Title
Thoughts Without Masters: Incomplete Understanding and the Content of Mind

Journal
UCLA School of Law Public Law & Legal Theory Series

Author
Greenberg, Mark

Publication Date
2010-12-18

THOUGHTS WITHOUT MASTERS: INCOMPLETE UNDERSTANDING AND THE CONTENT OF MIND

Mark Greenberg
Jesus College, Oxford
D.Phil., Trinity Term 2000

ABSTRACT

Thoughts Without Masters: Incomplete Understanding and the Content of Mind

Mark Greenberg
Jesus College, Oxford
D.Phil., Trinity Term 2000

When we explain Bruno's taking the train by attributing to him the belief that safety is more important than price, we attribute to him a thought involving the concept of safety, as opposed to (say) the concept of comfort. Most leading theories of mental content presuppose what I call *the achievement view* – that for a thought to involve a given concept is, roughly speaking, for the thinker to have a disposition to satisfy a condition associated with the concept, such as a disposition to make certain inferences. I show that it follows from the achievement view that a thinker must *fully grasp* (in a sense I make precise) a concept in order to have attitudes involving it (though most theorists would allow an exception for cases in which the thinker defers to others). I claim, to the contrary, that thinkers can have attitudes involving concepts no one fully grasps. Further, I argue that this phenomenon, which I call *incomplete understanding*, is the key to otherwise intractable problems in the philosophy of mind.

My attack on currently standard theories of content has two prongs. I first argue that such theories are unsatisfactory on their own terms – *i.e.*, apart from the issue of incomplete understanding – and that their problems can be traced back to the achievement view. I then turn to my argument that the theories cannot account for incomplete understanding. I marshal a range of examples in which thinkers plausibly have attitudes – not through deference to others – involving concepts they do not fully grasp. I argue that potential reinterpreting responses fail to provide a persuasive overall account.

In contrast to the achievement view, I propose *the responsibility view*: for a belief to involve a particular concept is for the thinker to be responsible to, or subject to, a standard associated with the concept. I argue that the responsibility view has the potential not only to avoid the difficulties encountered by the achievement view, but also to account for our having thoughts with incompletely understood contents. Since a thinker can be responsible to a standard she lacks the ability to satisfy, the responsibility view has better prospects than the achievement view for explaining how a thinker can have thoughts involving a concept that she does not fully grasp.

Contents

Acknowledgments

For much help and support and invaluable discussions, I am very grateful to a few good friends: Kinch Hoekstra, Brad Hooker, Harry Litman, Tim Macht, Nike Makris, and Nicos Stavropoulos. I am especially grateful to Andrea Ashworth for more things than I can say.

Many other friends and colleagues have read parts of this essay and given me helpful comments. They include Justin Broackes, Bill Child, Roger Crisp, Ram Neta, Christopher Peacocke, Paul Pietroski, and Barry Smith.

Georges Rey and Galen Strawson devoted a great deal of time and effort to reading this work and discussing it with me. I thank them for many valuable, and thoroughly enjoyable, hours of discussion.

I have been extremely lucky in my teachers and supervisors. Ronald Dworkin was always generous with his time. His genius for seeing the big picture helped me greatly. Joseph Raz was a wonderful teacher. He helped me to look at problems in new ways. I am grateful to Bernard Williams for many memorable supervision sessions. His penetrating comments left their mark on the thesis in its early stages.

I am especially indebted to Susan Hurley. She was my first supervisor in Oxford, and she had a great impact on the directions I took in philosophy. I am sure that her influence has affected my work in ways that I'm not even aware of.

Martin Davies was an extraordinary supervisor – a superdupervisor: insightful, clear, kind, patient, always constructive. He gave his time and help incredibly generously. There is no one better at taking up another person's point of view and helping to develop that person's position "from the inside."

I would like to thank the Marshall Commemoration Commission of the British Government for the Marshall Scholarship that brought me to Oxford in the first place and supported the first part of my graduate work. I am very grateful to Oxford University and Magdalen College for providing such a great place to do philosophy. I am indebted to Jesus College and to its Fellows for many things. They were open-minded (or crazy) enough to hire a philosophy graduate student from the United States as a law fellow. Later, their generous award of a Junior Research Fellowship funded my research during much of the writing of this thesis. Jesus College provided a wonderful working environment and tolerated my occasional interludes in the southern California desert and in Washington, D.C. Deep Springs College also provided an excellent, and unique, place to write. Finally, I dedicate the thesis to my mother Linn Turner.

Part I: Introduction

Chapter 1: The Three Problems

1.1 Introduction

This essay concerns the question of what it is for a person's thought to have a particular content. Mental states with propositional contents play a pervasive role in explanations of human beings and their behavior, both in everyday contexts and in psychological and other scientific theorizing. When we say that Bruno took the train because he believed that safety is more important than price, we attribute to him a thought involving the concept of safety, as opposed to (say) the concept of comfort. What is it for a person's thought to involve one concept rather than another, or, more generally, for it to have a particular content?[1] Theories of mental content are attempts to answer this question, which I call the content question. A working assumption of this essay is that a thought's having a certain content is not a basic, unanalyzable fact. The project is to seek an explanation of what it is in virtue of which a thought has one content rather than another.

Most leading theories of content share a very basic presupposition. In rough terms, the presupposition, which I call *the achievement view*, is that having a thought involving a particular concept is exercising a certain disposition or ability – a disposition to accomplish a mental achievement that is associated with the concept. The mental achievement might be, for example, making certain inferences or sorting objects into categories. In this essay, I challenge the achievement view and offer an alternative.

