidson's talk of a "constitutive ideal of rationality" and Frege's claim that logical laws prescribe how "one ought to think if one is to think at all" seem to raise precisely the problem I've identified with the dual role thesis. Both passages express commitment to the idea that there are certain norms with which our thought must be in accordance if we are to count as thinkers at all. These cognitive norms are constitutive of thought. Now presumably one is subject to cognitive norms only if one has certain requisite characteristics — chief among them, the characteristics that make one a thinker. And to regard certain principles as constitutive of thought is to regard their embodiment in one's thought as among those characteristics essential to one's identity as a thinker. But these considerations suggest that such constitutive principles, far from being cognitive norms themselves, instead specify characteristics one must have *if* one is to be subject to cognitive norms. There is, then, a very specific sense in

which one could not fail to think in accordance with such principles: one's identity as a thinker is linked to these principles in such a way that a failure to conform to them deprives one of that identity, so one cannot fail to think in accordance with them while remaining subject to cognitive norms.

It might be argued, however, that even granting OICN, the specific sense I have identified in which one could not fail to think in accordance with principles constitutive of thought is not sufficient to show that such principles cannot be normative. Consider the following analogy. Suppose that one is a member of an organization, and that one's membership in the organization is contingent upon one's engaging in some activity; this activity is constitutive of one's membership in the organization. Despite the constitutive nature of this activity for one's membership, might one not correctly regard the injunction to perform this activity as having a kind of normative authority over one's behavior, inasmuch as one wants to remain a member of the organization? And similarly, even if certain principles of thought are constitutive for one's identity as a thinker, might one not rightly regard those principles as possessing a kind of normative authority over one's thought, inasmuch as one wants to remain a thinker? What this analogy suggests is that one can fail to think in accordance with principles constitutive of thought in a way that is consistent with OICN.

Despite the apparent similarities between the two cases, however, there is a crucial dissimilarity: an organization member who fails to act so as to maintain her membership does not thereby cease to act. So the rules governing her membership truly are rules prescribing how she ought to act; she can follow them or not, depending

on whether she wants to maintain her membership. But do principles constitutive of thought tell one how one ought to think? The implication behind the idea that they do is that one could think differently. Just as one could act in violation of the organization rules, one could think in violation of principles constitutive of thought. But one cannot do the latter: the idea that these principles are constitutive, in the sense under consideration here, is that thought that fails to accord with them is not genuine thought. As constitutive principles, they tell one how one must think, where the 'must' is not normative but descriptive. One might regard it as important *that* one continue to think, and for that reason one might take very seriously considerations concerning what is essential to thought, but there can be no latitude, with respect to constitutive principles, *how* one ought to think.

4.2 A moderate version of the constitutive thesis

I explained at the end of the last chapter that my interest in the dual role thesis lies, at least in part, with its prospects for illuminating the categorical conception in a way that is responsive to Papineau's demand "that we should have some kind of understanding of the peculiar force that [sui generis, categorical] judgemental norms are supposed to exert on us" (NJ, 27-8). But in light of the problems I've introduced above, these prospects may seem quite dim. The idea that the dual role thesis can elucidate the categorical nature of cognitive normativity depends on the assumption that it is possible for a principle to be both constitutive and normative, and it is not clear whether this assumption is correct. Moreoever, even if it were correct, there is the further question of whether the constitutive thesis, an essential element of the dual role

thesis, depicts human thought in a manner so overly intellectualized that it has little bearing on its actual nature: in this case, even if the dual role thesis could shed light on categorical cognitive normativity, one might reasonably wonder what relevance this has for us, given that the constitutive thesis seems to describe idealized rational beings, rather than baser creatures like ourselves.

What I want to emphasize is that both of these problems depend on a strong reading of the constitutive thesis according to which principles constitutive of thought are ones to which thought must, in a descriptive sense, conform. The key to seeing our way around the problems I've sketched is formulating a more moderate version of the constitutive thesis — a version preserving the idea that these principles play a constitutive role in thought in some sense while leaving room for a normative role as well. It is to this task that I now turn.

4.2.1 Normative concepts

As a preliminary step in the articulation of a more moderate version of the constitutive thesis, I want to look more closely at a claim of Davidson's that we saw earlier:

> Perhaps it is obvious that definitional reduction [of mental concepts to physical] is out of the question; but why can't there be laws — strict laws — that connect each mental event or state with events or states described in the vocabulary of an advanced physics? When writing about this twenty years ago I said, in effect, that one can hope for strict connecting laws only when the concepts connected by the laws are based on criteria of the same sort, and so a strict law could not combine normative with with non-normative concepts. This answer still seems to me right as far as it goes. . . . [SIO, 170]

What I want to focus on in this passage is the distinction between 'normative' and

'non-normative' concepts. It's not entirely clear what this distinction comes to, or what it means for a concept to fall into one category or the other. In discussing this passage in Chapter Three, I said that normative concepts are those for which the conditions of attribution include normative principles, while non-normative concepts are those whose conditions of attribution do not include such principles. But this isn't much of an improvement: it's not clear what it means for a concept to have normative principles as conditions of attribution. If we think of 'conditions of attribution' as ways of specifying those characteristics something must have in order to fall under a particular concept — Frege's *Merkmale*, or 'characteristic marks' [1] — the idea that there might be normative principles among these conditions is obscure: what would it mean for an object to have a normative principle as a characteristic? Perhaps one could explain this idea in terms of the constitutive thesis: to fall under a normative concept might involve embodying the principle, or conforming to it, in one's behavior. But with this suggestion we return to the tension between the constitutive and normative theses that I described above: if the principle in question simply describes the behavior of what falls under the supposedly normative concept, it is no longer clear why that principle should be characterized as normative in the first place.

Despite its initial opacity, the notion of a normative concept can, I believe, be explicated in a way that is instrumental to the formulation of a more moderate version of the constitutive thesis. It is through an understanding of the concept of a thinker or rational being as normative that we can render coherent the dual role thesis that I have

[1] See FA, §53.

ascribed to Frege and Davidson. So in the next few pages I want to discuss a range of examples illustrating what Davidson may have in mind in speaking of normative concepts. After presenting the examples, I will try to draw some general conclusions about how best to understand such concepts.

1. Functional concepts

R. M. Hare discusses the notion of a functional word in *The Language of Morals*: "A word is a functional word if, in order to explain its meaning fully, we have to say what the object it refers to is for, or what it is supposed to do" (Hare, 100). Similarly, a functional concept is one that specifies a category of things by reference to some function they are supposed to perform. Hare points to 'auger' as an example of a functional word; the concept of an auger, then, is a functional concept. We haven't fully explained what an auger is simply by citing its characteristic physical features: a "long pointed shank with a cutting edge and a screw point, and a handle fixed at right angles to the top of the shank."[2] To explain the concept of an auger — to explain what an auger is — we must also explain its function, what it is for: namely, boring holes in wood.

One might even argue that I have misdescribed matters above: it is not that we haven't fully explained what augers are without mentioning their function; we haven't explained what augers are at all if we've simply cited their typical physical features. In many cases, perhaps, functional concepts pick things out as belonging to a particular type not in terms of physical characteristics, but solely in terms of function. A toddler's

[2] *Oxford English Dictionary*, 2nd edition.

toolset, containing soft plastic replicas of common toys, could very well contain something resembling an auger in many physical respects, but I think we would say that what we have in that toolset is not a genuine auger, but simply a toy. While it has the physical form of an auger, the material out of which it is constructed guarantees that it will not do what an auger is supposed to do. Having certain physical characteristics, then, is not sufficient for something to count as an auger. One might object, of course, that the substance out of which an auger is constructed cannot be left out of an account of the concept of such a tool: the child's toy auger is not a genuine auger precisely because it is made out of the wrong sort of stuff. But any attempt to specify the 'right sort of stuff' out of which real augers are made will inevitably appeal to what an auger is supposed to do: the 'right sort of stuff' is simply whatever enables the auger to do what it is supposed to do. So it is not clear that any list of physical characteristics can adequately specify a functional concept without covertly appealing to the function in question.

It is interesting to note, however, that although there seems to be a sense in which an item falling under these functional concepts must be capable of performing the relevant function, there is another sense in which this is not the case. Although the function of a watch is to tell time, performing this function well, or even at all, is not essential to the identity of a thing as a watch. And a broken auger is nevertheless an auger, even though it is no more able to perform its function than the plastic toy auger in a child's toolset. In these two examples it may be relevant that we're dealing with artefactual concepts: perhaps it is sufficient for something's being a watch that it was made for telling time, even if it is subsequently unable to serve that purpose. But

assuming this is a feature of concepts that are both functional and artefactual, the complex relationship to function that is characteristic of such concepts is not confined to them: a heart that ceases to pump blood remains a heart, even though it no longer serves its characteristic function within the organism.

2. *Explanation*

Now consider the concept of an explanation. Here we have what seems to be a functional concept: an explanation is "[t]hat which explains, makes clear, or accounts for."[3] And it's certainly true that an explanation needn't do its job well in order to count as an explanation: poor explanations can be explanations. But in contrast with the functional concepts discussed above, the concept of explanation does not seem to display the same tolerance with respect to functional failure: while a watch that fails to tell time may remain a watch, in at least some circumstances what appears to be an explanation may turn out not to be so if it fails to explain anything. Putnam offers the following example:

> Professor X is found stark naked in the girls' dormitory at 12 midnight. Explanation: (?) He was stark naked in the girls' dormitory at midnight - ϵ [where ϵ is some small interval of time] and he could neither leave the dormitory nor put on his clothes by midnight without exceeding the speed of light. But (covering law:) nothing (no professor, anyhow) can travel faster than light. [Putnam 1978, 42]

Putnam is unsure as to whether this is a 'terrible' explanation or no explanation at all, but claims that the latter interpretation embodies his "sense of both how the language

[3] Ibid.

works and how it should work" (ibid.). The line separating bad explanations from nonexplanations may be fuzzy, and it may be that Putnam's example doesn't fall clearly into one category or the other. But perhaps clearer examples are available of ostensible explanations whose status as explanations dissolves on further inspection. A young child, asked to explain why he hit his friend on the playground, offers the following response: "Because I did." One might say that the child has some grasp of the form of explanation-giving, but hasn't quite grasped what an explanation is supposed to do. I think we would be justified in such cases in saying that no explanation has been given. Or consider circular explanations generally. In at least some situations where we confront a set of statements purporting to explain something by appealing (overtly or not) to the very thing to be explained, it is reasonable to say that what we confront is not, in fact, an explanation. It appears, then, that there is some reason to think that it is essential to something's being an explanation that it perform, to at least some minimal degree, its function. Actual performance of function seems to be essential to the concept of explanation in a way that it is not to concepts like that of an auger or heart.

3. Father

What I have in mind here is not the concept of a biological father, but the concept involved in statements like the following: "He was a father to me as I was growing up, in a way my biological father wasn't." Whether one is a father in this sense is independent of biological relationship, and depends instead on one's attentiveness, the care and concern one demonstrates toward a child, the weight one gives one's parental

responsibilities in comparison to other commitments, and, in general, the ethical character of one's relationship to a child. It's unclear whether this is a functional concept, partly because the 'functions' of a father are so diffuse that it is difficult to specify what such a function might be. We might attempt to corral these functions under the rubric, "to raise a child properly", but is it essential that one raise a child in order to be a father to it in this sense? What seems clearer is that this concept categorizes persons in terms of a role they might serve or fail to serve — a role that might also be served in a more or less commendable fashion. So one can be a better or worse father in this sense, but sufficiently gross failure to adhere to the norms and obligations associated with this role can disqualify one from being a father entirely.

4. 'Life-form' concepts

Finally, I want to discuss briefly a class of concepts identified by Michael Thompson as 'life-form' concepts.[4] Thompson focuses on what he terms 'natural-historical judgments', the often tenseless judgments we make in characterizing life-forms or species. We say, for example, that "Dogs have four legs", or "Kokanee salmon mate in the fall". Such judgments, he claims, resist the logical and epistemological categories into which we might be tempted to put them. Despite their initial appearance, they cannot be assimilated to universal generalizations like "All dogs are animals with four legs" or "For all x, if x is a dog then x has four legs". Confronted with a three-legged dog, we are not inclined to give up the claim that dogs have four legs, but simply to conclude that the dog in front of us suffers from some kind of defect. Nor are natural-

[4] See, for example, Thompson 1995, pp. 280-91.

historical judgments statistical generalizations: what is statistically true varies with changes in the population over which we are generalizing, but even if it turned out that all dogs born between 2010 and 2030 had three legs as a result of a chemical spill with global environmental consequences, we would not say during that time that dogs have three legs.[5] And for reasons related to those I've just cited, natural-historical judgments are not obviously empirical, either: they are not responsive to empirical observation of dogs in the way one would expect of a straightforwardly empirical claim.[6]

Thompson's claim is that natural-historical judgments are judgments concerning species or life-forms, rather than the individuals belonging to those species or life-forms. Such judgments, and the sentences expressing them, which he terms "Aristotelian categoricals", embody "one's interpretation or understanding of the life-form shared by the members of that class" (Thompson 1995, 288). To put the point another way, these are judgments delineating the character of species or life-form concepts — concepts embodying this interpretation or understanding. Thompson defines a life-form or species concept as "a possible [logical] subject of the corresponding [natural-historical] form of judgment" (ibid., 292).

[5] Thompson also argues that such judgments are not properly understood as involving implicit *ceteris paribus* clauses, but his argument is too complex to reproduce here. See Thompson 1995, pp. 285-6.

[6] This is not to say that natural-historical judgments are not empirical, but simply that their relationship to observation is clearly much more complex than that of sentences like "The cat is on the mat" or "All of the students in the classroom are wearing shirts." Here I may diverge from Thompson; if I read him correctly, he strongly suggests in Thompson 2004 (pp. 57-8) that he regards natural-historical judgments as a priori.

Life-form concepts are not functional concepts, although presumably many natural-historical judgments identify teleological or functional traits of the individuals belonging to the species they describe. (Thompson's example: "They have blossoms of such-and-such type in order that such-and-such insects should be attracted and spread their pollen about" (ibid., 293).) Like the previous examples, however, life-form concepts allow us to make various sorts of evaluative distinctions among members of the species, including judgments of defect and soundness.

"Augers bore holes in wood." "Hearts pump blood." "Dogs have four legs." These sorts of judgments are essential to our concepts of such things, but, as functional and life-form concepts illustrate, that is not to say that for all x, if x is an [auger/heart/dog] then x [bores holes in wood/pumps blood/has four legs]. Given a particular tool that fails to bore holes in wood, we are licensed (depending on other features of the situation) to conclude either that it is not an auger or that it is a defective auger. But given a sample of liquid that turns out not to contain two hydrogen atoms for every atom of oxygen, we are not similarly licensed to conclude that what we confront might be a defective example of the chemical compound, water; the sample either contains water or not, and this one does not. What distinguishes the former from the latter is that the concept of an auger, along with that of a heart or a dog, involves a reference to a sort of standard, form, or ideal. A functional concept, for example, specifies what an object falling under it ought to do, not what it is in fact able to do. Thompson's reference to species concepts as 'life-form' concepts may also be relevant here: such concepts specify a form that particular individuals may embody to a greater or lesser

degree. And certainly the concept of a father involves the specification of a standard against which an individual's conduct can be measured.

This ideality, this reference to an ideal, form, or standard, is essential to the notion of a normative concept. And through this ideality, normative concepts underwrite distinctions between function and dysfunction, soundness and defect, health and sickness, and goodness and badness, among others — distinctions that are central to the normative judgments we make concerning objects falling under these concepts. The four examples I presented above are all examples of normative concepts, and demonstrate something of the range of concepts that fall under this heading. In many cases — those of the functional concepts — this ideality serves as a way of identifying what members of the class are supposed to do, and thereby provides a basis for evaluation of objects falling under the concept. In other cases — consider the concept of a father or Thompson's life-form concepts — it provides a basis for evaluation without specifying any precise function. But in still other cases — the examples of explanations and fathers — the ideality serves not simply as a basis for evaluation, but also as a criterion for categorization of a sort that is absent in the other cases: in both cases candidacy for inclusion under the concept seemed to involve a certain minimal level of approximation toward the ideal or standard embodied in the concept.

What characterizes normative concepts, I am claiming, is this element of ideality, and it is this element that non-normative concepts lack. The mass of an object, in itself, can be neither good nor bad, neither sound nor defective. The concept of physical mass does not provide a basis for evaluative discriminations, although we may

of course bring with us to the investigation of an object's mass (our own, for example) interests or expectations that may or may not be fulfilled by what we discover. Similarly, as Stephen Toulmin notes, "Planets do not have 'good' or 'bad' orbits: they simply move as they move."[7] The notion of an orbit, or a satellite, does not open up a space for normative distinctions of the sort that we find with the concepts of a heart, dog, or auger.