I show that it follows from the achievement view that a thinker must *fully grasp* (in a sense I make precise) a concept in order to have attitudes involving it (though most theorists would allow an exception for cases in which the thinker defers to others). Central to my attack on the achievement view are cases of thinkers who do not have mastery or full grasp of a concept but nevertheless have thoughts involving the concept. I call this phenomenon *incomplete understanding*. Because of Tyler Burge's influential arguments, most philosophers would accept that there are cases of incomplete understanding. It tends to be assumed that social deference – deference to other thinkers who fully

[1] I will use the term "thought" to mean a belief, judgment, or other attitude, *not* the content of the attitude. As a shorthand, when writing of a thought the content of which includes a particular concept as a component, I will

grasp the relevant concepts – is what makes such cases possible. I argue, however, that standard, achievement-view-based ("achievement-based," for short) theories cannot make sense of how deference could play this role. More importantly, I argue that thinkers can have attitudes involving concepts no one fully grasps.

My alternative answer to the content question is *the responsibility view*. According to that view, a thought's involving a particular concept is a matter of the thinker's being responsible to standards that are distinctive of the concept. The responsibility view explains incomplete understanding in a natural way. In general, people can be responsible to standards that they do not satisfy. Standards of rationality are an example. The fact that people are bad at reasoning does not itself exempt them from logical or rational standards (that is, does not make such standards inapplicable to them). The responsibility view explains thoughts involving incompletely grasped concepts as cases in which the thinker is responsible to standards the thinker does not satisfy. The relevance of deference to others is that it is one mechanism that can make a thinker responsible to the standards of a concept she incompletely grasps. In addition, I argue that the responsibility view has the potential to solve or avoid problems that seem intractable for standard, achievement-based theories of content.

The question of what it is in virtue of which a thought involves a certain concept bears on a wide range of philosophical issues. Issues about knowledge, disagreement, and interpretation are examples. Whether a person's beliefs are true depends on their content; thus, whether a person has knowledge (and, if so, what knowledge she has) depends in part on the correct description of the content of her beliefs. Similarly, in order for people to agree or disagree, they must think contents involving the same concepts. Finally, how we should interpret a text, a practice, or a theory will often depend in part on the contents of thoughts. These examples suggest that the correct account of the contents of thought matters in such diverse areas as epistemology, philosophy of science, ethics, anthropology, and literary and legal theory.

This chapter and the next give an overview of the thesis. This chapter introduces the content question and some standard attempts to answer it, and sketches difficulties that these attempts encounter. In the next section (1.2), I set out some methodological assumptions and get a few

---

write that the thought *involves* the concept.

terminological points out of the way. In section 1.3, I discuss what is involved in answering the content question, and show that the issues known as the "Kripkenstein" problem and the "disjunction" problem are just different formulations of the content question. Next, in section 1.4, I briefly describe the two most prominent types of answers to the content question: informational and inferential-role theories of content. In section 1.5, I introduce what I regard as three important problems that any such theory must face – *the Three Problems*. Section 1.6 provides capsule summaries of two prominent, and very different, theories of content that will serve as exemplary versions of informational and inferential-role theories. The following three sections (1.7-1.9) give a preview of my arguments with respect to each of the Three Problems. Finally, in section 1.10, I point out that the Three Problems are not merely a miscellaneous collection of difficulties, but reinforce each other.

Chapter Two introduces the achievement view and the responsibility view, shows that the Three Problems can be traced back to the achievement view, and outlines the reasons why the responsibility view has better prospects than the achievement view for dealing with the Three Problems.

1.2 Preliminaries

In this section, I set out some methodological assumptions and clarify my terminology. Although the term "content" may be philosophical jargon, the notion of a linguistic expression's or an attitude's having a particular content is not a specialized notion deriving from philosophy, but a familiar and indispensable notion in any realm concerned with human behavior. Everyday explanations of human behavior, literary and legal interpretations, and anthropological and psychological theories all make use of attributions of propositional content. To borrow Fred Dretske's (1988, pp. 79-81) example, it is the frequency of the opera singer's singing that shatters the glass, but it is the content that makes the audience cry. (Imagine trying to explain why the spectators at the football match suddenly roared without appealing to the content of their belief – that the home team had scored). In a very different context, a judge might argue for an interpretation of a statute by claiming that the legislators' intentions had a particular content. And psychologists routinely explain behavior – for example, precise differences in reaction times – by appealing to the contents of mental states.

Such attributions of content make use of sentences that express propositions. For example, the alleged content of Bruno's belief – the one that explains his taking the train – is the proposition expressed by the sentence "safety is more important than price." In this essay, I will take for granted that propositions are abstract objects with determinate identity conditions. I will also take for granted that propositional contents are composed of concepts – abstract objects that can be combined in different ways to form propositions. (As noted above, when writing of a thought the content of which includes a particular concept as a component, I will write that the thought *involves* the concept.) It is important to be clear that the notion of a concept with which I am working is therefore not a notion of a mental particular. To avoid confusion, when discussing the views of theorists who use the term "concept" to mean a mental particular, I will not use "concept" in their way, but will instead use a term such as "mental symbol" or "mental representation."

In order for propositional contents to be true or false and to be about objects, properties, or events, concepts must be such that various kinds of things can satisfy them or fall under them. For example, predicate concepts can have reference. In my terminology, the reference of a concept is a property. Thus, the reference of the concept of water is the property of being water. For our purposes, it will not be necessary to choose between competing views of the nature of properties. I will assume only that properties are individuated finely enough that a *necessary* condition on property identity is sameness of instances in all possible worlds. (Arguably, it is not a sufficient condition; for example, it is plausible that there can be non-identical, necessarily uninstantiated properties.) I will use the term "extension" in a standard way, such that the extension of a concept is the set of objects in the actual world that fall under it. So the extension of water includes all instances of water in the actual world. Thus, coextensionality is not sufficient for sameness of reference (and I am leaving open whether cointensionality – in the sense of sameness of instances in all possible worlds – is sufficient for sameness of reference).

In trying to understand the nature and status of content, I will begin at the source, with ordinary, non-philosophical attributions of content. These are the data from which the notion of content and the problems about it derive. It may be that ordinary and theoretical purposes ultimately diverge. We cannot know that, however, unless we first see how far we can go in making theoretical sense of ordinary attributions of content.