So when the astronomer glimpses something that seems to be a satellite of a particular planet, she may be faced with the question of whether it is or is not a satellite, but not with the further question of whether it might be good or bad, sound or defective. But when the archaeologist uncovers an object that is auger-like in some respects yet non-functional, the alternatives are somewhat different: it may be an auger, albeit defective, or it may be something else entirely, such as a ceremonial instrument of some sort. And when the biologist encounters an organism that resembles members of species x in many respects but diverges in others, there is once again the question of whether this is a defective x or a non-defective member of some hitherto unknown species.

In the case of the archeaologist or the biologist, what justifies one response — it is a defective auger, a defective member of species x — or the other — it is not an auger, not a member of species x? There is probably no general answer we can give here that would cover all cases. With functional concepts that are also artefactual, like that of an auger, the intentions with which the object was constructed and used will, to

[7] Toulmin 1990, 149.

a large extent, determine which response we find correct. In the biological case, on the other hand, whether or not we find other organisms with similar divergences from the members of species x will be an important consideration in determining our response.[8] But even if there is nothing general to say about what justifies one response or the other in all cases, there is something to be said about the general pattern that we confront in many such cases. In some ways, Plato's talk of particular things 'striving' to be like the Form they instantiate provides an apt metaphor for understanding this pattern.[9] In the archeaological and biological cases, and in others like them, the question we confront is whether the object under consideration can reasonably be regarded as 'striving' toward a particular form. The notion of striving is a way of articulating what I have referred to as the 'ideality' of normative concepts: such concepts categorize particular objects by reference to a form, standard, or ideal, so the question in identifying an object as falling under a normative concept is not whether it meets the form, standard, or ideal in all respects, but whether it is best seen as 'striving' toward it. To appeal to a different metaphor, we can think of a normative concept as defining an axis along which objects falling under the concept can be placed and evaluated as approximating to a greater or lesser degree to a standard, form, or ideal. But if an object is to fall under a particular normative concept and to admit of such evaluation, it must be capable of being placed along the appropriate axis; in other words, it must be possible to see that object as evaluable in terms of the ideal, standard, or form articulated in the concept.

8 Important, but not decisive. I say something more about this issue in the next paragraph.

9 *Phaedo*, 74e.

The question the biologist confronts concerns the axis on which she ought to place and evaluate the creature she has discovered. Are there other creatures like it that diverge in a relatively uniform way from species x? That might suggest that there is some distinct life-form — a hitherto unknown species y — in terms of which we ought to evaluate this particular creature. But if, in addition, there are contaminants present in the immediate environment that can cause deformities similar to the respects in which the organisms differ from members of species x, the most reasonable conclusion might be that these are deformed x's. The question is whether we are to see these organisms as striving with relative success toward form y, or as striving in adverse and abnormal circumstances toward form x.

I now want to turn back to examples 2 and 3 above — the concepts of explanation and fatherhood. I claimed earlier that these two concepts differed from the other examples in a particular way: while a broken auger remains an auger, an explanation must actually do some explaining, and to merit the characterization of 'father' (in the relevant sense) a person must behave in at least a minimally father-like manner. Inclusion under the concepts of explanation and fatherhood seems to require a certain minimal level of approximation toward the ideal or standard embodied in each concept. The reason for this requirement, I think, is that these concepts characterize activities (in the case of explanation) or roles involving activities (in the case of fatherhood). Part of what it means to engage in this activity or to serve in this role — part of what is necessary to place something or someone on one of the axes defined by these concepts — is to display a sensitivity and responsiveness to the aims and norms internal to the activity or role. So when a three-year-old, asked to explain

why he hit another child on the playground, says he did it because he did, the reason we are inclined to say that no explanation has been offered at all is because his attempt at explanation displays an obliviousness to the interests and aims an explanation serves: an interest for greater insight into what led to the hitting, perhaps with an eye to preventing such behavior in the future. And someone who continually shirks his responsibilities to his child may be a biological father, but is not a father in the sense specified above. Failure to reach a minimal level of approximation to the ideals or standards embodied in these concepts is prima facie evidence of an absence of the sensitivity and responsiveness requisite for an act to be one of explanation, or for a person to be a father.[10]

[10] It might be argued that I'm identifying only one alternative here where there are at least two. Failure to attain a minimal level of approximation to the standards embodied in these concepts *might* be evidence of an absence of sensitivity and responsiveness to those concepts, but it might also be evidence, not of a lack of sensitivity, but simply of a lack of responsiveness. By ignoring the latter alternative, the objection continues, I'm stacking the deck in favor of the categorical conception, ruling out, in an ad hoc manner, the possibility that one might be a thinker, and recognize such norms, but just not care about following them. The problem with this objection is that it misses the various examples of normative concepts in which sensitivity and responsiveness are both essential to individuals falling under them. Consider, for example, the notion of a baseball player. It is not sufficient to be a player simply to recognize the norms governing the game; a player who makes no effort to achieve the aims internal to baseball would, at a certain point, cease to be a player of baseball. Similarly, it's not sufficient for one's activity to be one of explaining that one recognize but not heed the aims and interests constitutive of explanation; even if the child recognizes these aims and interests, his failure to act in a way that is informed by them means that he still hasn't offered an explanation.

Now let me turn back to Frege and Davidson. We can make sense of the dual role thesis to which they are committed if we regard the notion of a thinker as a normative concept analogous to the concepts of explanation and fatherhood. Consider, for example, what Davidson has in mind when he characterizes mental or intentional concepts as normative. On Davidson's view, our application to others of mental or intentional concepts — the concept of belief, for example — is inextricably bound up with certain norms and ideals. To see someone as having beliefs, as engaging in thought, as rational, is to see them as sensitive and responsive to certain norms. Sufficient insensitivity and lack of responsiveness to these norms jeopardize a characterization of them as rational beings or thinkers. So the concept of a thinker is similar to that of an explanation or fatherhood, in that it characterizes an activity (or rather, a role defined in terms of an activity) with certain norms internal to it. To engage in the activity — to be a thinker — is to display some minimal level of recognition, tacit or otherwise, of these norms, and to adjust one's behavior accordingly.

Suppose, then, that we identify Davidson's principles of coherence and correspondence as among the norms internal to thought.[11] To be a thinker, then, would involve, among other things, making at least some adjustments in one's thought — in one's beliefs, one's judgments — in the face of recognized incoherence (i.e., inconsistency in one's thought) or failure of correspondence (i.e., beliefs that prove false). The principles of coherence and correspondence, given our supposition,

[11] For a discussion of these principles, see SIO2, 211.

articulate an ideal internal to the concept of thought — an ideal in terms of which we assess the thought of ourselves and others. In terms of a metaphor to which I appealed above, this ideal defines an axis along which we place thinkers in evaluating their approximation to the ideal. So these principles are normative in the sense that they establish a basis for assessing thought, for making normative judgments about it. But we also face the question of whether a particular individual can be placed on this axis — whether the individual is properly seen as oriented toward this ideal. These principles are also constitutive of thought, then, in the sense that some minimal level of sensitivity and responsiveness to them is essential to one's being a thinker.

There are at least two different models for how we might understand the relationship between cognitive principles and the ideal, form, or standard embodied in the normative concept of a thinker. On the one hand, we might think of these principles as analogous to Thompson's natural-historical judgments. On this view, it would perhaps be misleading to characterize them as norms without making some qualifications: rather than involving any explicit 'ought's, such principles would delineate a form or ideal toward which we could see individual thinkers as oriented — a form or ideal of cognitive health or proper functioning. On the other hand, we might regard the normative concept of a thinker as a functional concept specifying an end (perhaps true belief) toward which the cognitive principles in question are conducive. To be a thinker, on this view, would involve being oriented toward this end, and taking as authoritative principles conducive toward it.

Whatever the model we choose, the general point I want to make is that if we look at the notion of a thinker as a normative concept embodying an ideal, form, or

standard, then we can render coherent the idea that certain principles play both normative and constitutive roles in thought. Moreover, we can do this in a way that does not seem to commit us to the overly intellectualized conception of thought that the constitutive thesis initially seemed to suggest. Certain norms or principles are constitutive of thought not in the sense that anything that is to count as thought must rigidly conform to them, but in the sense that any thinker must regard these norms or principles as having authority over their thought. And what counts as regarding a norm or principle as authoritative for one's thought? What I have been suggesting is that sensitivity and responsiveness to the norm or principle are the requisite characteristics: one must demonstrate some minimal ability to recognize circumstances in which the norm or principle applies and a readiness to adjust one's behavior accordingly.

4.3 A normative concept of belief

In the last section I tried to show that principles constitutive of thought could also be normative for thought. My argument so far is hypothetical: if we regard the concept of a thinker (or the concept of thought, the activity in which a thinker engages) as normative, then we can render the dual role thesis coherent. But is there any reason for believing that we have a normative concept of a thinker or the characteristic activity in which a thinker engages? The answer, I think, is that there is, but my argument for that point must await the next chapter. Here I simply want to point out that this normative conception of thought is not without precedent.

There is a widely held view that distinguishes cognitive from conative attitudes

in terms of the notion of 'direction of fit': while cognitive attitudes such as beliefs seek to fit the world, or to be true of it, conative attitudes such as desires seek to have the world fit them; in belief a proposition is treated as true, while in desire a proposition is treated as to be made true. As J. David Velleman has emphasized, however, it's not sufficient to characterize belief as an attitude toward a proposition that involves treating that proposition as true. This characterization is not sufficient because it doesn't succeed in distinguishing belief from other similar cognitive attitudes, such as assumption: when one assumes that p, for example, one also treats p as true, so there must be something further to distinguish belief from assumption. Velleman's claim is that we must distinguish between different cognitive attitudes in terms of their different 'constitutive aims':

> The clearest way to analyze such differences between belief and the other cognitive attitudes is in terms of the subject's dispositions to regulate his acceptance of a proposition. When someone assumes a proposition, he or his cognitive faculties are disposed to regulate his acceptance of it in ways designed to promote the ends of argument or inquiry: he comes to accept the proposition when doing so seems conducive to scoring a point or making a discovery, and he is disposed to continue accepting it only insofar as doing so seems to serve such polemical or heuristic purposes. . . . When someone believes a proposition, however, his acceptance of it is regulated in ways designed to promote acceptance of the truth: he comes to accept the proposition, for example, when evidence indicates it to be true, and he's disposed to continue accepting it until evidence indicates otherwise. Part of what makes someone's attitude toward a proposition an instance of belief rather than assumption or fantasy, then, is that it is regulated in accordance with epistemic principles rather than polemics, heuristics, or hedonics. An attitude's identity as a belief depends on its being regulated in a way designed to make it track the truth. [Velleman 1992, 14]

So the aim of assumption tends to be 'polemical or heuristic', while the constitutive aim of belief is to reflect the truth. To have a belief is to regard a proposition as true with

the aim of being rightly guided by the way things are. To put the point another way, to have a belief is to have a propositional attitude that one regards as responsible to a variety of truth-conducive norms. What Velleman presents here is a concept of belief with the structure of a normative concept: belief is defined in terms of an ideal of truth-tracking, and to be a believer is to approximate toward this ideal through conformity to truth-conducive norms. This concept of belief would fit the second model I described above: the norms constitutive of belief would derive their authority from an end — truth — toward which they are conducive — an end that is internal to the nature of belief.

4.4 Does the dual-role thesis yield categorical normativity?

As a first step toward responding to the naturalists' challenge to the idea of categorical cognitive normativity, I argued that the dual role thesis is coherent, provided we regard the concept of thought or cognition as a normative concept. But even if we accept this proviso and the coherence of the dual role thesis, the argument is still incomplete. At least three tasks remain. The first, and most obvious, I will begin to undertake in the remainder of this chapter: it must be shown that principles to which the dual role thesis is applicable — principles playing both normative and constitutive roles in thought — would possess categorical rather than hypothetical authority over thought. The second task is perhaps less obvious, although no less important: having shown that principles with dual roles would be categorical, it still remains to show how this account of the categorical nature of certain cognitive norms can meet Papineau's demand that "we should have some kind of understanding of the peculiar force that

[sui-generis, categorical] judgemental norms are supposed to exert on us." Finally, as I pointed out in the preceding section, my argument thus far is hypothetical, simply assuming that it is reasonable to regard our concept of a thinker as normative, so it also remains to show that there is some justification for so regarding it. I will take up the second and third tasks in the following chapter.

What reason, then, is there to think that principles with dual roles would be categorical? To say that such principles would be *hypothetical* is to say that their authority over an individual, qua thinker, is contingent upon that individual's desires. To put it another way, hypothetical cognitive norms do not have authority over thinkers as such, but only over thinkers with some particular desire. To identify a norm as categorical, on the other hand, is to say that its authority over one's thought is not contingent upon one's desires: categorical cognitive norms have authority over an individual qua thinker, not qua thinker-desirous-of-x. It would seem that norms constitutive of thought would be categorical, simply because part of what it is to be a thinker is to take these norms as authoritative. In other words, these are norms whose authority depends only on one's status as a thinker, and not on any particular desires one might happen to have.

There is good reason, then, to regard such norms as categorical, but perhaps it would nevertheless be constructive to look at the situation from another angle: are there any considerations militating against a view of constitutive norms as categorical? It might seem that there is at least one, and to bring it out it will help to look at a discussion in which Peter Railton describes the concerns of a fictional student, Gary, who persistently questions the authority of epistemic norms:

> Gary asked why he should pay attention to epistemic norms. If we reply that this is necessary in order to be a believer and thus to be an agent, he can respond: "But just how severe a cost does this threaten me with? Somewhere on the continuum between the ideal type of belief, on the one end, and clear non-belief on the other, there is a region which forms the borderland of genuine belief. I want to know why my attitudes should be on one side rather than the other of that borderland. The claim that I would cease to be an agent on one side of the region sounds dramatic. But if life on the believer side of the borderland has certain pluses and minuses, how do we know in advance that the balance must be worse on the other side? Mightn't it even be better, on the whole?" [Railton, 73-4]

There are two similarities worth noting between the account of belief at work in this passage and the account of thought that I have been developing in this chapter. First, both accounts regard commitment to certain epistemic or cognitive norms as constitutive of belief or cognition; one should pay attention to epistemic norms, on the account Railton envisions here, "in order to be a believer and thus to be an agent." And second, the concept of belief embraces a continuum, ranging from an ideal of complete conformity to epistemic norms to a minimal level of adherence necessary to sustain one's status as a believer.

The question of the authority of cognitive or epistemic norms is tied, in this passage, to the question of whether one wants to be a thinker (or believer), and tied to it in such a way as to suggest that the constitutive account merely pushes the question of their authority back a step. Gary's response to the constitutive answer to the question, "Why should I follow these norms?", is to ask "Why should I be a believer (or agent)?" Similarly, one might argue that the account of thought I've sketched is open to the same charge of merely postponing the real question of normativity at issue here. The real question would be "Why be a thinker?", and the suspicion might be that

the real authority of norms constitutive of thought is, in fact, conditional upon a desire — namely, the desire to be a thinker. Absent such a desire, one might argue, cognitive norms have no authority. Unless Gary is given some good reason for thinking he ought to be a thinker — and perhaps, as Railton's responses to Gary suggest, it's hard to imagine reasons here that don't show how being a thinker appeals to Gary's desires in some fashion — why should he take seriously such norms?

The criticism I'm considering here is related to an issue I will take up in the next chapter: the question of how the account I'm developing of categorical cognitive normativity might meet Papineau's demand for an account of how categorical cognitive norms are supposed to motivate us. So we will revisit some of these issues later. But at this point I want to respond at least in a preliminary way to the criticism. The concern behind Gary's response may be something along these lines: by harnessing cognitive norms and their authority to an account of the nature of thought (or belief), we succeed in explaining how such norms might be categorical, but at the cost of raising the further question of what normative claim the status of 'thinker' or 'believer' has over us.[12] Since the only account of categorical normativity on the table so far appeals to principles playing a constitutive role in thought (or belief), and we're now questioning the value of being a thinker (or believer), it seems unlikely that whatever answer we give to the question, "Why should I be a thinker (or believer)?", could plausibly identify categorical reasons for valuing that status. Hence, it seems that any reasons we offer at this level will have to be hypothetical (appealing perhaps to one's

[12] As we will see in the next chapter, the concern Gary articulates here is quite similar to a concern that lies at the center of Stich's *The Fragmentation of Reason*.

desire to be a thinker), and if it is correct to think that it is at the level of questions concerning our reasons for being thinkers or believers that questions concerning the authority of cognitive norms are resolved, then it would seem that cognitive normativity must, at a fundamental level, be hypothetical.