One concern is that ordinary attributions of content are subject to pragmatic constraints. Our willingness to use a particular sentence in a "that"-clause to ascribe a thought to a person may be sensitive to context in a variety of ways that are not relevant to the theoretical project of answering the content question. An objector may claim that the further the subject is from us in time or culture the more relaxed are our standards for judging the appropriateness of apparent ascriptions of content. Thus, we cannot assume that ordinary attributions of content are univocal. In the case of people from very different cultures, children, or non-human animals, for example, we may use propositional-attitude talk loosely, using it merely as a convenient way of indicating the rough subject matter of the person's or animal's thought.

This important concern deserves fuller consideration than I can give it in this essay. Let me make a few remarks at this point.[2] First, reinterpretation is always a possible reaction to philosophical data that conflict with an established theory. Whether reinterpretation is the best reaction in particular cases will depend on how otherwise successful the motivating theory is. As we will see, I rely on certain data of ordinary attributions of content that leading theories of content cannot explain. Philosophers committed to those theories will thus be tempted to reinterpret the data. I attack the theories in question, and the fundamental assumption they share, on grounds that are independent of the contested data. To the extent that my attack is successful, the motivation for reinterpreting my data is removed.

Second, the idea that our standards for the use of propositional-attitude language are looser as we move to distant contexts presupposes that there is a home context, and a central or core use of propositional-attitude language that is required in that home context. There must be a home context for the notion of distant contexts to make sense. And the idea that we loosen standards as we move away from the home context implies the existence of a strict or core use. How could we make the appropriate pragmatic adjustments if there were no starting point from which to adjust? If an expression or notion has no literal content, it can hardly be used more or less loosely or strictly. The effectiveness of attitude-ascription talk thus depends on there being a stable core to the ordinary

[2] See also section 6.8.

notion of a thought with a particular content. We can therefore proceed by considering ordinary attributions of content to see what theoretical sense can be made of that notion. Along the way, we will need to be alert to the possibility that our ordinary practice is affected by pragmatic considerations.

The fundamental commonsense notion in this area is that of a person's having an attitude with a particular content. Since the main question of the essay is what it is for a thought to have a particular content, it is important that we not take for granted a position on that issue. Also, I will not assume answers to questions about how concepts must be individuated in order for them to play their role as constituents of the propositional contents we attribute to each other. For example, I will not take for granted an answer to the question whether concepts are individuated at the level of reference or, more finely, at the level of cognitive significance.

Three more preliminary points will be useful. First, for simplicity of exposition, I will often frame the discussion in terms of one concept, asking not what it is in virtue of which a person has an attitude with a particular *content*, but what it is in virtue of which a person's attitude involves a particular *concept*. I intend this way of talking to be neutral with respect to holistic and anti-holistic views of concepts and of content attribution. Second, I will refer to concepts and contents either by writing the concept terms in upper case letters – as, for example, WATER – or by using the locution "the concept of," as in "the concept of water."[3] Third, I will use the phrase "a *constitutive* account of content" as a shorthand for "an account of what it is in virtue of which a thought involves a particular content" and, similarly, will write that a condition is "constitutive of content" to mean that satisfying that condition is what makes it the case that a thought has a certain content.

1.3 The content question

1.3.1 Reduction and other constraints

At least since Saul Kripke's (1982) reflections on Wittgenstein's rule-following discussion, the content question has held a central place in philosophy of mind and psychology. Much of the debate has been dominated by two camps. On the one hand, there are those scientifically-minded philosophers who have attempted to give reductive accounts of content. For example, some aim to

[3] When I use single letters to refer to concepts and sometimes when I refer to whole propositional contents, I will use italics, as in "the concept *C*."

use only terms acceptable to the physical sciences. On the other hand, there are diverse philosophers – it is not really apt to call them a "camp" – who hold that content is "normative" in some essential respect and that the kind of explanation appropriate to such a normative domain is fundamentally different from that appropriate elsewhere. This group tends to be quietist about content, rejecting the idea that we can give an answer to the content question or reducing the ambitions of attempted answers.

It is evident that there is large middle ground between reductivism and normativist quietism. For example, it is possible to recognize that content has normative features, but to seek an explanation of those features. We should reject the assumption that if content were normative, it would be outside ordinary explanatory paradigms. More strongly, I argue that an appeal to normativity can be explanatorily fruitful.

The goal of explanation comes with constraints that are related to those of naturalism. For example, in order for an answer to the content question not to assume what it is trying to explain, it must not appeal, at the end of the day, to unanalyzed intentional facts or properties. An important motivation for naturalism is that the natural sciences are a source of resources with which an explanatory account can avoid circularity.[4] For several related reasons, however, we should not begin by insisting on strong reductive requirements, such as a restriction to the terms of physical science.

First, these requirements are not respected by successful sciences. High-level sciences, such as geology, dynamic systems theory, biology, ethology, neuroscience, and even macro-mechanics, fruitfully employ notions of which they cannot provide a reductive account. Consider, for example, the notions of tectonic plate, dynamic feedback loop, organism, species, and evolutionarily stable strategy. A notion can earn its keep by yielding explanatory gain without our having a reduction of it. To take a historical example, if Darwin had restricted himself to notions for which physicalistic reductions were available, he would have denied himself access to notions, such as heredity with variation, necessary for the theory of evolution by natural selection. Later developments in molecular biology explained the mechanism of heredity. Molecular biology does not provide a reduction of the theory of evolution by natural selection, but in combination with that theory

[4] Rey makes this point in forthcoming a; forthcoming b.

promises a full naturalistic explanation of the origin of species.

Second, as we will see, a non-reductive account of content can have explanatory value by showing that content is an instance of a broader phenomenon with which we are familiar – a standard kind of explanation in science. Galvani's recognition that neurons make muscles contract by transmitting electrical stimuli was explanatorily powerful, despite the fact that the nature of electrical phenomena was not yet understood.