The key move in this criticism is the contention that the constitutive account of normativity 'postpones' or 'pushes back' the real question concerning the nature of cognitive normativity. It is alleged to postpone the question in the following sense: even if we have identified certain cognitive norms as constitutive, and thereby secured for them a sort of categorical authority, we still haven't fully explained the nature of cognitive normativity until we address what is, in essence, Gary's question: "Why be a thinker (or a believer)?"

Although a full response to this criticism must await the next chapter, it is still worth emphasizing at this point some reasons for resisting it. First, it does not directly challenge the idea that certain cognitive norms might be categorical, and in fact it seems to leave that view untouched. As I have indicated, Gary's example might seem to suggest that any normativity in cognition must ultimately be contingent upon one's desire to be a thinker. But if my appeal to the notion of a normative concept is at all plausible as a way of making sense of the dual-role thesis — if it is part of the concept of thought that to be a thinker is to take certain norms as authoritative — then there are certain norms whose authority over thought is not contingent upon one's having a particular desire, the desire to be a thinker or any other. There are surely plenty of thinkers who have never considered whether or not they want to remain thinkers, but if my account is correct they are still subject to the authority of these norms.

A somewhat different suggestion would be that while cognitive norms do not depend for their authority on one's having a particular desire, they do depend upon one's *not* having a particular desire: namely, the desire to end one's status as a thinker. Just as a person intending to drop her membership in a particular organization might regard its rules as lacking any authority over her behavior, Gary might come to see epistemic norms as similarly lacking in authority if he determines that life on the non-believer side of the borderland really is preferable. But this suggestion has its problems too. While it's certainly true that Gary would cease to be subject to cognitive norms were he to cease to be a cognitive agent — if, that is, he were no longer capable of forming beliefs, reasoning about them, and engaging in whatever other activities we take to be essential to cognition — it's not clear that this fact, or Gary's desire to cease being an agent in this sense, has any bearing on whether he is subject to such norms as a thinker. If the account I have sketched is correct, then such norms possess authority over Gary's thought so long as he remains a thinker, and, as Railton points out[13], deference at the very least to truth-conducive norms and norms involved in instrumental reasoning seems to be presupposed in the very question he asks: whether it is preferable for Gary to be a believer or a non-believer is settled, it would seem, only by trying to ascertain the truth concerning which form of existence is most conducive to what Gary wants. So one might argue that Gary's even raising the question of whether or not life on the non-believer side of the borderland is preferable presupposes that cognitive norms retain their authority over him.

[13] Railton, 75.

What I've said here in response to Gary's objection only scratches the surface of what is — so I will argue — a complex set of objections to an account of categorical cognitive normativity of the sort I am developing here. I will take up discussion of this objection again in the next chapter, where I will look at a version of it that arises out of Stich's work.

4.5 Conclusion

I have argued in this chapter that if we regard certain cognitive concepts, such as the concept of a thinker or believer, as normative concepts, then (a) we can render coherent the dual-role thesis that we find in Frege and Davidson, and (b) we have the beginnings of an account of how certain cognitive norms could be categorical. Obviously there is a great deal more to say about (b), and in the next chapter I will try to flesh out the account of categorical cognitive normativity that I have begun here.

Chapter 5: In Defense of the Dual Role Thesis

5.1 Gary's objection

At the end of the last chapter I looked at an objection, arising out of a discussion of Peter Railton's, to the idea that we might ground the categorical authority of cognitive norms in their constitutive relation to thought. The gist of the objection — which I'll refer to as "Gary's objection", in honor of its fictional author — was that this sort of attempt to gain a foothold for categorical norms within cognition fails, because it merely pushes the question of the authority of such norms back a step: if we try to account for the authority of cognitive norms by pointing to their role as constitutive of thought — in essence, claiming that one must heed these norms if one is to be a thinker — we succeed only in raising the question of what reason one has to be a thinker. Because the only account of categorical normativity on offer appeals to principles constitutive of thought, it's not clear how we might appeal to anything beyond hypothetical reasons in trying to answer the latter question. And if it is true that we still haven't explained the nature of cognitive normativity until we answer the question, Why be a thinker?, then it might still seem that cognitive normativity is fundamentally hypothetical.

I made, at the time, two observations about this objection. First, it does not, in any obvious way, undermine the account of categorical cognitive normativity that I offered in the last chapter, for it does not challenge the idea central to that account: that there might be principles associated with a normative conception of thought that

exert an authority over us conditional only upon our status as thinkers, and not upon any desires we might or might not have. Second, as I pointed out in the preceding paragraph, it presupposes that we haven't explained the nature of cognitive normativity until we answer the question, Why be a thinker? I now want to look more closely at these two observations, beginning with the second.

5.1.1 The presuppositions behind Gary's objection

First, let me revise my second observation slightly. There are, it seems, two presuppositions built into Gary's objection, at least as I've explained it. First, it seems to presuppose that a full accounting of cognitive normativity must answer the question, Why be a thinker? And second, it takes for granted that it is the answer we get to *this* question — or rather, the sort of reasons, categorical or hypothetical, to which we must appeal in answering this question — that will determine the fundamental character of cognitive normativity. Neither presupposition, however, is obvious.

Concerning the first, we might profitably ask what audience we have in mind for these questions concerning cognitive normativity. Although this point may not take us very far, it is worth noting that presumably anyone capable of entertaining the question, Why follow these (cognitive) norms?, must be a thinker or a cognitive agent, someone capable of forming beliefs and reasoning about them. Moreover, anyone capable of understanding or arriving at a reasoned answer to this question will also be a thinker. Given that the audience for this question does and must consist solely of thinkers, who cannot relinquish that status without ceasing to grasp the question and the reasons on one side or the other, one might reasonably ask why we should assume

that a justification for following cognitive norms must be given from a perspective which does not take for granted participation in the activity those norms govern. What is at issue here, after all, are norms governing *thinkers*, and it does not seem completely implausible to maintain that we should keep separate the question how one ought to think, given that one is a thinker, and the question whether one has good reason to be a thinker.[1]

My concern with the second presupposition is related to this last point. Suppose, as I've suggested Gary's example assumes, that any attempt to answer the question, Why be a thinker?, can appeal only to hypothetical reasons for preferring thinkerhood to nonthinkerhood or vice versa. Would this demonstrate that cognitive normativity must be fundamentally hypothetical? To answer this question it may help to consider an analogy. Our actions are often guided by moral commitments, but it is

[1] One reason for not keeping these questions separate would be a sense of what we're trying to do in theorizing about the authority of norms in various areas: we're trying to find some way of balancing competing normative demands on our action. Given this view of such theorizing, we should be less interested in the question, 'Given that I'm a thinker, how should I think?', and more interested in finding some way of balancing one's responsibilities as a thinker with one's responsibilities as, say, a religious believer, a spouse, or a decent human being. Doing the latter requires seeking some determination of the weight one's role as a thinker should have in one's deliberations in comparison to the other roles one inhabits. Where I take issue with this view, however, is in its conception of thinkerhood as just one more role that one might (or might not) take up alongside others — a role that we can bracket in weighing the normative demands issuing from the various areas of our lives. One's identity as a thinker is more fundamental than, say, one's identity as a spouse, for thought, and therefore the norms governing thought, are implicated in all other normative arenas as well.

arguably possible to bracket these commitments and assess, from a purely self-interested standpoint, whether one is better off taking these commitments seriously or not. Does the possibility of such an assessment — one that does not take for granted the authority of a moral perspective — demonstrate that the norms associated with that perspective are merely hypothetical or prudential? The answer, I think, is that it does not: instead, it merely shows that the perspective of morality is not the only normative perspective from which one might evaluate one's behavior. Similarly, the possibility of adducing hypothetical reasons for or against being a thinker in no way shows that cognitive norms are not categorical. At most, it shows that there are various normative perspectives from which we might assess cognition. But it is no part of the categorical conception to deny that considerations drawn from other normative arenas might also have some bearing on how we think, and on our conception of how we ought to think.

5.1.2 A hollow victory?

Although it appears that Gary's objection leaves untouched the account of cognitive normativity that I offered in the last chapter, my responses to it may nevertheless seem dissatisfying; it may still seem that there is something fundamentally correct about his skepticism concerning the constitutive strategy that my responses gloss over. After all, what occasions the sort of concern Gary expresses is the very move that is central to the constitutive account of cognitive normativity: the association of certain norms with the nature of thought. So if we were skeptical of the authority of these norms in the first place, the constitutive strategy may succeed only in extending this skepticism, or

relocating its focus, from cognitive norms to the status of which they are alleged to be constitutive — the status of being a thinker. What is needed is something to address this skepticism — an account demonstrating how categorical norms, as I have conceived them, could plausibly attract and retain our allegiance. Without such an account, it may be that we can still hold onto the view of categorical cognitive normativity that I offered in the last chapter — it seems that Gary's objection does not jeopardize that account — but this success will represent only a hollow victory over proponents of the hypothetical conception, for we will still be left without any explanation of the hold such norms might have over us. In other words, although Gary's objection can be taken as targeting the categorical conception, it is perhaps more plausible, and more serious, when understood as pointing to a weakness inherent in the constitutive account of cognitive normativity: namely, that it does not by itself meet Papineau's demand that "we should have some kind of understanding of the peculiar force that [sui-generis, categorical] judgemental norms are supposed to exert on us." The constitutive account shows us how there might *be* such norms, but, as it stands, it does not yet explain why we might take such norms as authoritative.

As a way of amplifying Gary's objection, as well as indicating why it might seem difficult employing the constitutive strategy to overcome it, I want to turn back to the view Stich defends in *The Fragmentation of Reason.* The view he defends there, I will argue, is similar in many ways to Gary's objection, and it poses a serious challenge for attempts to defend the categorical conception via a constitutive account of categorical norms.

5.1.3 Conceptual claims and Stichean objections: a Stichean Gary

What I argued in the last chapter is, in essence, that categorical normativity is grounded in conceptual truths — truths about the concepts in terms of which we think about cognition: the idea that there are categorical norms governing cognition is tied, on this account, to the status of our concept of thought, and that of related concepts, as normative. But it is this feature of the view I have defended that makes it seem particularly susceptible to a line of criticism advanced by Stephen Stich and Hilary Kornblith, the thrust of which is that conceptual or analytic truths are not adequate to explain cognitive normativity. As Kornblith puts it "Semantic considerations alone. . . cannot explain the normative force of epistemic terms" (Kornblith 1993, 362).[2]

For both Stich and Kornblith, this concern with the effectiveness of semantic or conceptual analysis arises out of a dissatisfaction with the methods of analytic epistemology. In particular, they argue, its methods are not adequate to its normative ambitions: while analytic epistemology seeks, among other things, to tell us "which ways of going about the quest for knowledge. . . are the good ones [and] which are the bad ones" (FR, 1), and in doing so to tell us how we ought to conduct our epistemic and cognitive lives, the analyses upon which it relies to reach these conclusions cannot provide any compelling justification for them. The analytic epistemologist, Stich explains, "proposes to evaluate. . . different cognitive processes by explicating our intuitive notions of cognitive evaluation, and then exploring which inferential processes fall most comfortably within the extension of those notions" (FR, 20). In

[2] The discussion here and in the next two paragraphs of the objections advanced by Stich and Kornblith covers some of the same ground as my discussion in Chapter Two, pp.21-23.

other words, the norms governing cognition are to be drawn from the concepts in terms of which we already think about and evaluate cognition — our concepts of belief, rationality, and justification, for example. This prompts Stich to ask why anyone "would care whether their reasoning falls within the boundaries of the intuitive notion of rationality" (ibid.). So, given an account of cognitive evaluation guided by the concepts in terms of which we think about cognition, the question, in essence, is why anyone would want to be a thinker, or a rational being, in this intuitive sense. Stich's move here recalls a similar move on Gary's part: faced with the idea that the authority of epistemic norms might be tied to the nature of belief, Gary takes this not to vindicate such norms but to cast doubt on the value of belief. And Stich's answer reflects a concern similar to Gary's: "It is hard to see why most people would *care* very much whether a system of cognitive processes falls within the extension of some ordinary notion of epistemic evaluation" (ibid.). Far from vindicating our ordinary notions of cognitive evaluation, pointing out that these notions are bound up with our concept of belief succeeds only in pushing back concerns about these notions to a level — that of our commonsense concepts — at which they are unanswerable.

What is it, though, about the role of "intuitive notions" or "commonsense concepts" in analytic epistemology that casts doubt on its normative conclusions? Why, on Stich's view, are such considerations insufficient to justify the normative claims based on them, or to address concerns about those claims? Stich's argument isn't as clear as it might be, and passages like the one near the end of the preceding paragraph might seem to suggest that his concerns here are fueled by nothing more than his own intuitions concerning what "most people would care very much" about.

The following passage, however, suggests a more substantive argument:

> [I]magine that we have located some exotic culture that does in fact exploit cognitive processes very different from our own and that the notions of epistemic evaluation embedded in their language also differ from ours. Suppose further that the cognitive processes prevailing in that culture accord quite well with *their* evaluative notions, while the cognitive processes prevailing in our culture accord quite well with *ours*. Would any of this [i.e., considerations concerning our cognitive concepts and the notions of cognitive or epistemic evaluation embedded in them, such as the points I've made about the normative concept of belief] be of any help at all in deciding which cognitive processes we should use? Without some reason to think that one set of evaluative notions was preferable to the other, it seems clear that for most of us it would be of no help at all. [FR, 92-3]

What Stich seems to have in mind (and Kornblith, too, inasmuch as he follows Stich in these matters[3]) is something along the following lines. What undermines the idea that cognitive normativity can be grounded in the concepts in terms of which we think about cognition is the possibility of alternative such concepts. If our reply to a Stichean version of Gary is that he ought to take seriously these cognitive norms because doing so is essential to being a thinker, he is likely to point out that our account of the authority these norms possess merely raises the further question of why he ought to be a thinker, as opposed to a thinker* or thinker$^{*\ldots*}$, where the asterisks indicate that we are referring to normative concepts and associated notions of cognitive evaluation belonging to exotic cultures. To endorse certain norms and the cognitive processes they sanction simply because they are products of the concepts in terms of which we think about cognition is, as Stich puts it, to display "epistemic chauvinism". But this is, in effect, what the analytic epistemologist does: he privileges our epistemic concepts, and the norms to which they give rise, without offering any justification for preferring

3 See Kornblith 1993, pp. 361-3, where he explicitly endorses Stich's arguments on this score.

those concepts and norms over their rivals. It is for this reason that Kornblith maintains that "semantic arguments cannot explain the source of epistemic normativity" (Kornblith 1993, 362): they cannot explain why we might prefer one set of cognitive norms over another. Whatever authority a set of cognitive norms has over us is not the product of its relation to a normative concept, but of our having some independent reason for favoring that normative concept over others. Stich and Kornblith maintain that such reasons will be hypothetical, appealing to the desires served by one set or another.

The line of argument that Stich and Kornblith advance is similar to Gary's objection. Although, unlike Gary, they do not explicitly frame their objection as a criticism of the constitutive strategy, their arguments are easily adapted to that purpose, and seem to reach the same pessimistic conclusions concerning that strategy as Gary's objection: what initially seemed a strength of the constitutive strategy — that it provides us with a basis for making sense of the idea that cognitive norms might be categorical — turns out to be a liability, for that basis is insufficient to explain the normative force of the principles which rest on it. Once again, the problem is that the constitutive strategy pushes back a step questions concerning the nature of cognitive normativity without really addressing them: Gary's claim was that it simply transfers skepticism concerning normative principles to the status of which these principles are alleged to be constitutive, while the argument that Stich and Kornblith advance suggests that the strategy simply raises questions concerning the authority of the concepts that are supposed to yield categorical cognitive norms.

But the Stich-Kornblith argument goes beyond Gary's objection, and in doing

so offers what is potentially a much more serious indictment of the account of cognitive normativity I introduced in the last chapter. For while Gary's objection identifies a shortcoming with the constitutive strategy as I've explained it so far — namely, that it does not address Papineau's demand — the Stich-Kornblith argument suggests, in addition, that this is a shortcoming that the account lacks the resources to overcome. By relying on claims about the concepts in terms of which we think about cognition, their argument suggests, this account cannot give a satisfactory explanation of why we should take as authoritative the norms it identifies as categorical: it not only *does not* meet Papineau's demand; it *cannot* meet it.

5.2 Disambiguating the Stichean Gary

Stich and Kornblith are not alone in expressing concern about the prospects for an attempt to provide conceptual foundations for normativity. Railton, for example, voices a similar worry concerning the constitutive strategy, claiming that "[i]t would be surprising if we could give an answer [to something like Gary's skepticism toward the authority of practical and epistemic norms] with nothing more than a few definitions. To be genuinely responsive to the concerns expressed, constitutive arguments must capture a substantive — not merely linguistic — necessity" (Railton, 70).