Third, projects that insist on one-step reductions risk missing possibilities for explanation. It may be best for the overall progress of explanation to make a connection that in the present state of knowledge does not provide reduction. It is thus prudent to begin with a flexible approach to explanation; accounts in terms of high-level notions may not only offer understanding, they may later be combined with accounts of particular notions to provide explanations that would not be available without the intermediate, high-level accounts.

Thus, I will not assume that the content question demands a reductive answer. Rather, the crucial question in deciding whether we can make use of a notion, property, or fact will be whether it does explanatory work. For example, I will not take for granted a ban on normative notions. They are not themselves content-involving and it is not evident that normativity depends on content. Even if it does, there are ways in which normativity could depend on content that would not give rise to circularity. For example, it may be that if there were no creatures with intentional states, there would be no moral requirements, but it does not follow that it is in virtue of the contents of mental states that particular moral facts obtain. And even an explanatory circle may be acceptable if it is large enough.

I am thus proposing to proceed in a way that does not accept, on the one hand, a restriction to the terms of the natural sciences and, on the other, quietism about the content question. The account that I will offer is sympathetic to the naturalistic project of giving an answer to the content question in non-intentional terms. But I argue that taking account of the normative aspects of thought is crucial to the project of answering the content question.

The content question can be presented in the following straightforward form, though, as we will see, that is not the form in which it has typically been addressed in the philosophical literature.

A person is claimed to have (correctly) applied a concept *C* to an object. The attribution of a content involving *C* characterizes a state of a flesh-and-blood human being in terms of one abstract entity rather than others. What is it in virtue of which the concept the person applied was *C*?

Before reviewing some of the standard ways of trying to answer the content question, we should consider two familiar versions of it. First, Kripke's (1982) discussion of Wittgenstein's rule-following arguments posed a problem that has received wide attention – the "Kripkenstein" problem. Second, a central focus of much of the literature on content is the "disjunction" problem. The Kripkenstein and disjunction problems are merely special versions of the problem posed by the content question, though this may not be obvious at first glance. I will consider them in turn.

1.3.2 The "Kripkenstein" version of the content question

Kripke's illustration concerns a person who has apparently added correctly two natural numbers, 68 and 57, that are larger than those the person has previously added. The challenge is to say what makes it the case that the person meant ADDITION and thus reached the correct result, as opposed to meaning QUADDITION, where quaddition is a function that is precisely similar to addition until one of the arguments reaches 57, from which point on the function's value is 5. Obviously, all the evidence of the person's previous putative addings is consistent with both the addition and quaddition hypotheses.

There are three features of Kripke's discussion that can make it difficult to see that the Kripkenstein problem is really just a colorful way of posing the content question.[5] First, Kripke places weight on what he calls the *normativity* of meaning: he emphasizes that what the person means determines not what the person will do, but what the person should do (pp. 11, 24, 28-29, 37). But this is just another way of saying that an answer to the content question must differentiate between an incorrect application of one concept and a correct application of a different concept. It is because of the possibility of such an incorrect judgment that what the person means does not determine what the person actually will do (p. 37). For example, Kripke famously argues that

[5] In addition to the features discussed in the text, there are two more minor ways in which Kripke's exposition may obscure the fact that the question at issue is the content question. First, he sometimes talks in epistemic terms. He makes clear, however, that the real question is what makes it the case that the person meant ADDITION rather than QUADDITION, not how we can know that the person meant ADDITION. Kripke accordingly specifies that any facts, not just those to which we have access, may be used to respond to the challenge (1982, pp. 11, 21, 38-39). Second, in order to minimize the difficulties of using language to shed doubt on meaning, Kripke presents the problem as an issue about what makes it the case that the person meant addition in the past (1982, p. 12). The problem is no different in the present.

appealing to dispositions cannot help because having a disposition to do something does not justify doing it. Again, however, this is just another way of arguing that a theory of content must allow for a distinction between correctness and incorrectness (pp. 24, 28-29).[6]

It is sometimes supposed that the Kripkenstein problem depends on some claim about the normativity of content that is stronger than the claim that a theory of content must distinguish between correctness and incorrectness. Such a claim might be, for example, that a person should use a given word with the same meaning as in the past or in accordance with his past intentions.[7] Or the claim might be that attributions of content are a way of placing a person within a normative framework of reasons that has no mirror at the physical level and thus cannot be captured in non-intentional terms.[8] Perhaps some such supposition explains why much of the literature on naturalizing content focuses on the disjunction problem, rather than the Kripkenstein problem (*e.g.*, Dretske 1981; Fodor 1990). As we have seen, however, the Kripkenstein problem will arise given only the assumption that the concept that is applied to an object determines the conditions under which the judgment is correct or incorrect. If a concept determines certain standards of correctness, an adequate explanation of what it is in virtue of which a thought involves that concept must account for the appropriate standards of correctness. Thus, the core of the Kripkenstein problem does not depend on a stronger claim about the normativity of meaning than the claim that content determines correctness and incorrectness.

Second, it is sometimes assumed that the problem depends fundamentally on there being infinitely many numbers. This assumption is understandable since Kripke focuses on a mathematical example, appeals to the fact that there will always be numbers greater than any the person has previously added, and argues that past performance and dispositions cannot solve the problem because both are finite (pp. 8, 18, 26-28). In fact, however, though the infinitude of the natural numbers makes the presentation simple and compelling, the central problem is not special to numbers or infinitude.[9] In the first place, Kripke could have presented the problem for the concept of addition without relying on the infinitude of numbers. Consider an example involving numbers the person had

---

[6] There is, however, more than one strand to the argument that Kripke bases on the claim that a theory of content must distinguish between correctness and incorrectness. See section 6.7 below.
[7] See, *e.g.*, McGinn (1984, pp. 147, 174).
[8] See, *e.g.*, McDowell (1985). I do not mean to suggest that McDowell thinks that the Kripkenstein problem depends on such a claim.
[9] This is not to deny that contents involving numbers may present special problems because of infinitude. In general, I should note that Kripke's rich discussion raises many different issues; my point in this section is

previously (apparently) added: the "bent" quaddition-style rule could specify that the correct answer is 5 after a certain date or in a particular context. Such an example depends only on the possibility of a situation that is in some way different from past situations in which the concept was applied; as Goodman's grue example illustrates, there is nothing about this structure that is peculiar to mathematics. Since the question of the application of any concept, whether mathematical or not, can always arise in a new situation, the problem is completely general. Kripke himself points out that the problem applies to all language and gives an illustration involving the concept TABLE (pp. 7, 19.) More fundamentally, though Kripke does not make this point, the problem does not even depend on the possibility of a new situation: the problem can be posed with an example involving a person who *always* made a particular apparent error in adding or in applying some non-mathematical concept. The issue would be whether the putative systematic error was in fact an error or a correct application of a different concept.