I tried in the preceding section, and at certain points in Chapter Two, to make sense of the criticism implicit in Stich and Kornblith of the use of conceptual claims in arguments concerning cognitive normativity. In order to respond to their criticism effectively, it is of course essential to have a clear view of what that criticism is. But

despite the work I've already done trying to clarify their criticism, it will be useful at this point to back up and look more closely at it, for I think it is possible to see more than one criticism concealed there. Much of the rest of the chapter will therefore be devoted to distinguishing, clarifying and responding to the various criticisms underlying the contention, common to Stich, Kornblith, and Railton, that conceptual claims are not sufficient to address concerns about normative authority.

5.2.1 *Quid juris?*

Kant introduces the idea of a deduction of the categories by means of a legal analogy: "Jurists, when speaking of rights and claims, distinguish in a legal action the question of right (*quid juris*) from the question of fact (*quid facti*); and they demand that both be proved. Proof of the former, which has to state the right or the legal claim, they entitle the *deduction*" (Kant 1929, A84/B116). He goes on to draw the connection between this legal notion of deduction and a deduction of concepts. I will quote this passage at length, because I think it sheds light on one way of understanding the charge brought by Stich et al., against the constitutive strategy:

> Many empirical concepts are employed without question from anyone. Since experience is always available for the proof of their objective reality, we believe ourselves, even without a deduction, to be justified in appropriating to them a meaning, an ascribed significance. But there are also usurpatory concepts, such as *fortune*, *fate*, which, though allowed to circulate by almost universal indulgence, are yet from time to time challenged by the question: *quid juris*. This demand for a deduction involves us in considerable perplexity, no clear legal title, sufficient to justify their employment, being obtainable from experience or from reason. [ibid., A84-5/B116-17]

"To be genuinely responsive to the concerns expressed" about the normative authority

of a principle, Railton maintains, "constitutive arguments must capture a substantive — not merely linguistic [or, we might add, conceptual] — necessity". The constitutive strategy attempts to secure the categorical authority of a norm by arguing that the norm is constitutive of thought. And, as I argued in the last chapter, we can make sense of the idea of a norm's being constitutive of thought by appealing to the notion of a normative concept: to regard certain norms as constitutive of thought is to regard the concept of thought, or of a thinker, as normative — as requiring sensitivity and responsiveness to certain norms as a condition of something's falling under the concept.

One way of construing Railton's claim would be to see him as insisting on the necessity, to put the matter in Kantian terms, of a proof of the "objective reality" of the normative concept of thought — as insisting, in other words, on the need for a deduction of this concept. I've demonstrated, I hope, the possibility of such a concept — we *can* think of cognition in terms of normative concepts, and doing so seems to secure a place for categorical cognitive norms — but not its legitimate application to human cognition — is human thought such that we *ought* to think about it in terms of normative concepts? The former — what I have demonstrated — is analogous to Kant's question of fact; the latter, to the question of right. Having merely shown that a normative concept of thought is possible, and would explain categorical cognitive norms, it remains to show that it actually captures something of the nature of thought — that it identifies, as Railton puts it, a substantive and not merely a conceptual necessity. Conceptual claims, in the absence of some demonstration that the concepts involved are the appropriate ones in terms of which to think of human thought, cannot

underwrite normative claims concerning how we ought to think.

5.2.2 An argument from relativity

On the face of things, the last objection to the constitutive strategy involved no appeal whatsoever to the possibility of alternative concepts of cognition. But as I have said, it's clear that such an appeal is involved in some way in the concerns expressed by Stich, Kornblith, and Railton about grounding cognitive normativity in conceptual or linguistic analysis. Stich, for example, claims that "the most intuitive way" to see the inadequacy of the analytic epistemologist's attempts in this regard "is to begin by noting how the specter of culturally based cognitive diversity lends a certain urgency to the question of which cognitive processes we should use" (FR, 91). Kornblith cites Stich's argument approvingly, and points to the availability of "conceptions of justifiedness other than the ordinary one" as evidence that "[s]emantic considerations alone thus cannot explain the normative force of epistemic terms" (Kornblith, 362). Finally, even Railton suggests a connection between normative diversity and skepticism toward familiar norms: Gary's failure to suggest "anything like the possibility of an alternative to familiar forms of practical reasoning" reflects poorly on his skepticism concerning the norms governing such reasoning (Railton, 75).

What this emphasis on normative diversity and conceptual alternatives suggests is an argument from relativity. As Mackie explains it, "[t]he argument from relativity has as its premiss the well-known variation in moral codes from one society to another and from one period to another, and also the differences in moral beliefs between different groups and classes within a complex community" (Mackie, 36). The

conclusion, at least with respect to moral beliefs and judgments, is that we should not treat them "as apprehensions of objective truths" (ibid.). The route from premise to conclusion is not deductive; we should look at the argument instead as abductive, contending that the best explanation for the diversity of moral opinion is that there is no truth at which this opinion aims. If there were objective moral truths to which we had access, then we would expect more agreement in moral beliefs. Given widespread *disagreement*, the best explanation is that there are no objective moral truths to which we have access.

The situation with respect to cognitive norms is, at least as Stich sees it, similar. His appeal to the "specter of culturally based cognitive diversity"[4] suggests that we face a potential variety of conceptions of the nature of cognition and of how we ought to think, with no clear prospects for adjudicating between rival conceptions. And "if we cannot say why our cognitive processes are any better than those prevailing elsewhere, it suggests that it is ultimately no more than an historical accident that we use the cognitive processes we do or that we hold the beliefs those processes generate, just as it is an historical accident that we speak English rather than Spanish and wear trousers rather than togas" (FR, 91). So, in the absence of some justification of our concepts as against others, we cannot regard our commitment to particular conceptions of cognition or particular cognitive norms as the product of any apprehension on our part of truths concerning the nature of thought or the normative

[4] Not to mention his talk of "the idiosyncratic nature of our notions of epistemic evaluation, and the existence of a wide range of possible alternative notions" (FR, 96).

standards governing it.[5] The problem with the attempt to ground cognitive normativity in conceptual claims, then, is that "the existence of a wide range of possible alternative [cognitive] notions" (FR, 96) undermines any presumption of authority that our own cognitive concepts, or norms associated with them, might have.

5.2.3 Internalism and the motivational impotence of conceptual claims

Thus far the objections I've explained raise, in somewhat different ways, the question of whether particular cognitive concepts in terms of which we do (or might) think of cognition have any claim to be the concepts in terms of which we *ought* to think about it. The final objection I want to introduce takes a different tack. The sort of concerns expressed by Stich, Kornblith, and Railton about conceptual claims can be taken not as questioning the legitimacy of particular cognitive concepts, but instead as questioning whether any conceptual claims can have motivational force.

Take 'internalism', in the present context, as the view that "the claim of any class of considerations to provide reasons for action is hostage to the motivational

[5] Even though Stich doesn't explicitly present an argument from relativism, it is worth noting that the conclusion of such an argument neatly lays the groundwork for Stich's pragmatic account of cognitive evaluation: if we cannot rely on the conceptions of cognition or cognitive evaluation with which we are familiar to disclose anything of value concerning how we really ought to think, Stich argues, one obvious alternative is simply to assess various conceptions in pragmatic terms, to see which ones are most effective in getting us what we want. This is not to say that I fully endorse Stich's argument, or that it is without problems, but by ruling out the idea that these conceptions have any intrinsic value, we're left with the alternative of assessing them in terms of their extrinsic value — i.e., their conduciveness to what we desire.

efficacy of those considerations" (Scheffler, 74). If we construe 'action' broadly, so as to include cognitive 'actions' — the formation of beliefs, insofar as that is volitional, as well as reasoning about one's beliefs — internalism is the view that if any consideration is to count as a cognitive reason, then it must be capable of motivating us. A fortiori, if any consideration is to count as a *categorical* cognitive reason, then it too must be capable of motivating us. This view, it is worth noting, is one element of what I described in Chapter Two as the Humean theory of motivation: there I described Hume's assertion that "morals excite passions, and produce or prevent actions" as a specific version of the more general view that it is characteristic of reasons that they are capable of motivating us to action.

The question, then, is whether conceptual considerations — considerations about the concepts in terms of which we think about cognition — have any motivational efficacy. It seems reasonable to read Stich, at certain points, as claiming that they do not: "It is hard to see why most people would *care* very much whether a system of cognitive processes falls within the extension of some ordinary notion of epistemic evaluation" (FR, 20).[6] In other words, the mere fact that some cognitive process is sanctioned by a particular concept is not something about which most people would care, and therefore it is not by itself something that would motivate most people to adjust their cognitive behavior. But if conceptual claims lack motivational force, then, according to internalism, they cannot yield reasons for action, cognitive or

[6] Perhaps we can see a view similar to Stich's in Railton's assertion that "[i]t would be surprising if we could give an answer [to something like Gary's skepticism toward the authority of practical and epistemic norms] with nothing more than a few definitions" (Railton, 70).

otherwise. And if conceptual claims cannot yield reasons for action, then any attempt, like that of the last chapter, to ground the normative authority of certain cognitive principles in conceptual claims will be a nonstarter.

5.3 Responding to the Stichean objections

The objections I rehearsed in the preceding section fall naturally into two categories. The first two objections can both be understood as emphasizing the importance of addressing Kant's 'question of right': particularly in light of the alleged diversity of cognitive concepts, so the concern goes, it is absolutely essential that we have some justification of the concepts in terms of which we seek to understand thought, some demonstration that these concepts capture a substantive and not just a linguistic necessity. If we are to rely on conceptual claims to underwrite the authority of cognitive norms, we must demonstrate that the concepts involved limn the nature of thought, rather than merely expressing our parochial understanding of it.

The third objection is not unrelated to the first two. It too questions whether conceptual claims can be used to underwrite cognitive normativity: if such claims are motivationally impotent, and considerations appealing to them cannot therefore constitute reasons for action, then it would seem that any strategy that seeks to underwrite cognitive normativity by appealing to the concepts in terms of which we understand thought must fail. One way of looking at this objection would be to see it as making the same point as the first two objections: absent some reason for thinking that our concepts capture substantive necessities concerning thought, they cannot underwrite the authority of cognitive norms, for mere concepts or "nothing more than

a few definitions" will not suffice. But the difference between the two sets of objections lies in their respective accounts of *why* mere concepts or "nothing more than a few definitions" will not suffice: the first two objections argue that conceptual claims are inadequate for this task because substantive claims are indispensable to it, while the third argues for their inadequacy by pointing to their motivational impotence. It is to these two concerns about my strategy in Chapter Four that I will respond in what follows.

5.3.1 Objections from substantive indispensability[7]

So far my primary aim has been simply to secure conceptual space for the view that some cognitive norms might be categorical, and to do that I found it necessary in the preceding chapter to introduce the notion of a normative concept. But thus far I have not defended any particular view concerning *which* norms might be categorical for cognition, nor have I defended a view of any particular cognitive concepts as normative.[8] The limitations of my project up to this point, however, put me in an awkward position with respect to the demand for some justification of the concepts in terms of which I am suggesting we understand thought: if I'm not defending any specific cognitive concepts, how can I possibly meet this demand?

[7] I will sometimes refer to these — the first and second objections, discussed in 5.2.1 and 5.2.2 — as the SI objections.

[8] Granted, I do discuss in Chapter Four Velleman's claims concerning the 'constitutive aim' of belief. My aim there, however, is not to defend that particular conception of belief, but rather to cite it as an example of how we might regard certain cognitive concepts as normative.

One way of answering this question goes back to Papineau's distinction between naturalism and non-naturalism. Naturalism, as Papineau describes it, holds that "sensitivity to norms of judgement is an addendum to the possession of beliefs itself" (NJ, 32), while non-naturalism insists that we must appeal to a "peculiar species of content-constituting norms" (NJ, 22) in explaining the intentional content of such things as beliefs. Papineau's non-naturalism, then, seems to be a label well-suited for the view I described in Chapter Four: I am arguing for a view according to which sensitivity and responsiveness to norms of judgement (or, more generally, cognitive norms) is not an addendum to thought, but essential to it. So here would be one way of answering the question: while I'm not arguing for any specific cognitive concepts, I am arguing for a conception of thought that has consequences for which specific concepts we regard as plausible. It is this conception, associated with the view that certain cognitive concepts must be normative, that requires some justification.

The problem with this suggestion is that it is not clear to me how to defend non-naturalism in the abstract, without defending the normative status of some particular cognitive concept. So what I intend to do is to return to the example I presented in Chapter Four: the concept of belief. This concept is certainly central to our understanding of cognition, so if a case can be made for the normative status of cognitive concepts, it seems to me that it ought to be made here. And I think that a plausible case can, in fact, be made for the idea that our concept of belief is a normative concept. Moreover, it is primarily in terms of belief and the epistemic norms governing it that Stich and Kornblith frame their discussions, so focusing on belief will make it easier to frame my own position in a way that lines up with their concerns. The

remaining question, then, is how to defend the idea that we should look at the concept of belief as a normative concept.

I suggested above that we can look at Railton's demand for substantive necessities, and not just conceptual or linguistic claims, as analogous to Kant's emphasis on the need for a deduction to demonstrate the objective reality of certain concepts. This analogy might also seem to indicate one way in which we could respond to this demand and justify a normative conception of belief: if a transcendental deduction is Kant's answer to the question of right concerning the categories, why not consider a transcendental argument as a way of addressing Railton's concern? Showing that sensitivity and responsiveness to certain norms are a necessary condition for having beliefs would vindicate the normative concept of belief, demonstrating that the necessity here is not just conceptual but substantive.[9] Moreover, there is precedent for understanding the constitutive thesis in the context of some sort of transcendental argument: Davidson himself suggests that his argument against conceptual relativism, in which the constitutive thesis figures prominently, should be read as a transcendental argument.[10]

Despite the appeal of such arguments, neither their nature nor the force of their

[9] One could look at the dual role and constitutive theses *as* transcendental arguments, or at least as parts of such arguments. To say that certain principles or norms are constitutive of thought is, one might argue, simply to say that conformity to the principles or responsiveness to these norms is a necessary condition for the possibility of thought.

[10] ITI, 72.

conclusions is clear.[11] Furthermore, it's unclear whether such arguments would have any traction against an opponent like Stich. One of the standard concerns about transcendental arguments is that they are infected by an 'irremediable parochiality', rendering them "incapable of yielding conclusions applicable beyond some limited sphere characterized variously as, say, 'my experience', 'our form of life', 'our conceptual scheme', or 'the species to which we belong'" (Bell, 192). Given such concerns, transcendental arguments hardly seem a promising route for responding to worries about our parochial conceptions of cognition (Stich, Kornblith) or the demand for substantive rather than conceptual necessities (Railton).

So my strategy in responding to the first two objections will be more modest. There are two points I want to make in what follows. First, I will argue that it is, in fact, quite important to the force of the SI objections that they not merely assert the bare possibility of alternative conceptions of thought or belief, but that they demonstrate the existence of actual alternatives, and this they fail to do. Second, I will point to some difficulties with a non-normative conception of belief that militate in favor of a normative conception.

5.3.1.1 A "wide range of possible alternative notions"?

Mackie explains that the conclusion of the argument from relativity is that our moral beliefs are not "apprehensions of objective truths". If we understand Stich's focus on the parochiality of our cognitive notions and the availability of alternatives as part of an argument from relativity, its conclusion will be that our beliefs concerning how we

[11] See Stroud 1968 for a seminal critique of such arguments.

ought to think should not be regarded as apprehensions of the intrinsic value of the processes sanctioned by those beliefs: instead, "it is ultimately no more than a historical accident that we use the cognitive processes we do or that we hold the beliefs that those processes generate, just as it is an historical accident that we speak English rather than Spanish, and wear trousers rather than togas" (FR, 91).

One can imagine how a toga-wearer's discovery of trousers might feel liberating, and constitute an occasion for some serious reflection on the relative merits of the two forms of dress for the activities in which she engages. Similarly, Stich intends his work both to liberate us from the notion that there is something intrinsically valuable about the conceptions of cognition to which we are committed, and to convince us of the view that the proper way to assess the rival claims of various conceptions of cognition is to look at their extrinsic value: to weigh their relative merits with respect to the attainment of our multifarious goals. But note that it is important to the story of the toga-wearer's liberation that she discover an alternative to togas. If one were living in the benighted land of toga-wearers, where no one has any inkling of anything other than togas, the mere suggestion, in the absence of examples, that there might be alternative forms of dress would be unlikely to occasion the same feelings of liberation or soul-searching concerning the practical merits of various outfits.