Third, Kripke initially describes the problem as a problem about rule-following because Kripke is presenting his version of Wittgenstein's rule-following argument (pp. 7-8). This characterization may impede recognition that the problem is that posed by the content question. The phrase "following a rule" is often used colloquially such that a necessary condition for John to be following a rule is that his behavior accord with the rule. Moreover, it may even be *sufficient* that his behavior accord with the rule; that is, he can count as following a rule even if his behavior is not guided by the rule. (For example, when we ask whether you are following the speed limit rules, we are interested only in whether your speed accords with the rule; if your speed is below the limit, you are following the rule, even if you were unaware of the rule, and if your speed is over the limit, you are not following the rule, even if you were trying to conform to the rule. Your intentions might be relevant to the question of legal liability, but not to the question of whether you were following the rule in this colloquial sense.)

Under the notion of rule-following with which Kripke is concerned, however, that your behavior conforms to a rule is neither necessary nor sufficient for you to be following that rule. The results of your computations may be correct under the rule for addition, yet you may not be adding but quadding. Even if your results would be incorrect under the rule for quaddition, you could be

---

simply that a central theme of Kripke's discussion is the content question.

following the rule for quaddition, *i.e.*, you could be quadding; perhaps you were distracted. The possibility of error cannot be ruled out if we are to have any prospect of an account of content. Thus, Kripke's question is what distinguishes a person who is following a particular rule and acting in accordance with it from a person who is following a different rule but not acting in accordance with it.[10] The question of what it is in virtue of which a person is correctly applying one concept (ADDITION) rather than incorrectly applying some other concept (QUADDITION) is a specific version of this question.

In fact, it becomes clear immediately that the principal case of rule-following with which Kripke is concerned is the case of a person's applying a word in accordance with what that person means by it (pp. 8-9). Although the discussion is focused on the application of a linguistic expression, the word "plus," Kripke emphasizes that the problem concerns the *content* of the person's mental state of meaning ADDITION or QUADDITION, or of the person's intentions with respect to the word "plus" (*e.g.*, pp. 8, 11). The same basic problem applies to the content of any propositional attitude. Just as we can ask what determines that the content of a person's intentions with respect to a word is such that the application of the word is a correct application of one concept rather an incorrect application of another, we can ask what determines that a belief about some object is a correct belief involving one concept rather than an incorrect belief involving a different concept.

1.3.3 The "disjunction" version of the content question

The other well-known form of the problem posed by the content question is the disjunction problem. The problem can be seen most clearly in the context of a theory that attempts to account for content in terms of covariation between bearers of content and objects or properties in the world. (I discuss such *informational* theories of content in section 1.4 and in chapters 3 and 4 below.) For simplicity, we can take the bearers of content to be (subpersonal) mental representations or symbols.[11] The presence of an object that is within the extension of a mental representation (*i.e.*, the extension of the concept that the representation stands for) is not the only thing that can (reliably) cause that mental representation to be tokened – *i.e.*, to occur or to be triggered. Other mental states, such as a desire for an absent object can cause it to be tokened. Symbols can be tokened mistakenly, as the result of

[10] The basic question can be posed even in cases where the person's actual behavior conforms to both rules, but it dramatizes the problem to consider examples in which the behavior would be incorrect under one of the rules.
[11] As with words, mental representations can be tokens or types; they are type-individuated by their content.

mistaken perception of an object not within the concept's extension. And such mistakes can occur systematically. Conversely, objects within the representation's extension can systematically fail to cause occurrences of the representation. The problem for an informational theory is to specify what makes it the case that some of the things that (reliably) cause a representation to be tokened fall under the concept that the representation stands for, while others do not. Something must determine that the thinker is using a particular concept rather than a more (or less) inclusive one. In other words, something must determine that the concept being used has an extension that includes *these* things and not *those*, where some of these things do *not* reliably cause occurrences of the mental symbol that represents the concept, and some of those things *do* reliably cause occurrences of it.

The problem is dubbed "the disjunction problem" because typical examples raise the question whether a thinker is using a concept whose reference is a certain (intuitively) non-disjunctive or normal property or a more inclusive, disjunctive concept. For example, suppose a thinker reliably tokens a mental representation in response to horses, but also to cows on dark nights. The problem is to say what distinguishes the incorrect use of one concept – HORSE – from the correct use of a different concept – HORSE-OR-COW-ON-A-DARK-NIGHT (*e.g.*, Dretske 1981; Fodor 1990).

Although the more inclusive concept's extension can be specified with a disjunction, the problem need not involve a concept with an intuitively disjunctive or strange content. A well-known example concerns whether the content of a frog's perception is FLY or MOVING BLACK SPOT, given that frogs reliably snap at all moving black spots, not just those that are flies. More theoretically, even if a concept's extension can be captured with a disjunction, it does not follow that the concept in fact has a disjunctive compositional structure or is intuitively strange.

As further discussed in the next chapter, what is essentially the disjunction problem, though with appropriate adjustments, can be posed for less inclusive or bent concepts as well as for more inclusive ones. Suppose, for example, that some, perhaps particularly bovine, horses reliably fail to cause occurrences of the putative HORSE symbol (though nothing that is not a horse reliably causes occurrences of that symbol). What is it in virtue of which the person is using the broader concept – HORSE – as opposed to a narrower concept – a concept that applies only to horsey looking horses?