I think there is an analogous point to be made about the concepts in terms of which we understand cognition. According to the SI objections, conceptual claims are inadequate to account for the authority of cognitive norms without some convincing arguments to the effect that these claims capture substantive necessities about the nature of cognition. On the one hand, I think there is something to this point: I

certainly don't want to deny the importance, for example, of ensuring that theoretical or conceptual investigations have some grounding in the actual objects or events they are about. On the other hand, though, there is something peculiar about these objections. Conceptual claims, after all, express our understanding of the nature of a thing, and the emphasis in SI on the 'merely linguistic' or conceptual suggests that we can prize apart the thing itself and the concepts in terms of which we understand it. It is not that we cannot do this, but rather that we typically do it under the pressure of considerations suggesting that our understanding of a thing could in some way be defective, or might not capture its nature.[12] In the absence of such considerations, the demand for a justification of the concepts in terms of which we understand a thing may seem unmotivated.

Now one sort of consideration that might lead us to look with suspicion on the concepts in terms of which we understand something is a recognition that other seemingly reasonable people have quite different ways of understanding it. This seems

[12] Consider, for example, the way in which Kant's treatment of the concept of causation in the *Critique of Pure Reason* is a response to Hume's claim that we can find nothing in our experience of the outside world corresponding to the idea of a necessary connection between cause and effect. One way of looking at the point I'm making in this section, particularly in the next paragraph, is to see it as maintaining that we can't really keep Railton's objection about the indispensability of substantive claims (Section 5.2.1) separate from the argument from relativity (Section 5.2.2): the idea that conceptual claims require some substantive justification is typically motivated by, among other things, the sort of considerations concerning conceptual variation to which the argument from relativity appeals. Absent some considerations pointing to a need for substantive justification, the sort of demand that Railton emphasizes is unmotivated.

to be the precisely the sort of consideration that Stich has in mind in characterizing our cognitive concepts as parochial and commonsense. It seems worth asking, then, whether we can find in Stich, Kornblith, or Railton a non-normative conception of belief — as Papineau would put it, a 'naturalist' conception of belief according to which sensitivity and responsiveness to norms are in no way constitutive of belief. Such a naturalist conception would be the sort of genuine alternative to a normative conception of belief that would motivate the demand for a substantive justification of the latter conception.

The answer, I think, is that there is no naturalist rival to the normative conception of belief evident in Stich, Kornblith, or Railton. But I want to spend some time looking more closely at the conception of belief underlying Stich's *The Fragmentation of Reason:* it is the closest thing to such a conception in their works and is easily mistaken for precisely the sort of non-normative, naturalist conception that might constitute a serious challenge to the normative conception of belief.

Stich maintains that "to have a belief. . . is to have a token of a well-formed formula stored appropriately in one's brain" (FR, 109). On Stich's view, to have any propositional attitude involves having a well-formed formula stored somehow in one's brain, so it is the notion of a token's being 'stored appropriately' that is the key to the way in which Stich distinguishes belief from other sorts of propositional attitudes. Different sorts of propositional attitudes are stored in different 'boxes':

> In thinking about the assignment of content to mental sentences, it proves enormously useful to adopt the myth that mental sentences are inscribed in some readily recognizable orthography. For example, Stephen Schiffer, who has made particularly vivid use of this myth, asks that we imagine a person who is

> psychologically quite ordinary, save for the fact that he has a huge, transparent head in which two boxes are plainly visible — one marked 'Beliefs', the other 'Desires'. In these boxes are an enormous number of sentence tokens in some unfamiliar but easy-to-recognize orthography. [ibid., 33]

For one's mental state to be a belief, then, as opposed to a desire, is for it to appear in one's belief box. Now the notion that our propositional attitudes are sorted into kinds via a system of clearly labeled boxes is obviously metaphorical, and must be cashed out in more realistic terms. As Fodor suggests in the following passage, this metaphor is usually spelled out in terms of functionalism, the view that mental states are to be explained in terms of their causal roles. The notion of a 'box', then, is the notion of a functional type that a token stored in one's brain instantiates.

> [F]or present purposes at least, everybody is a functionalist, which is to say that we all hold that mental states are individuated, at least in part, by reference to their causal powers. (This is, of course, implicit in the talk about 'intention boxes' and the like: To be — metaphorically speaking — in the state of having such and such a rock in your intention box is just to be — literally speaking — in a state that is the normal cause of certain sorts of effects and/or the normal effect of certain sorts of causes.). [Fodor 1987, 138]

So to be in a state of believing that p is to be in a state that is typically correlated with certain sorts of inputs and outputs, and it is the functional role of this belief — that is, the distinctive set of inputs and outputs that tends to be associated with it — that distinguishes it from the state of desiring that p or, for example, believing that q. So, to take a simplistic example, the belief that there is ice cream in the freezer is normally caused by certain perceptual experiences (e.g., removing the ice cream from the grocery bag and placing it in the freezer, or seeing it there next to the frozen peas) and, together with the desire for ice cream, normally eventuates in one's removing the ice

cream from the freezer and eating some of it.

The functionalist's ambition is to explain propositional attitudes in purely causal terms — the content and type of a propositional attitude are to be explained in terms of the attitude's functional or causal role — and so functionalism initially seems to represent a genuine alternative to a normative conception of belief. But it's worth noting the qualifications that accompany the functionalist's generalizations about the propositional attitudes. Here, for example, is Fodor's schematic explanation of what it is to intend that p:

> [Y]ou put into the intention box a token of a mental symbol that means that P. And what the box does is, it churns and gurgles and computes and causes and the outcome is that you behave in a way that (ceteris paribus) makes it true that P. [ibid., 136]

Note the ceteris paribus clause or, in the earlier formulations of functionalist generalizations, the references to 'normal' or 'typical' inputs and outputs.[13] These qualifications serve at least two purposes. First, they represent an acknowledgment of the enormous complexity of our mental economy. For example, belief-desire pairs 'normally', in a vacuum as it were, produce actions: my desire for ice cream, in combination with my belief that there is ice cream in my freezer, 'normally' produces as output my removing the ice cream from the freezer and eating some of it. But to consider such pairs in a vacuum is to abstract from the complexity of an actual

[13] Elsewhere Fodor suggestions that such qualifications are essential to the sort of generalizations characteristic of functionalism: "If there are psychological laws, then they must be nonstrict; they must be 'ceteris paribus' or 'all else equal' laws. There couldn't, for example, be a mental state whose instantiation in a creature literally guarantees a subsequent behavior" (Fodor 1991, 21).

individual, where further circumstances may very well intervene to prevent the output that one would expect from such a pair were one to consider it in isolation: the belief-desire pair under consideration will likely not lead to my eating ice cream if (a) I desire to be thin, and believe that eating ice cream will frustrate that desire, or (b) I behave irrationally, and simply fail to take the means to my end.

The second purpose served by these qualifications is more complex. Consider, once again, a particular desire. Considered in isolation, we might expect it to yield certain outputs, but desires are of course always the desires of particular people varying in their rationality and possessing a variety of other desires, beliefs, and propositional attitudes, interacting in complex and often unexpected ways. So, in reality, a particular desire may yield a wide range of outputs. To take another example, a belief, in reality, may be the result of a great variety of inputs: my belief that there is ice cream in the freezer may be the result of my having seen ice cream there, or the result of my having seen a container of fake ice cream planted there to fool me; rather than seeing it, I may have simply been told it was there, or I may have acquired the belief as a result of a dream concerning ice cream in my freezer; or I may simply have suffered from some mental glitch that led to my having the belief. Despite the enormous range of possible inputs and outputs, functionalism identifies certain of these inputs and outputs as 'normal', as indicative of the functional nature of the propositional attitude in question. The qualifications — 'normal', 'typical', 'ceteris paribus' — are all ways of gesturing toward those causal roles that are taken as indicative of the content and nature of a propositional attitude and distinguishing them from the causal roles that are not so indicative.

But on what basis do we selectively identify some inputs or outputs as indicative of the functional nature of a particular sort of propositional attitude, as 'normal'? The answer, I think, is that functionalism relies on assumptions about the proper or 'normal' functioning of the various propositional attitudes — assumptions concerning, for example, the circumstances that ought to give rise to a belief or its abandonment, or what behavior desires (at least in conjunction with the relevant beliefs) ought to yield. Any functional accounts of belief and desire, for example, will need to make generalizations like the following:

A. "[A] belief that p tends to go out of existence in the presence of a perception with the content that not p, whereas a desire that p tends to endure, disposing the subject in that state to bring it about that p" (Smith, 115).
B. Perceptual experiences of ice cream in the freezer normally result in a belief that there is ice cream in the freezer.
C. The desire to eat ice cream, together with the belief that there is ice cream in the freezer, normally results in one's taking ice cream out of the freezer and eating it.

All of these generalizations involve assumptions about how a cognitive agent ought to be forming these states, and what she ought to be doing with them; it is from these assumptions that functionalism derives its characterizations of the propositional attitudes. To put the point another way, the account that a functionalist offers relies on considerations concerning the norms governing the propositional attitudes in characterizing their nature. But what this means, if I'm right, is that functionalism is *not* a genuine rival to a normative conception of, say, belief. For the functionalist tacitly assumes, along with the proponent of the normative conception, that the identification of propositional attitudes cannot proceed independently of normative considerations:

we cannot identify someone as having beliefs independently of considerations concerning whether that person is sensitive and responsive to, for example, truth-conducive norms governing belief.[14,15]

I want to be clear about what I'm trying to accomplish in this discussion of Stich's functionalism. I'm *not* claiming to have demonstrated that there are no rivals to a normative conception of belief. Demonstrating this would require, at a minimum, surveying the range of theories that Papineau identifies as non-naturalist — "such theories as indicator semantics. . ., success semantics. . ., teleosemantics. . ., and Fodor's 'asymmetric dependence' theory of content" (NJ, 21) — and arguing, in the way I've done with functionalism, that each theory presupposes certain normative criteria in its identification of the propositional attitudes. That is not a task that I can reasonably undertake in this context.

To see what I am doing here, let me first review some conclusions from earlier chapters. The hypothetical conception entails that all normative principles are extrinsic to cognition: the nature of cognition is one thing; the cognitive goals one adopts, and

[14] See Zangwill 1998 for an argument that functionalist generalizations are false unless regarded as normative characterizations of *rational* cognitive behavior. There is at least the following difference between functionalism and the normative conception. The account I developed in Chapter Four takes as basic the question of what it is for an individual to be a believer, to possess beliefs, while functionalism seems to focus primarily on the question of what it is for a mental state to be a belief with such-and-such content.

[15] I should point out that my claims about functionalism apply only to those versions that take commonsense or 'folk' psychology as the underlying theory from which they derive their generalizations.

the norms one therefore follows, another thing entirely. It is this consequence of the hypothetical conception that places it at odds with a normative conception of thought, for this conception holds that certain norms are intrinsic to the nature of cognition: a normative conception of belief, for example, would hold that belief involves as part of its nature a sensitivity and responsiveness to certain norms. This normative conception of thought entails the categorical conception, for if sensitivity and responsiveness to certain norms are intrinsic to thought, then the norms in question have authority over one's thought simply by virtue of one's status as a thinker; their authority is not contingent upon one's being a thinker with any particular desire. So we can look at the dispute between the hypothetical and categorical conceptions as also playing out at another level, in a dispute over whether or not there are norms intrinsic to the nature of thought.

Granted, in arguing that our own cognitive concepts are parochial and that there is a wide range of possible alternatives, Stich does not have in mind the specific sort of concepts I'm considering here — what I've referred to as normative concepts, or the normative conception of thought. But his aim in making this argument is to undermine the idea that any of our conceptions of how cognitive evaluation ought to proceed have anything more than the weight of history behind them — our preferences when it comes to cognitive norms are no different from our preferences for trousers over togas, he maintains. So our normative intuitions in this area cannot be regarded as apprehensions of the intrinsic value of the processes they sanction; to put the point another way, these intuitions cannot be regarded as apprehensions of the categorical nature of the norms that sanction these processes. This line of argument,

then, is meant to support the hypothetical conception and to undermine the categorical conception. But if the dispute between these two conceptions also plays out at another level, in the dispute between normative and non-normative conceptions of thought, then Stich cannot fully address the former dispute without also addressing the latter. Moreover, and more to the point for the present discussion, I've claimed that it is reasonable to see Stich's line of argument in favor of the hypothetical conception as involving, in essence, an argument from relativity. But I've also claimed that such arguments depend, not on the mere assertion of an abstract possibility of conceptual diversity, but on our facing genuine alternatives. Mackie's argument, for example, has some plausibility because we can, with little difficulty, cite actual instances illustrating the diversity of moral opinion. But if my argument concerning the nature of functionalism is correct, then far from demonstrating the existence of alternatives to a normative conception of belief, Stich actually presupposes such a conception. Our situation, then, is akin to that of the toga-wearer who is informed that there are alternative forms of dress without being given any specific alternatives: there may very well be alternatives, but in the absence of specific and plausible examples there is little to do but to stick with what we know. Of course, it may be that there is, in any case, a need to demonstrate that a normative conception of belief has some grounding in the nature of belief — and it is to that task that I turn next — but the SI objections leveled by Stich, Kornblith, and Railton do little to make that need seem more urgent.

5.3.1.2 Aberrant belief, perceptual belief, and action

My argument in the last section was negative: it provided little positive support for the

normative conception of belief, and instead merely tried to fend off a particular sort of objection to that conception. I now want to turn to positive considerations supporting that conception, but I should emphasize that my conclusions here are tentative. A proper defense of a normative conception of belief, never mind of such a conception of thought generally, would require considerably more argument, and considerably more time surveying the naturalist alternatives, than I can give it here. Nevertheless, I want to point to two considerations that militate in favor of a normative conception of belief — one consideration based on the difficulties confronting a non-normative conception, and a second consideration appealing to the nature of action.

In Chapter Four I distinguished two different models for how we might understand the norms involved in cognition, on the assumption that the normative conception is true. We could view such norms as instrumental to the attainment of some end, such as true belief, commitment to which would be constitutive of thought. Alternatively, we could view adherence to them as constitutive of thought independently of any end to which they are conducive, in which case we might think of them as articulating a standard or ideal of cognitive health or proper function. As I mentioned above, I am focusing here on the notion of belief, and I will rely on a version of the first model, according to which an orientation toward truth, and sensitivity and responsiveness to truth-conducive norms, is constitutive of belief.

The obvious question to ask in assessing this view is whether an orientation toward truth, and the attendant sensitivity and responsiveness to truth-conducive norms, really is constitutive of one's having beliefs. But it may be useful approaching this question from another angle: can we make sense of a view according to which this

orientation, sensitivity, and responsiveness are *not* constitutive of belief? In other words, can we make sense of Papineau's view that "sensitivity to norms of judgement is an addendum to the possession of beliefs itself"? The idea that we might give a purely causal characterization of belief encourages us to think that we can, as does the mythology of beliefs as mental sentences stored in clearly marked mental boxes. But when we turn to actual examples things begin to look more complicated.

Imagine, for example, the following scenario: one evening I decide that I'd like some ice cream and, believing that there is some in the freezer, I open the freezer door. It turns out that there is no ice cream in the freezer, but this fact does not dissuade me from my conviction that the freezer contains ice cream. This scenario certainly could be taken as illustrating a case in which belief involves no sensitivity to epistemic norms: my belief, in these circumstances, seems to be unresponsive to obvious facts concerning the way things are. So, the naturalist concludes, belief need not be responsive to epistemic norms.

But if we look more closely at the example it's less clear what conclusions it supports. One thing to say about the example is that it is underdescribed, and raises more questions than it answers: What behavior surrounds the seemingly irrational belief that there is ice cream in the freezer? Attempts to eat the imaginary ice cream, attempts to justify the belief that the ice cream is there, or something else entirely? Is this behavior with respect to ice cream in the freezer isolated, or part of a wider pattern of irrational behavior? What evidence is there that I *believe* the freezer contains ice-cream, as opposed to fervently wishing that it contained ice-cream? Regardless of how we answer these questions, though, there is a more fundamental question that

goes to the heart of the naturalist account's plausibility: assuming that what is in question here is genuinely belief, in the absence of some complicated and difficult-to-imagine justification for my belief concerning the presence of ice cream in my freezer, why should we not regard a belief like this as a sort of aberration, a delusion to be explained in terms of some psychological disorder?

If beliefs like this are regarded as aberrations, as signs of psychological disorder, what does this mean for the naturalist, non-normative conception of belief? On the one hand, it does suggest that we can make sense of isolated cases of belief that are not responsive to epistemic norms. On the other hand, however, it suggests that we can make sense of such cases only against the backdrop of a vast number of beliefs that are *not* aberrant — beliefs that are treated as responsive to such norms. This is first and foremost a logical point — there must be such a thing as 'normal' belief in comparison to which we judge certain beliefs as aberrant — but there may be room for a point about the nature of belief here as well. If we follow Davidson, to have a belief, however irrational, concerning the presence of ice cream in one's freezer, one must also have various other beliefs concerning freezers, ice cream, and related matters that *are* well-grounded in reality.[16] If all of my beliefs had as tenuous a connection to reality as

[16] See, for example, Davidson's discussion of the idea that "[b]eliefs are identified and described only within a dense pattern of beliefs" (ITI, 200). Perhaps an analogy between beliefs and maps is helpful here. To see a particular map as depicting a particular area, it must be possible to identify at least general features of the map that correspond to general features of the area it depicts. It is only against the backdrop of such general agreement that we can make sense of other features of the map as erroneous. Similarly, to understand someone as having beliefs about their circumstances, and to identify certain of their beliefs as erroneous, it must be possible to identify certain general points of

my beliefs concerning the non-existent ice-cream in my freezer, a Davidsonian might argue, it's not clear that I could be understood as having any beliefs at all. Whether we accept the Davidsonian point or not, the idea that beliefs so recalcitrant to epistemic norms should be regarded as psychologically aberrant is a double-edged sword for the naturalist: it lends some legitimacy to the idea that beliefs need not, in all cases, be responsive to epistemic norms, but at the cost of undermining the view that "sensitivity to norms is an addendum to the possession of beliefs." It might be an addendum to the possession of certain isolated beliefs, but it cannot be an addendum to the possession of beliefs in general.

contact between their beliefs and the way things are. Again, it is against the backdrop of such contact between their beliefs and reality that we identify other beliefs they hold as failing to connect properly with reality. The disanalogy between maps and beliefs, of course, is that we can still identify something as a map without being able to identify any particular area that it depicts, while, according to the normative conception of belief, it would seem that a similar inability to identify any features of reality to which a person's beliefs correspond at a certain point jeopardizes the idea that we confront a person with beliefs. The explanation for this disanalogy, on the normative conception, would be that there are certain conventions that maps typically follow, and we can often identify something as a map, independently of being able to identify what it is a map of, because it follows these conventions. But the normative conception of belief maintains that one of the identifying features of beliefs (or better: of someone's having beliefs) is a sensitivity and responsiveness to cognitive or epistemic norms on the part of the person holding them. So, at a certain point, if we cannot understand a person as oriented in their thought toward the truth concerning their circumstances — if we cannot detect a sufficient number of points of contact between what we take to be their beliefs and their environment — we cannot regard them as possessing the requisite sensitivity and responsiveness for possession of beliefs.

The naturalist view, of course, is not just that we can identify isolated instances of belief unresponsive to epistemic norms, but that belief in general need not be responsive to such norms. Just as history might have unfolded in such a way that we were wearing togas rather than trousers, we might have evolved, biologically or culturally, in such a way that our beliefs were responsive only to pragmatic norms. And presumably the view of beliefs unresponsive to epistemic norms as aberrant is to be regarded not as a product of insight into the nature of belief, but merely as a product of history having followed one course, and having bequeathed to us one set of intuitions, rather than another. But can we really make sense of the view that we might have evolved so as to be responsive in our beliefs only to pragmatic norms? Again, there is a danger here that appeals to what we can or cannot make sense of are merely rehearsing the parochial intuitions that Stich decries, but this view does raise questions to which the naturalist owes us some answers. Many of our beliefs, for example, are perceptual. Is it really plausible to think that we might have evolved so as to have no beliefs that were responsive to our perceptions of the way things are in our immediate environments? Surely this sort of responsiveness to the deliverances of perception is one basic way in which we would want to start making sense of the idea of beliefs as oriented toward the truth, but to think that beliefs might have evolved so as to lack such responsiveness is, it seems, to imagine beliefs as disconnected from the means by which we try to navigate our physical surroundings. It is hard to reconcile such imaginings with a view of belief as bound up with action. Or, leaving aside the question of perceptual beliefs, how, in general, are we to distinguish between beliefs, on the one hand, and desires, hopes, or wishes, on the other? By the lights of a

normative conception of thought, a propositional attitude that is in no way sensitive to the way things are looks very much like a desire or wish rather than a belief. But it's difficult to see how we are to distinguish between such propositional attitudes given a non-normative conception of thought according to which *none* of these states involves any orientation toward the truth, or sensitivity to truth-conducive norms.

What I've been doing in the last few paragraphs is raising questions about the plausibility of a non-normative or naturalist conception of belief, rather than arguing directly against it. But the questions I've raised point to areas in which this conception, without further development, lacks plausibility: it's unclear how this conception can make sense of perceptual beliefs, or the connection between such beliefs and action; nor is it clear how this conception can distinguish between beliefs and others propositional attitudes like desires or wishes; finally, it's unclear why we should take the non-normative conception as articulating a conception of normal belief, as opposed to aberrant or psychologically disordered belief. Of course these points don't show that a non-normative account of belief is impossible, but they do point to some of the difficulties involved in such an account.

Before turning to a consideration of the motivational objection to the normative conception, I'd like to mention briefly one more reason for skepticism concerning the non-normative conception of belief. Part of what it is to act is to have goals or ends that one seeks to realize in or by means of one's actions, and having goals involves seeking, or at least desiring, effective means to their realization. As agents, we are not, nor can we be, completely indifferent to the success of our actions; identifying and executing effective means to our ends is typically the measure of success. Agents must

therefore weigh different courses of action, different means to their ends, on the basis of how effective they will be at achieving those ends. The courses of action they select are those that, other things being equal, they believe will be effective at achieving their ends. So the selection of a course of action, a means, expresses (again, other things being equal) a belief that it will be effective in achieving an end. At least some true beliefs, then — beliefs concerning the nature of one's goal and what means will be effective in achieving it, for example — seem essential to any successful action, and therefore essential to what we seek as agents. Can we then make sense of people acting yet not caring about the truth of their beliefs? I think the answer is no, for not to care whether one's beliefs were true or not would be not to care whether the means one has selected to one's goals are effective or not.

Insofar as we are agents who seek to realize our goals in the world, therefore, it is difficult to imagine our being indifferent to the truth of our beliefs. Now does this show that a orientation to truth is essential to belief, or part of its nature? Perhaps not. It could be that sensitivity to norms is an addendum to the possession of beliefs in creatures that do not act — purely contemplative creatures — or whose beliefs are not bound up with their action (assuming we can make sense of such beings). But in creatures like us — creatures whose actions are, at least sometimes, informed by beliefs, or, more generally, creatures whose thought is often hostage to the exigencies of action — we cannot distinguish so clearly between having beliefs and being sensitive to their truth. So long as we think of beliefs in terms of the role they play in human action, and not just as psychological states with propositional content, we are pushed in the direction of a normative conception of belief.

5.3.2 Objection from motivational impotence

I want to turn now to the objection that conceptual claims are motivationally impotent. To some extent, this seems to be the objection that conceptual claims, *as opposed to substantive claims*, are motivationally impotent. This is certainly the thrust of Railton's criticism:

> Gary. . . sought answers to seemingly substantive practical and epistemic questions: 'Why do things that way?', he wanted to know. It would be surprising if we could give an answer with nothing more than a few definitions. To be genuinely responsive to the concerns expressed, constitutive arguments must capture a substantive — not merely linguistic — necessity. [Railton, 70]

Of course, if we take the objection this way, my arguments in the last section — arguments meant to establish, at least in a preliminary way, some of the substantive credentials of a normative conception of belief — may go some way toward meeting it. But it's not clear whether the objection, so conceived, is really plausible, for what it seems to presuppose is that substantive claims have some motivational force that conceptual claims lack — that we ought to be able to persuade Gary, for instance, of the authority of practical and epistemic norms, provided we rely on substantive rather than conceptual claims. But as Railton's argument later suggests, such expectations concerning our ability to persuade Gary underestimate his obstinacy in questioning these norms: later, on the assumption that we have identified norms responsiveness to which really is essential to being a believer, Gary's question is essentially, Why be a believer?[17] This question is a substantive version of the same questions Stich asks: Why should anyone care about the cognitive processes recommended by our ordinary

[17] Railton, pp. 73-4.

notions of thought? Why should anyone care about the norms associated with a commonsense conception of belief? But if Stich's questions concerning the motivational force of conceptual claims can be transposed into a substantive key, it's no longer clear what it means to say that substantive claims have a motivational efficacy that conceptual claims lack.[18]

If substantive and conceptual claims are equally susceptible to challenges to their normative authority or motivational force, where does that leave the account of the last chapter? The claim I made there was that we can understand the authority of cognitive norms as categorical on the assumption that certain cognitive concepts are normative concepts. And to describe a concept as normative, in this sense, is to say that there are normative standards built into the concept in such a way that a prerequisite for something's falling under the concept is that it demonstrate a sensitivity and responsiveness to these standards. The charge of motivational impotence, however, is that conceptual claims, or truths about the concepts in terms of which we understand something like belief, lack motivational force. The upshot of this charge, if correct, is that the attempt to understand cognitive normativity in terms of normative concepts

[18] Certainly demonstrating that a particular concept — say, that of belief — lacked any basis in reality might undermine the authority of norms bound up with the concept. If it could be shown that we lack beliefs, as they are usually understood, then any normative demands contingent upon our status as believers would lose their authority over us. But neither Stich nor Kornblith demonstrate any such thing, and it's not clear exactly what would be involved in such a demonstration. Eliminativists, of course, hold that our folk psychological concepts lack any basis in reality, but this contention is premised on the idea that if there were propositional attitudes they would be neurophysiological states — an idea which is far from unproblematic.

fails: motivational efficacy is essential to something's being a reason — reasons must be able to motivate us — and if conceptual truths lack motivational efficacy, then they cannot supply reasons, categorical or otherwise, for our thinking one way or another. Simply showing that there are reasons for doubting the motivational efficacy of substantive as well as conceptual claims doesn't clear the account of the last chapter of this charge that has been leveled against it.

Responding to this charge requires addressing the contention that conceptual claims lack motivational force. Stich, Railton, and Kornblith offer little in the way of arguments defending this contention, so rather than trying to rebut imaginary arguments, I want to sketch a view of normativity on which conceptual claims can possess both motivational efficacy and categorical normative authority. The key to this view will be the notion of a self-conception: if we understand normative concepts as tied to an individual's self-conception or self-understanding, we can make sense of how conceptual claims might have motivational efficacy.[19]

I want to begin with a passage from MacIntyre's *After Virtue*:

> For I am never able to seek for the good or exercise the virtues only *qua* individual. This is partly because what it is to live the good life concretely varies from circumstance to circumstance even when it is one and the same conception of the good life and one and the same set of virtues which are being embodied in a human life. What the good life is for a fifth-century Athenian general will not be the same as what it was for a medieval nun or a seventeenth-century farmer. But it is not just that different individuals live in different social circumstances; it is also that we all approach our own circumstances as bearers of a particular

[19] My emphasis in the remainder of this chapter on the importance of self-conceptions bears some similarities to a view defended by Christine Korsgaard in *The Sources of Normativity.* See, in particular, Lecture Three.

> social identity. I am someone's son or daughter, someone else's cousin or uncle; I am a citizen of this or that city, a member of this or that guild or profession; I belong to this clan, that tribe, this nation. Hence what is good for me has to be the good for one who inhabits these roles. As such, I inherit from the past of my family, my city, my tribe, my nation, a variety of debts, inheritances, rightful expectations and obligations. These constitute the given of my life, my moral starting point. This is in part what gives my life its own moral particularity. [MacIntyre, 220]

We can see in what MacIntyre says here the seeds of an account of certain kinds of normativity as grounded in one's social identity, in the roles one inhabits. These roles, MacIntyre explains, give rise to the rights, obligations, and duties that form "the given of my life, my moral starting point." As an account of how the norms associated with these roles gain their hold over someone — whence they derive both their motivational efficacy and their normative authority — MacIntyre's account here is incomplete. After all, simply being "someone's son or daughter", "a citizen of this or that city, a member of this or that guild or profession", or merely belonging "to this clan, that tribe, this nation" — none of these attachments, by themselves, explain how one might come to regard as authoritative the norms to which they give rise: one can belong to a family, city, guild, profession, clan, tribe, or nation in one sense while feeling, in another sense, that one does not belong, that one is alienated from both the practices and norms associated with these roles.

There is a sense of 'belonging', though, that is capable of explaining one's taking seriously the norms associated with a role — a sense more specific than simply "being a member/citizen/practitioner of x". To belong, in the relevant sense, is to identify with a particular role; it is for that role to be part of one's self-conception or self-understanding. MacIntyre, I would argue, registers the importance of such

identification in his shift to the first-person: it is not enough for these attachments to generate obligations, and to motivate me to take them seriously, that someone else is able to say of me, for example, that "He is Harold's son"; *I* must be able to express these attachments, to acknowledge them as part of my identity, my self-understanding: "I am someone's son or daughter, someone else's cousin or uncle; I am a citizen of this or that city, a member of this or that guild or profession; I belong to this clan, that tribe, this nation."

It is important to recognize that it is characteristic of the social identities or roles with which MacIntyre is concerned that there are norms internal to them. What this means is that our concepts of at least some of these roles may be normative: inhabiting a role of this sort involves, among other things, having some orientation toward the goods internal to them. One's understanding of a role and what it entails may, of course, be more or less adequate. An improper or inadequate understanding can result in one's failing to grasp fully what is involved in one's inhabiting a particular role. Consider, for example, a physician. Arguably, the concept of a physician is a normative concept: a physician is not simply someone who holds a particular degree, or possesses particular skills and knowledge, but someone who regards her training and abilities as subservient to a specific set of ends and who recognizes that the ends to which she is committed as a physician yield certain norms to which she ought to adhere. But there are varying degrees to which one might understand one's commitments as a physician, and therefore varying degrees to which one might appreciate the goods internal to the profession and the normative demands that it involves.

So to explain how one might be motivated by the norms internal to a role, we must presuppose not just an identification with that role on the part of the agent, but also an understanding of the nature of the role. Given an understanding of the role with which one identifies, however, nothing more is necessary to explain how one might see oneself as bound, and how one might be motivated, by the norms internal to that role. These norms are constitutive of the role with which one identifies, and one's understanding of that role together with one's conception of oneself as inhabiting it are sufficient to explain how one could be motivated by such norms. For an agent occupying such a role, to perceive an action as required or forbidden is simply to see it as either essential to one's understanding of who one is or inconsistent with that understanding.

This last thought requires further development. In particular, there may be a temptation at this point to object that I have left out the only thing that could truly explain one's motivation to adhere to the norms associated with a particular role, and that is a desire to continue occupying this role, or a desire to maintain a self-conception in which this role is essential. According to this objection, we should amend the final sentence of the last paragraph to read as follows: for an agent occupying such a role, to perceive an action as required or forbidden is simply to see its performance or nonperformance as essential to one's desire to maintain that role and the self-conception with which it is associated. But it is important to recognize that the appeal to desire here is superfluous. That is not to say that one *couldn't* be motivated by desire in the way described in the objection, but merely that this sort of appeal to desire isn't essential to understanding how one might be motivated in such circumstances.

Consider the mother of an infant who is not yet sleeping through the night. Awoken for the third time in a single night by a hungry newborn, she likely has little desire to get up yet again and feed her child, and the question of whether she desires to maintain a self-conception of herself as a parent, or even whether she desires to remain a parent, are either far from her mind or, if present in her thoughts at all, beside the point. The consideration that compels her to leave her bed and feed her child is that she *is* a parent — regardless of what she might desire — with responsibilities to her child; it is her conception of who she is that prevents her from regarding as a serious option staying in bed and ignoring her child's cries. It is through her identification with this role, and her understanding of what it entails, that she is sensitive to a range of considerations as placing demands on her action — considerations to which she would not be sensitive had she not identified with this role.

What I am proposing is a framework for thinking about normativity within which we can make sense of motivation that is (a) partly grounded in conceptual truths, and (b) not contingent upon desires. The elements of this framework include the notion of a normative concept characterizing a role (perhaps also an activity or practice), together with someone's understanding and identifying with that role. My suggestion is that if we can make sense of someone's understanding and identifying with a role that is characterized in terms of a normative concept — a role, therefore, that involves some minimal level of sensitivity and responsiveness to certain norms — we can make sense of someone's being motivated by considerations that satisfy both (a) and (b). These considerations are grounded in conceptual truths, thereby satisfying (a), in the following sense: part of what it is to identify with a role or to conceive of

oneself in terms of it is to have at least a minimal grasp of truths concerning how that role is to be understood — a grasp, in other words, of conceptual claims concerning that role. This conceptual understanding plays a role in explaining one's actions within that role: I am a father (identification); fathers do x and refrain from doing y (a grasp of conceptual claims concerning fatherhood); so I must do x and refrain from doing y.