Despite their basic equivalence, the Kripkenstein and disjunction problems tend to be discussed separately and to evoke different kinds of reactions. In part, the reason is that, as we have seen, neither problem is posed in a fully general way. Kripke focuses on an example in which the challenge is to say why the thinker's apparently *correct* application of an *unbent* concept is not, on the latest occasion, an incorrect application of a bent concept. The disjunction problem is typically posed in terms of a challenge to say why apparently *systematically incorrect* occurrences of an intuitively non-disjunctive concept are not correct applications of a different, *more inclusive* concept (*i.e.*, a concept whose extension includes the extension of the other concept as a proper subset).

The disjunction problem manages to seem less fanciful than the Kripkenstein problem. It is widely understood that the disjunction problem can arise in the context of ordinary mistakes and without appeal to queer concepts. In contrast, philosophers strongly associate the Kripkenstein problem with implausible scenarios and weird quaddition-style concepts. As noted above, however, the force of the Kripkenstein problem is not restricted to cases in which the thinker's past applications are assumed to be correct, and she is confronted with a wholly new situation. Thus, the Kripkenstein problem can turn on the possibility of the plausible, ordinary sort of mistake that is typically used to set out the disjunction problem. (Of course, our intuitions about what concept the thinker is using may vary depending on how plausible or far-fetched the putative error is.)

Other factors that may impede recognition that the two problems are essentially variants of the same problem are the special features of Kripke's exposition discussed above, most importantly the weight he places on "normativity." Despite the different emphasis and terminology of the typical presentation of the disjunction problem, however, that problem also concerns normativity in the relevant sense. As discussed above, the only normative claim required to generate the Kripkenstein problem is that content determines correctness and incorrectness of applications. Since the disjunction problem concerns how to give an account of what distinguishes correct applications of one concept from incorrect applications of a different concept, it involves the same normative claim.

1.4 Answers to the content question

Now that we have seen that the Kripkenstein and disjunction problems are forms of the more general problem posed by the content question, we can illustrate possible answers to that question without adverting specifically to the peculiarities of typical presentations of the Kripkenstein and disjunction problems.[12] An answer to the content question will provide solutions to both of them. As discussed above, the content question asks what it is in virtue of which a state of a human being is correctly characterized in terms of one abstract entity – a particular content or concept – rather than others. Various kinds of answers have been given.

We can distinguish between, on the one hand, accounts of content that give a blanket answer to the content question, an answer that is the same for every content, and, on the other, accounts that give answers that are specific to each content (or at least a program for providing such answers). An example of the former kind of account is the Davidsonian view that a person has (an attitude with) a particular content in virtue of that attribution's being part of the best interpretation of the person. The best interpretation of the person is the total set of attributions of propositional attitudes to the person that would make the person as intelligible as possible. Such an account tells us nothing about the individual contribution of each concept to overall intelligibility. In contrast, an example of the latter, *concept-by-concept* kind of account would be an account that holds that a content involves the concept HORSE in virtue of a specified causal relation between a particular mental representation and the worldly property of being a horse.

The latter kind of account aims closer to the ideal of explaining what it is in virtue of which a person's thought involves one concept rather than another. It has a distinct story to tell with respect to each concept. In contrast, the Davidsonian account presupposes that there is a fact of the matter about the contribution each concept makes to a person's overall intelligibility, but it assumes that nothing can be said about that contribution on a concept-by-concept basis. We should at least begin by aiming for the more ambitious type of account. The following two chapters therefore focus on the two most influential types of concept-by-concept answers to the content question: inferential-role and informational theories.[13]

An inferential-role theory sets up a correspondence between concepts and dispositions by means of an isomorphism between, on the one hand, the abstract relations between concepts or

[12] Also, it will not be necessary to treat the Kripkenstein and disjunction forms of the problem separately, and I will sometimes use the terms "Kripkenstein problem" and "disjunction problem" interchangeably to refer to the general problem.
[13] On "teleological" theories of content – theories that appeal to biological function – see sections 7.2, 7.7 below.

propositions, and, on the other, certain causal relations between states of the thinker – in effect the thinker's dispositions to move between mental states (Block 1986; Loar 1981). (The states of the thinker that are *relata* could be either attitudes themselves or mental representations.) We can make an analogy to the way in which numbers are deployed to characterize masses or other physical magnitudes of physical objects. The point of characterizing an object as having a mass of 55 grams is to rely on the relations between the number 55 and other numbers to capture the relations between the object's mass and the masses of other objects. Similarly, according to an inferential-role theory, the point of characterizing a thought as having a certain content is to rely on the relations between the thought's content and other contents, which derive from the relations between their component concepts, to capture the causal relations between the thought and other thoughts.

There are several dimensions along which inferential-role theories can vary. For example, such theories can disagree about which connections between concepts are defining or constitutive. Further, a theory may or may not allow some form of idealization of dispositions, may or may not include among the relevant dispositions causal relations between states of the person and states or properties outside the person, and may or may not individuate dispositions broadly (in terms of how things are in the surrounding world).

A sophisticated form of inferential-role theory relies, not on facts about what inferences the thinker makes, tends to make, or would make under various conditions, but on the causal grounds that lie behind and explain those facts (*e.g.*, Peacocke 1992). I will say that a thinker has a *conditional behavior* to φ under condition C to mean that the thinker would φ under that condition. I will use the term *ability* as a technical term for the underlying causal ground for the conditional behavior. (These usages are purely stipulative, adopted in order to have available a convenient way of distinguishing between theories that appeal to counterfactual behavior and theories that appeal to underlying causal grounds; I don't intend to take a position here in the debate about the nature of dispositions.) So, in my terminology, my ability to understand French is part of the causal explanation of my conditional behavior to respond in certain ways when asked questions in French. (It should be clear that I am not using the term *ability* in the very weak sense in which every normal human being has the ability to speak German or to swim, even if he or she doesn't know how to do so.)

An informational theory relies on a causal (or nomic) relation between bearers of content and worldly objects or properties. For a mental representation (or other bearer of content) R to stand for a particular concept *C* is for the thinker to be disposed to token R in a specified causal relation to objects that fall under *C* (or to the property of being C). For the thinker's thought to involve *C* is for thinker to token R (in a way that is appropriately related to other parts of the thinker's mental economy). (Because informational theories rely on a direct, causal relation between a mental representation and its content, to the extent that a theory is informational – hybrids are possible – it must maintain that there is no more to content than reference. In other words, it cannot allow that a concept's content is in part constituted by connections to other concepts.)

Informational theories allow a great deal of variation along several dimensions. The most straightforward option is to take mental representations to be the bearers of content, but it is also possible to do without mental representations by relying on relations between attitudes themselves and the world.[14] Theories can take different views about, for example, which concepts are primitive (complex concepts derive their content from primitive ones of which they are composed). Most importantly, theories can take different positions on the precise nature of the covariation relation that is claimed to constitute (or at least be sufficient for) reference. For example the relation could be causation, counterfactual covariation, nomic dependence, or some specified, more complex relation. A related point is that informational theories, like inferential-role theories, can appeal to ideal conditions: a person's mental representation refers to a certain property just in case the person's disposition to token the mental representation bears the appropriate relation to the property *when certain ideal conditions hold*. The ideal conditions can themselves be specified in diverse ways. Another proposal is to rely on a *ceteris paribus* clause, rather than attempting to specify ideal conditions.[15]

Finally in this section, we should note that there is also the possibility of a hybrid between an informational and an inferential-role theory, combining, for instance, an inferential-role theory of logico-mathematical concepts with an informational theory of other primitive concepts.[16]

---

[14] Davidson has suggested, though not spelled out, a position that would appeal to causal relations between attitudes and events or states of affairs (Davidson, 1986a; 1986b).
[15] See Pietroski and Rey 1995.
[16] Another dimension along which theories of content can vary is whether they rely on community rather than individual dispositions. Community-based versions of many of the theories mentioned in the text are possible.

1.5 The Three Problems for standard accounts of content

I will raise three problems for currently popular theories of content (*the Three Problems*, for short). First, I argue that those accounts fail even on their own terms. In theory, of course, the accounts have the aspiration to give a complete answer to the content question – that is, to say what makes it the case that thoughts have their particular contents.[17] In practice, however, leading theories typically assume that what they have to do in order to show that they have answered the content question is to solve *the problem of error*.

Let me explain. A theorist of content proposes a condition or condition schema – a way of generating a specific condition for each content – that is claimed to be constitutive of an attitude's having a particular content (a *condition for content*, for short). As I discuss in section 2.1, leading theories of content share a deep presupposition, *the achievement view*: in very crude terms, what makes it the case that a thought involves a particular concept is that the thinker has (and activates) a disposition that corresponds to the concept in a specific, perspicuous way. For example, if a concept is individuated by its connections with other concepts, the relevant disposition is a disposition to make the appropriate inferences; if concepts are individuated by the properties that are their references, the relevant disposition is a disposition to apply a mental representation to the appropriate property.

Leading theories of content thus begin by proposing a condition for content that conforms to the achievement view. They then take their further task to be to show that the condition can allow for the possibility of error – that is, to show that a thinker who has the supposedly content-determining dispositions can nonetheless be disposed to make systematic errors in applying the concept, or, to put it another way, that the proposed condition would not have the consequence that whatever a thinker was disposed to do was correct. (The tendency of theorists to address the content question in the form of the disjunction and Kripkenstein problems is a manifestation of the focus on the problem of error.) As we will see (sections 2.4.1 and 2.6), the achievement view is what makes the problem of error so hard to solve.

In sum, theorists typically confine themselves to offering an achievement-based proposal and

[17] Or, at least, to give a condition that would be *sufficient* for content. See section 4.4.

arguing that it allows for error. As my discussions of the second and third of the Three Problems will make clear, there are other possible ways of supporting a theory of content. When I say that theories of content fail *on their own terms*, I mean that, even leaving aside other challenges for a theory of content, theorists fail in their attempts to support their proposals, in particular in their attempts to solve the problem of error. Moreover, the nature of the failure suggests that the defect is basic to the general types of theories I discuss, and thus that more than mere tinkering with the details is required.

I further argue that even if a theorist offered an account that succeeded on its own terms, the theorist would not have done enough to make it plausible that she had answered the content question. The second and third of the Three Problems involve ways in which theorists have underestimated what is required to show that an answer the content question is successful.

My discussion of the second problem takes off from consideration of the question, pressed by Kripke (1982) and Boghossian (1989), of accounting for the normativity of content. In addition to arguing that we cannot specify dispositions that correspond to correct applications, Kripke made the apparently different claim that, because content is fundamentally normative, dispositions are the wrong kind of material with which to fashion an account of content. As the above discussions of the Kripkenstein and disjunction problems make clear, an answer to the content question must determine conditions of correctness and incorrectness. To take a silly example, the fact that I am using the concept of addition makes it correct for me to reach certain results. But, arguably, the fact that I am disposed to reach those results when the moon is full does not make it correct for me to reach those results. Therefore, the fact that I am disposed to reach those results when the moon is full cannot be what it is in virtue of which I am using the concept of addition. Have leading theories of content failed to explain how the properties that they claim constitute content could make anything correct or incorrect?