Moreover, one's motivation in such circumstances is not conditional upon the presence of desires, thereby satisfying (b). One's taking seriously the norms internal to such roles is not contingent upon one's wanting to occupy the role: it is not because I want to be a father that I take seriously my responsibilities, but because I *am* a father. Recognizing that one is a father, or a doctor, or a military officer — regardless of whether one has any thoughts concerning whether the continuation of this identification would be welcome — is at least sometimes sufficient to explain why one regards the norms associated with the role as authoritative. Nor need we appeal to desires to explain individual actions undertaken as part of a role. As Samuel Scheffler puts it, discussing the way in which internalization of an ideal can serve as a desire-independent source of motivation,

> An action may be performed or rejected because it is perceived as required or excluded by considerations whose salience for the individual is explained by his or her internalization of the ideal. I do not mean that, case by case, one is motivated by the desire to live up to the ideal. The idea is not that one reasons: "I want to be a certain kind of person, to achieve a certain ideal. Therefore I must treat considerations of this type as decisive. These considerations militate in favor of act X and against act Y. So I will perform act X." Rather, the internalization of the ideal supplies the psychological background against which considerations of certain types are perceived as in themselves providing reasons for action. In other words, part of what it *is* to internalize an authoritative ideal is to come to perceive considerations of some kinds as providing reasons for action, and considerations of other kinds as having no deliberative weight whatsoever.

> Accordingly, for one who has internalized such an ideal, certain sorts of action present themselves as mandatory, while others are viewed as simply out of the question, as not representing serious options. [Scheffler, 89-90].

Scheffler distinguishes between what he terms 'desire-based motivation' and 'authoritative motivation', and argues that we can make sense of the latter along Freudian lines, in terms of the internalization of an ideal. The point I am making here is similar to Scheffler's: just as internalization of an ideal can provide "the psychological background against which considerations of certain types are perceived as in themselves providing reasons for action", it is against the backdrop of one's acceptance of a particular role that actions demanded by the role can "present themselves as mandatory", in spite of what one might otherwise desire.

My claim is not that the demands placed on someone occupying these roles will always serve as a source of desire-independent motivation; we can imagine, for example, someone disillusioned with the practice of medicine who is able to continue as a physician only out of a desire for the material benefits of the profession. But if is possible for someone to be motivated by role-related norms, independently of what they happen to desire, that is enough to prove my point: contrary to what Stich, Kornblith, and Papineau argue, we can make sense of categorical norms, and contrary to what Stich, Kornblith, and Railton maintain, we can understand their authority and motivational efficacy, at least in part, in terms of the concepts with which such norms are associated. It is possible to dismiss conceptual truths as lacking motivational efficacy only by ignoring the powerful influence that self-conceptions exert over our behavior.

Chapter 6: Conclusion

What I have tried to do in the dissertation is to develop a framework within which we can make sense of certain cognitive norms as categorical. Central to this project is the following idea: if there are norms constitutive of thought — if, that is, the dual role thesis is true — then these norms will have categorical rather than hypothetical authority over thought. The dual role thesis holds that there are principles playing both normative and constitutive roles in thought, and I have argued that the key to understanding this thesis is the notion of a normative concept. If we regard certain cognitive concepts as normative — the concept of belief, for example — then falling under that concept — having beliefs, being a believer — involves orientation toward a standard or ideal embodied in the concept: in the case of belief, I have suggested, this orientation will involve sensitivity and responsiveness to truth-conducive principles. Such principles would be normative in the sense that they specify constraints on what we ought and ought not to believe, and constitutive in the sense that one must demonstrate a minimal level of sensitivity and responsiveness to them in order to be a believer. Because they are constitutive of belief, these norms have authority over believers not by virtue of desires they may or may not have, but simply by virtue of their status as believers. It's possible to see the argument I have constructed as involving two steps: from the notion of a normative concept to the dual role thesis, and from the dual role thesis to the categorical conception. And my project here has involved clarifying and defending these steps, as well as defending the argument as a

whole from objections advanced against the categorical conception by its opponents.

At the outset of the dissertation I introduced this project against the backdrop of a widespread sense that certain conceptions of normativity — in particular, a conception according to which there are norms with intrinsic, objective, or categorical authority — are inconsistent with naturalism. In an effort to identify a theme common to many of the views described as naturalistic, I characterized naturalism as the view that science is our only reliable guide to what the world is like. Now of course this characterization, as broad as it is, encompasses a wide variety of views — views differing, for example, in what they count as science and in how they characterize the methods appropriate to it. Perhaps some of these views truly are inconsistent with a categorical conception of normativity, but — and this explains why I haven't spent time trying to provide a more careful characterization of naturalism — if the views I discussed in Chapter Two are any indication, the resistance typically displayed by naturalists toward the idea of categorical normativity owes little to a commitment to the methods or findings of science. Instead, what seems to drive this resistance is the view that there is no plausible account of how categorical norms or reasons might engage with our motivational capacities. And fueling this view, in turn, is a picture of desires as essential to human motivation — a picture within which the very idea of categorical or desire-independent motivation is not merely mysterious but unintelligible.

There is nothing especially naturalistic about this picture of motivation, at least

so long as we confine ourselves to naturalism as I've defined it.[1] Whether or not one holds that supernatural entities and explanations have some legitimacy in our understanding of things seems to have little bearing on whether one accepts a view of human motivation in which desires are essential. And the conception of categorical normativity that I developed in the preceding chapters should raise no red flags for the naturalist who is willing to countenance beliefs and normative concepts of the sort I described in Chapter Four, and who isn't independently predisposed to reject the idea that our self-conceptions can serve as a desire-independent source of motivation.

So the real obstacle to the categorical conception is not naturalism per se, but a theory of motivation, perhaps often held by naturalists, according to which desires are essential to motivation. In the preceding chapter I presented an alternative to this theory according to which normative concepts of the sort I discussed in Chapter Four are capable of motivating us. If we accept internalism, which I described as the view that it is characteristic of reasons that they are capable of motivating us, then any attempt to defend the status of certain reasons as categorical must also show how such reasons are capable of motivating us. My efforts in Chapter Five to defend the motivational force of conceptual claims are essential, then, to my defense of the categorical conception, given that I've tried to make sense of that conception in terms of the idea that certain key cognitive concepts, such as that of belief, are normative.

There are, however, several issues that remain to be addressed in developing

[1] There is a historical connection between this picture and naturalism: we're indebted for the account of motivation to David Hume, who aspired to be a "Newton of the mind". For an account of Hume's naturalism, see Stroud 1977.

the account of cognitive normativity that I've begun here, and I want to indicate briefly three of the more important of these outstanding issues and where my responses to them might lead.

1. Universality

Two of the three issues concern the extent to which the view I've defended here really deserves to be called a defense of a *categorical* conception of cognitive normativity. What might occasion worries on this score is this: the sense in which certain norms can be described as categorical, given the views I have defended here, is very specific, and does not capture everything that is traditionally associated with the notion of the categorical. On the account I am defending, to say that a reason or a norm is categorical is to say that it is capable of motivating someone independently of their desires. But this is not to say that it is capable of motivating anyone, conditionally only upon their rationality: its motivational efficacy is contingent upon one's identification with a certain role and one's understanding of that role as making certain demands on one's behavior.

One important respect in which this account of the categorical is limited in comparison to traditional accounts is that it does not, at least as I have presented it here, secure for categorical reasons or norms the sort of universality that Kant sought. The problem here, if it is a problem, is due to the way in which I've tied categorical normativity to the concepts in terms of which we understand cognition: concepts can differ from person to person, both in the sense that some people may have concepts

that others lack, and in the sense that understandings of a single concept may diverge.[2] What the dual role thesis promises — the idea that there are certain principles any thinker ought to heed because they are essential to one's status as a thinker — my emphasis on concepts in developing that thesis might be said to take away, since concepts (including, of course, normative concepts and self-conceptions) can differ in the ways I've noted.

This problem is a product, at least in part, of the way I have conceived of my project here: my primary aim has been to secure space for a view of certain cognitive norms as categorical. The outstanding issue here is determining whether there are any norms that fill that space. I have suggested that there are reasons for favoring a normative concept of belief involving orientation toward truth, but this falls far short of demonstrating the universality either of such a concept or of the norms associated with it. And considering only the latter shortcoming, there are various views concerning what counts as an orientation toward truth, and therefore what norms count as truth-conducive.

Addressing this problem requires demonstrating that there are some self-conceptions, yielding determinate norms, that are indispensable for cognitive agents. This may seem like a tall order, but let me sketch one possible route toward such a conclusion. Assume, for the sake of argument, that some normative conception of belief must be correct, even if it is not the truth-oriented one I've specified here. In other words, a minimal level of sensitivity and responsiveness toward *some* set of

[2] Both of these points raise the issue, which I'm not prepared to address here, of how we individuate concepts.

principles is constitutive of belief. Now this assumption may not seem to get us very far — after all, it seems likely that there is a wide array of possible concepts of belief, even if we restrict ourselves to normative concepts — but I think we may be able to get more mileage out of it than it seems at first glance. The key is unpacking what is involved in 'sensitivity and responsiveness'. Responsiveness is measured, in part, in terms of consistency between one's actions and the principles or norms to which one is committed. Sensitivity involves attentiveness to what these principles or norms require, as well as an awareness of whether one is being responsive to them — attentiveness and awareness that are both directed toward the goal of forestalling or correcting errors of responsiveness. To say that a believer must demonstrate such sensitivity and responsiveness, therefore, is to say that a believer must show some concern for the consistency of her actions with the normative principles to which she is committed. Any believer, therefore, must be concerned with consistency. Now it is not obvious that the sort of consistency at issue here — consistency between actions and principles — is the same as logical consistency, but if it is, or if there is a sufficiently strong connection between the two notions, then there is reason to think that commitment to certain fundamental logical principles or norms of rationality — at the very least, principles or norms concerning consistency — is constitutive of one's status as a believer.

The argument I've just sketched suggests the following: if we have a normative conception of belief at all, then it must be a conception central to which are at least certain basic logical principles or norms of rationality. This argument, if successful at

all, is dependent for its success on two assumptions: first, that the proper concept of belief is a normative one, and second, that there is a sufficiently strong connection between logical consistency and consistency between actions and principles. But if these two assumptions are defensible, then this argument also shows how we might begin to realize the universalist ambitions of a traditional view of categorical normativity.

2. *Motivational force versus normative authority*

There is a further respect in which the account of the categorical that I defend here is limited in comparison to traditional conceptions of the categorical. What I have tried to do here, particularly in Chapter Five, is to explain how it is possible for us to be motivated to take seriously principles that do not appeal to antecedently-held desires. But it could be argued that this is only part of what an account of categorical normativity should explain. What seems characteristic of an unqualifiedly categorical account is that it explains not only why we *happen* to be motivated to take seriously certain norms, which is what I've done here, but also why we *ought* to be so motivated. If we take Kant's views as a model, for example, his account of the categorical nature of moral imperatives is clearly meant to address the latter issue by showing that moral principles are derived principles of rationality, and therefore share the latter's authority. To put the complaint about the view I've defended here another way, while it is clearly important to show how categorical norms might fit into our motivational economy — how they might have motivational force for us — accomplishing that task by itself leaves us with a very anemic conception of the categorical, absent some

additional account showing why we ought to take such norms as authoritative for our thought.

It might even be argued that the absence of such an account provides another opening for the Stichean rejoinder that the only way of demonstrating the normative authority of certain epistemic or cognitive principles is to show that adhering to them is likely to be conducive to the satisfaction of our desires: if (a) cognitive norms derive their authority from the normative concepts with which they are associated, (b) there are various such concepts, and (c) there is no way of demonstrating the intrinsic superiority of one concept over another, then it may seem that we're right back in the situation that Stich describes, where (d) we must determine which normative concepts, and therefore which cognitive norms, to accept as authoritative on the basis of extrinsic factors.

But if, on the other hand, there are certain principles or norms that are essential to any plausible conception of cognition, Stich's rejoinder will be blunted: in keeping with (b), there may be some level at which we could speak of a variety of competing conceptions of cognition, but these conceptions will share a core of assumptions about the nature of cognition and the principles to which it is responsible. That there is such a core is precisely what I tried to suggest above in my argument responding to the complaint that my account does not realize the universalist ambitions of traditional accounts of the categorical: we have no choice but to take certain norms seriously, because they are implicated in every area of thought, and in every plausible conception of what counts as thought. If we look at the Stichean rejoinder as a version of an

argument from relativism, as I suggested we might understand Stich's overall strategy in Chapter Five,[3] then pointing to such a core is a way of seeking to undermine the premise asserting the diversity of possible cognitive concepts that allows the argument from relativism to get off the ground. It is also, however, a way of showing why we *ought* to take certain cognitive norms seriously: if we can't help but understand cognition in certain terms, then if we are to take seriously our status as thinkers at all, we must take seriously the terms in which we understand that status.

3. A cognitive self-conception?

Let me make one final point, concerning the account of motivation that I presented in the last chapter. To see that account as applicable to the case of cognition, we must make two assumptions: first, that we have a self-conception as thinkers or believers, and second, that we have some understanding of what this identity entails. These assumptions do not seem obviously false — after all, we seem to have a capacity for second-order reflection on our beliefs, which suggests that we must have an awareness of ourselves as having beliefs, as believers — but it is not clear to me at this point how to argue for them.

As I've made clear, there are issues that remain to be settled in the account of cognitive normativity that I have developed here, but I hope to have made a promising start in developing an account of cognitive normativity as categorical. I hope, in other words,

[3] See Chapter Five, footnote 5, and the text to which it is attached.

to have gone some way toward addressing Papineau's concern that "it is surely desirable that we should have some kind of understanding of the peculiar force that judgemental norms are supposed to exert on us" (Papineau, 28). I do not think, however, that the interest or importance of my argument in the dissertation is limited to the question of how we conceive of cognitive normativity, so before closing let me point to two ways in which the claims I defend here may have some broader significance.

First, there is nothing in my account of cognitive normativity that would prevent the extension of this account into other normative arenas: if there is any plausibility to the sort of constitutive account of the nature of cognitive normativity that I offer here, it may be that a similar account could be employed to explain the categorical character of norms in other areas.[4]

The second point is somewhat more complicated and, at this point, more speculative. The notion of a normative concept is central to my effort to show that the dual role thesis is coherent. And in Chapter Five, I argued for a normative concept of belief, according to which the possession of beliefs requires sensitivity and responsiveness to certain norms. Essential to the possession of beliefs, or to one's status as a believer, on this view, is engagement in a certain sort of activity — an activity demonstrating this requisite sensitivity and responsiveness. Now there is a widely held view in contemporary philosophy that beliefs are states of the brain, and

[4] Christine Korsgaard could be understood as employing just such an account in her 2002 Locke Lectures, in which she argues that commitment to certain normative standards is constitutive of agency.

what I want to suggest is that the normative concept of belief that I defend may have consequences for this widely held view. If having a belief requires sensitivity and responsiveness to certain norms, then the idea that beliefs are brain states cannot be correct, or at any rate does not tell the whole story about beliefs and their nature. For brains and their states, I would argue, do not exhibit or fail to exhibit such sensitivity and responsiveness; whole persons do.[5]

[5] The view I am suggesting here is similar to one defended in Baker 1995. See Chapter Six.

Bibliography

I. Abbreviated Works

References to the following works by Davidson, Frege, Papineau, and Stich are abbreviated in the text.

A. Works by Davidson

EAE *Essays on Actions and Events*. Oxford: Clarendon Press, 1980.

POI "Paradoxes of Irrationality", in *Philosophical Essays on Freud*. Edited by Richard Wollheim and James Hopkins. New York: Cambridge University Press, 1982.

RA "Rational Animals." *Dialectica* 36 (1982): 317-27.

FPA "First Person Authority." *Dialectica* 38 (1984): 101-11.

ITI *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984.

KOM "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (1987): 441-58.

EE "Epistemology Externalized." *Dialectica* 45 (1991): 191-202.

CSR "Could There Be a Science of Rationality?" *International Journal of Philosophical Studies* 3 (1995): 1-16.

SIO "Subjective, Intersubjective, Objective", in *Current Issues in Idealism*. Edited by Paul Coates and Daniel D. Hutto. Bristol, England: Thoemmes, 1996.

SIO2 *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press, 2001.

B. Works by Frege

TPW *Translations from the Philosophical Writings of Gottlob Frege*. Edited by Peter Geach and Max Black. New York: Philosophical Library, 1952.

FA *The Foundations of Arithmetic*. Translated by J. L. Austin. Evanston, Illinois: Northwestern University Press, 1959.

BLA *Basic Laws of Arithmetic*. Translated by M. Furth. Berkeley: University of California Press, 1967.

B *Begriffsschrift, a Formula Language, Modeled upon that of Arithmetic, for Pure Thought*. In *Frege and Godel: Two Fundamental Texts in Mathematical Logic*. Edited by Jean van Heijenoort. Cambridge, Mass.: Harvard University Press, 1970.

PW *Posthumous Writings*. Translated by Peter Long and Rober White. Oxford: Basil Blackwell, 1979.

PMC *Philosophical and Mathematical Correspondence*. Translated by Hans Kaal. Oxford: Basil Blackwell, 1980.

CP *Collected Papers on Mathematics, Logic, and Philosophy*. Translated by Max Black. Oxford: Basil Blackwell, 1984.

C. Works by Papineau

NJ Papineau, David. "Normativity and Judgement." *The Aristotelian Society,* Supplementary Volume LXXIII (1999): 17-43.

D. Works by Stich

FR Stich, Stephen. *The Fragmentation of Reason*. Cambridge: The MIT Press, 1990.

II. Other Works

Aach, John. "Psychologism Reconsidered: A Re-evaluation of the Arguments of Frege and Husserl." *Synthese* 85 (1990): 315-38.

Allen, Colin and Marc Bekoff. *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. Cambridge, Mass.: MIT Press, 1997.

Allison, Henry. *Kant's Theory of Freedom.* New York: Cambridge University Press, 1990.

Baker, Lynn Rudder. *Saving Belief: A Critique of Physicalism*. Princeton, New Jersey: Princeton University Press, 1988.

________. *Explaining Attitudes: A Practical Approach to the Mind.* New York: Cambridge University Press, 1995.

Barbosa de Oliveira, Marcos. "Logic and Cognitive Science: Frege's Anti-Psychologism." *Manuscrito* 17 (1994): 65-96.

Bealer, George. "The Incoherence of Empiricism", in *Naturalism: A Critical Appraisal*. Edited by Steven J. Wagner and Richard Warner. Notre Dame: University of Notre Dame Press, 1993.

Beck, Lewis White. *A Commentary on Kant's Critique of Practical Reason.* Chicago, Illinois: University of Chicago Press, 1960.

________. "A Prussian Hume and a Scottish Kant", in *Essays on Kant and Hume*. New Haven, Connecticut: Yale University Press, 1978.

Bell, David. "Transcendental Arguments and Non-Naturalistic Anti-Realism", in *Transcendental Arguments: Problems and Prospects*. Oxford: Clarendon Press, 1999.

Bermudez, José Luis. "Psychologism and Psychology." *Inquiry* 42 (1999): 487-504.

Boghossian, Paul A. "The Rule-Following Considerations." *Mind* 98 (1989): 507-49.

Brockhaus, Richard. "Realism and Psychologism in 19th Century Logic." *Philosophy and Phenomenological Research* 51 (1991): 493-524.

Brueckner, Anthony. "Conceptual Relativism." *Pacific Philosophical Quarterly* 79 (1998): 295-301.

Burge, Tyler. "Intellectual Norms and Foundations of Mind." *The Journal of Philosophy* 83 (1986): 697-720.

Cerbone, David R. "Don't Think, But Look: Imaginary Scenarios in Wittgenstein's Later Philosophy." *Inquiry* 37 (1994): 159-83.

________. "How to Do Things With Wood: Wittgenstein, Frege and the Problem of Illogical Thought", in *The New Wittgenstein*. New York: Routledge, 2000.

Cherniak, Christopher. *Minimal Rationality*. Cambridge, Massachusetts: The MIT Press, 1986.

The Church Hymnal Corporation. *The Hymnal 1982, according to the use of the Episcopal Church*. New York: The Church Hymnal Corporation, 1985.

Cohen, Jonathan. "Frege and Psychologism." *Philosophical Papers* 27 (1998): 45-67.

Conant, James. "The Search for Logically Alien Thought: Descartes, Kant, Frege, and the *Tractatus*." *Philosophical Topics* 20 (1992): 115-80.

Conee, Earl. "Normative epistemology", in E. Craig, ed., *Routledge Encyclopedia of Philosophy*. London: Routledge, 1998. Retrieved July 01, 2006, from http://0-www.rep.routledge.com.library.acaweb.org:80/article/P054.

Currie, Gregory. "Frege and Popper: Two Critics of Psychologism", in *Imre Lakatos and Theories of Scientific Change*. Boston: Kluwer, 1989.

Erdmann, Benno. *Logik*. Halle: Max Niemayer, 1892.

Evnine, Simon. *Donald Davidson*. Stanford: Stanford University Press, 1991.

Fodor. *Psychosemantics*. Cambridge, Mass.: MIT Press, 1987.

________. "You Can Fool Some of the People All of the Time, Everything Else Being Equal: Hedged Laws and Psychological Explanations." *Mind* 100 (January 1991): 19-34.

Forster, Michael N. "On the Very Idea of Denying the Existence of Radically Different Conceptual Schemes." *Inquiry* 41 (1998): 133-85.

Gambel, E. H. "The Normativity of Meaning." *Philosophical Studies* 86 (1997): 221-42.

Gardiner, Patrick. "German Philosophy and the Rise of Relativism." *Monist* 64 (1981): 138-54.

George, Alexander. "On Washing the Fur Without Wetting It: Quine, Carnap, and Analyticity." *Mind* 109 (2000): 1-24.

Gibbard, Alan. "Meaning and Normativity." *Philosophical Issues* 5 (1994): 95-115.

Goldfarb, Warren. "Frege's Conception of Logic", in Floyd and Shieh, eds., *Future Pasts*. New York: Oxford University Press, 2001.

Guignon, Charles. Introduction to *The Grand Inquisitor*, by Fyodor Dostoevsky. Indianapolis: Hackett Publishing Company, 1993.

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

Hare, R. M. *The Language of Morals*. Oxford: Clarendon Press, 1952.

Hesse, Mary. "The Strong Thesis of Sociology of Science", in *Revolutions and Reconstructions in the Philosophy of Science*. Bloomington: Indiana University Press, 1980.

Hill. Thomas E. "The Hypothetical Imperative", in *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca, New York: Cornell University Press, 1992.

Hollis, Martin and Steven Lukes. *Rationality and Relativism.* Cambridge, Massachusetts: The MIT Press, 1982.

Horgan, Terence. "Nonreductive Materialism and the Explanatory Autonomy of Psychology", in *Naturalism: A Critical Appraisal*. Edited by Steven J. Wagner and Richard Warner. Notre Dame: University of Notre Dame Press, 1993.

Hume, David. *A Treatise of Human Nature*. New York: Oxford University Press, 1990.

Husserl, Edmund. *Logical Investigations*, trans. by J. N. Findlay. London: Routledge & Kegan Paul, 1970.

Hutto, Daniel D. "Was the Later Wittgenstein a Transcendental Idealist?", in *Current Issues in Idealism*. Edited by Paul Coates and Daniel D. Hutto. Bristol, England: Thoemmes, 1996.

Kant, Immanuel. *Critique of Pure Reason*, trans. by Norman Kemp Smith. New York: Macmillan, 1929.

________. *Logic*, trans. By Robert S. Hartman and Wolfgang Schwarz. New York: Dover, 1974.

________. *Groundwork of the Metaphysics of Morals*, trans. by Mary Gregor. New York: Cambridge University Press, 1997.

Kim, Jaegwon. "Psychophysical Laws", in *Actions and Events*. Edited by Ernest LePore and Brian P. McLaughlin. New York: Basil Blackwell, 1985.

________. "What is 'Naturalized Epistemology'?" *Philosophical Perspectives* 2 (1988): 381-405.

Kitcher, Patricia. "Changing the Name of the Game: Kant's Cognitivism vs Hume's Psychologism." *Philosophical Topics* 19 (1991): 201-36.

Kitcher, Phillip. "The Naturalists Return." *The Philosophical Review* 101 (1992): 53-114.

Kornblith, Hilary. "The Laws of Thought." *Philosophy and Phenomenological Research* 52 (1992): 895-911.

________. "Epistemic Normativity." *Synthese* 94 (1993): 357-76.

________. "Naturalism: Both Metaphysical and Epistemological." *Midwest Studies in Philosophy* XIX (1994): 39-52.

Kornblith, Hilary, editor. *Naturalizing Epistemology.* Cambridge: MIT Press, 1994.

Korsgaard, Christine. "Skepticism about Practical Reason." *The Journal of Philosophy* 83 (1986): 5-25.

________. "Kant's Analysis of Obligation: the Argument of *Foundations* I." *Monist* 72 (1989): 311-40.

________. *The Sources of Normativity*. New York: Cambridge University Press, 1996.

________. *Self-Constitution: Agency, Identity, and Integrity.* 2002 Locke Lectures. Retrieved July 1, 2006, from http://www.people.fas.harvard.edu/~korsgaar /index.html.

Krausz, Michael, editor. *Relativism: Interpretation and Confrontation*. Notre Dame: University of Notre Dame Press, 1989.

Krausz, Michael and Jack W. Meiland, editors. *Relativism: Cognitive and Moral*. Notre Dame: University of Notre Dame Press, 1982.

Kraut, Robert. "The Third Dogma", in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell, 1986.

Kripke, Saul. *Wittgenstein on Rules and Private Language.* Cambridge: Harvard University Press, 1982.

Luper, Steven. "Naturalized Epistemology." *Routledge Encyclopedia of Philosophy*. New York: Routledge, 1998.

Luper, Steven. "Naturalized epistemology", in E. Craig, ed., *Routledge Encyclopedia of Philosophy*. London: Routledge, 1998. Retrieved July 01, 2006, from http://0-www.rep.routledge.com.library.acaweb.org:80/article/P033.

MacIntyre, Alasdair. *After Virtue: A Study in Moral Theory*. Notre Dame: University of Notre Dame Press, 1981.

Mackie, J. L. *Ethics Inventing Right and Wrong*. New York: Penguin, 1977.

Maker, Williams. "Davidson's Transcendental Arguments." *Philosophy and Phenomenological Research* 51 (1991): 345-60.

Marsonet, Michele. "Linguistic Idealism in Analytic Philosophy of the Twentieth Century", in *Current Issues in Idealism*. Edited by Paul Coates and Daniel D. Hutto. Bristol, England: Thoemmes, 1996.

McDowell, J. "Functionalism and Anomalous Monism", in *Actions and Events*. Edited by Ernest LePore and Brian P. McLaughlin. New York: Basil Blackwell, 1985.

________. *Mind and World*. Cambridge: Harvard University Press, 1994.

McLaughlin, Brian P. "Anomalous Monism and the Irreducibility of the Mental", in *Actions and Events*. Edited by Ernest LePore and Brian P. McLaughlin. New York: Basil Blackwell, 1985.

Meiland, Jack W. "Psychologism in Logic: Husserl's Critique." *Inquiry* 19 (1976): 325-39.

Mill, John Stuart. *A System of Logic, Ratiocinative and Deductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green, & Co., 1884.

Miscevic, Nenad. "Should Reason Be Fragmented?" *International Studies in the Philosophy of Science* 10 (1996): 23-36.

Mohanty, J. N. *Husserl and Frege*. Bloomington, Indiana: Indiana University Press, 1982.

Nagel, Thomas. *The Last Word*. New York: Oxford University Press, 1997.

Nunn, Remmel T. "Psychologism, Functionalism, and the Modal Status of Logical Laws: A Reply to J. W. Meiland's Objections to Psychologism." *Inquiry* 22 (1979): 343-9.

Picardi, Eva. "Sigwart, Husserl and Frege on Truth and Logic, or Is Psychologism Still a Threat?" *European Journal of Philosophy* 5 (1997): 162-82.

Putnam, Hilary. *Meaning and the Moral Sciences*. Boston: Routledge & Kegan Paul, 1978.

________. "Truth and Convention: On Davidson's Refutation of Conceptual Relativism", in *Relativism: Interpretation and Confrontation*. Edited by Michael Krausz. Notre Dame: University of Notre Dame Press, 1989.

________. "Rethinking Mathematical Necessity", in *Words and Life*. Cambridge: Harvard University Press, 1994.

Quine, W.V. "Epistemology Naturalized," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 1969.

________. "Reply to Morton White," in *The Philosophy of W. V. Quine*. Edited by Lewis Hahn and Paul Schilpp. La Salle, Ill.: Open Court, 1986.

________. *Pursuit of Truth*. Cambridge: Harvard University Press, 1990.

________. "Commensurability and the Alien Mind." *Common Knowledge* 1 (1992): 1-2.

Railton, Peter. "On the Hypothetical and Non-Hypothetical", in *Ethics and Practical Reason.* Edited by Garrett Cullity and Berys Gaut. New York: Oxford University Press, 1997.

Ramberg, Bjørn. *Donald Davidson's Philosophy of Language: An Introduction*. New York: Basil Blackwell, 1989.

Ricketts, Thomas. "Frege, the *Tractatus*, and the Logocentric Predicament." *Noûs* 15 (1985): 3-15.

________. "Generality, Meaning and Sense in Frege." *Pacific Philosophical Quarterly* 67 (1986): 172-95.

________. "Objectivity and Objecthood: Frege's Metaphysics of Judgment," in *Frege Synthesized: Essays on the Philosophical and Foundation Work of Gottlob Frege*. Edited by Leila Haaparanta and Jaako Hintikka. Boston: Kluwer, 1986.

Root, Michael. "Davidson and Social Science", in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Cambridge: Blackwell, 1986.

Rosenberg, Alex. "A Field Guide to Recent Species of Naturalism." *The British Journal for the Philosophy of Science* 47 (1996): 1-29.

Scanlon, T. M. *What We Owe to Each Other*. Cambridge, Massachusetts: Harvard University Press, 1998.

Scheffler, Samuel. *Human Morality*. New York: Oxford University Press, 1992.

Schnädelbach, Herbert. *Philosophy in Germany 1831-1933*. New York: Cambridge University Press, 1984.

Sellars, Wilfrid. *Empiricism and the Philosophy of Mind*. Cambridge: Harvard University Press, 1997.

Sheffer, Harry. "Review of *Principia Mathematica*." *Isis* 8 (1926): 226-31.

Siegel, Harvey. "Naturalism, Instrumental Rationality, and the Normativity of Epistemology." *Protosoziologie* 8/9 (1996): 97-110.

Sluga, Hans. *Gottlob Frege*. Boston: Routledge & Kegan Paul, 1980.

Smith, Michael. "The Humean Theory of Motivation." *Mind* 96 (1987): 36-61.

________. *The Moral Problem*. Cambridge: Blackwell Publishers, 1994.

Sober, Elliott. "Psychologism." *Journal for the Theory of Social Behaviour* 8 (1978): 165-91.

Stich, Stephen. "Relativism, Rationality, and the Limits of Intentional Description." *Pacific Philosophical Quarterly* 65 (1984): 211-35.

________. "Evaluating Cognitive Strategies: A Reply to Cohen, Goldman, Harman, and Lycan." *Philosophy and Phenomenological Research* 51 (1991): 207-213.

Stroud, Barry. "Transcendental Arguments." *Journal of Philosophy* 65 (1968): 242-56.

________. *Hume*. London: Routledge & Kegan Paul, 1977.

________. "Evolution and the Necessities of Thought", in *Meaning, Understanding, and Practice: Philosophical Essays*. New York: Oxford University Press, 2000.

Tanney, Julia. "Normativity and Thought." *The Aristotelian Society*, Supplementary Volume LXXIII, 1999: 45-61.

Thompson, Michael. "The Representation of Life", in *Virtues and Reasons : Philippa Foot and Moral Theory.* Edited by Rosalind Hursthouse, Gavin Lawrence, Warren Quinn. New York: Oxford University Press, 1995.

________. "Apprehending Human Form", in *Modern Moral Philosophy.* Edited by Anthony O'Hear. New York: Cambridge University Press, 2004.

Toulmin, Stephen. *Human Understanding*, Volume I. Princeton: Princeton University Press, 1972.

________. *Cosmopolis: The Hidden Agenda of Modernity*. Chicago: University of Chicago Press, 1990.

Velleman, J. David. "The Guise of the Good." *Nous* 26 (March 1992): 3-26.

Viale, Riccardo, Daniel Andler, and Lawrence Hirschfeld, eds. *Biological and Cultural Bases of Human Inference*. London: Lawrence Erlbaum Associates, 2006.

Wedgewood, Ralph. "The A Priori Rules of Rationality." *Philosophy and Phenomenological Research* 59 (1999): 113-32.

Wilson, Bryan R., ed. *Rationality*. Oxford: Basil Blackwell, 1970.

Wittgenstein, Ludwig. *Remarks on the Foundations of Mathematics*. Cambridge: MIT Press, 1978.

Wright, Georg Henrik von. *Norm and Action: a Logical Enquiry.* Routledge & Kegan Paul: London, 1963.

Zangwill, Nick. "Direction of Fit and Normative Functionalism." *Philosophical Studies* 91 (1998): 173-203.