The argument that standard theories of content fail to account for the normative force of content raises difficult questions about what is required of a theory of content. Must a theory provide a condition for content such that it is *a priori* that a thought that satisfies the condition will have certain normative features? Or is it sufficient to provide a condition such that a thought that satisfies the condition will have those normative features in all nomologically or metaphysically possible worlds? I avoid this issue by arguing that, at least given the current state of play, unless a theorist can offer an intuitive explanation of the normativity of content, the theorist will encounter difficulty making it plausible that her proposed conditions for content are successful. To put it another way, even if, as a metaphysical matter, a theory of content need only specify a condition that entails the normative features of content, showing that a proposed condition has an explanatory connection to those normative features is, in practice, the only way to support the theory. Leading theories of content fail to gain any plausibility in this way, however.

As a shorthand, I will refer to this set of issues as the problem of *normativity*.

Third, I argue that the leading naturalistic theories have taken too narrow a view of the task in a different respect. They do not attempt to account for a significant part of the data of attitude attributions – in particular, attributions involving concepts that the thinker has not fully mastered. In chapter 2, I make a crucial distinction between, on the one hand, a thinker who has full mastery of a concept, but nevertheless makes mistakes in applying it (for example, because of distraction or interferences of other kinds) and, on the other, a thinker who lacks full mastery of the concept. I use the term "incomplete understanding" for the latter kind of case. When theories do not ignore incomplete understanding altogether, they tend to suggest that it is a derivative and peripheral phenomenon that can be addressed by adding to an account of fully understood contents some kind of proviso about *deference*. In other words, they assume that an incompletely understood content is possible because the thinker with incomplete understanding defers to other thinkers with full understanding. As a preliminary matter, I argue that standard accounts cannot adequately explain how deference can enable a thinker to have an attitude involving an incompletely understood content. My central argument is that incomplete understanding is not limited to cases that are dependent on deference (*dependent incompletely understood contents*, or *DIUCs*); the standard theories do not even recognize the existence of *independent incompletely understood contents* (*IIUCs*) and have no prospect of accounting for them.

In sum, the first of the Three Problems is the problem of the failure of standard theories of content on their own terms. Theorists of content fail to support even the limited claims they make for their proposals. The second problem is the problem of normativity. The bottom line of my discussion of normativity will be that theorists of content encounter an epistemic difficulty in trying to convince us that they have provided constitutive accounts of content. The third problem is that leading theories of content are unable adequately to account for DIUCs and IIUCs.

The remainder of this chapter is devoted to a preview of the Three Problems. I discuss the third at somewhat greater length than the other two. There are several related reasons for this difference in treatment. Readers are likely to be familiar with the first problem and are even likely to agree that existing accounts do not succeed on their own terms. Of course, theorists interpret this lack of success in different ways. Those who take the view that content cannot be naturalized take it to be confirmation of their view. In contrast, those engaged in trying to provide a naturalistic account tend to assume that we have simply not got the details right – we just have to find the right conception of ideal conditions or *ceteris paribus* clauses or whatever. Whether I am successful in showing not only that the accounts are unsuccessful, but that the problems are fundamental to the standard kinds of accounts, will depend on the details of my arguments in chapters 2 through 5. The second problem lies in difficult and highly controversial territory, but the nature of the problem can be relatively briefly set out, and, in the end, my argument places little weight on the second problem.

In contrast, it requires substantial discussion simply in order to explain the nature of the third problem – that standard accounts have failed to address incompletely understood contents. Since the claim is that standard accounts have neglected a significant portion of the data, it is to be expected that there will be controversy about the proper characterization of the attitude attributions in question. Thus, this chapter does some preliminary work to clarify the nature of the supposed problem.

Moreover, as I explain further below, the third problem is more central to my overall argument than the first two. My argument that standard theories of content fail even on their own terms is intended to prepare the ground for my arguments that such theories do not account for incompletely understood contents. (It also provides important support for my argument that the standard theories have difficulty supporting their proposed conditions without demonstrating an intuitive connection between those conditions and the apparently normative features of content.) An attempt to revise the way in which a problem is conceived is less likely to obtain a sympathetic hearing if established theories are thought to be adequate. Thus, if current theories were satisfactory on their own terms, my central argument would face a more difficult task.

1.6 Fodor's and Peacocke's theories

In discussing the Three Problems, I will focus in particular on two accounts, Christopher Peacocke's inferential-role account and Jerry Fodor's informational account. Each is a highly sophisticated and developed example of its type, and the two theories are about as different from each other as two (concept-by-concept) accounts of content can be.

Peacocke's account holds roughly that concepts are individuated by their inferential roles. In contrast, Fodor holds that concepts are individuated exclusively by the properties to which they refer. On a different dimension, Peacocke holds that the primary level at which attitudes are predicated is the personal level, while Fodor holds that the primary bearers of content are sub-personal mental representations. Accordingly, Peacocke's account appeals to a personal-level notion – the thinker's finding certain inferences, or transitions between intentional states, natural[18] – while Fodor appeals only to relations between mental representations and worldly properties, explicitly excluding any reference to personal-level notions, such as the way in which a person thinks of a property. A related point is that while Fodor attempts to account for content in non-intentional terms, Peacocke's view is non-reductive, holding that intentional contents are ineliminable.

Despite the polarity of the two views in these respects, both theorists are engaged in the same project at a very general level. Both hold that beliefs and desires cause and causally explain behavior in roughly the way that folk psychology supposes. Further, both take beliefs and desires to be implemented by actual inner states, though this is not essential to Peacocke's view. And they both provide concept-by-concept, rather than blanket, accounts of what it is for an attitude to involve a particular concept.

1.7 The first problem: current theories fail on their own terms

Chapters 3, 4, and 5 argue that currently standard theories of content fail even on their own terms. They fail adequately to address the problem of error and are thus unsatisfactory even apart from the problems of normativity and incomplete understanding.

[18] Peacocke's term is "primitively compelling." A thinker finds a transition primitively compelling, according to Peacocke's definition, if (roughly) she finds it compelling, and does so not based on inference (1992, p. 6). See also section 5.2.3 below.

In chapter 3, I review and develop Boghossian's (1989) classification of theories of content. I agree with Boghossian that there are problems with all the categories of theories he discusses. In particular, he presses objections to all "dispositional" theories of content, including both informational and inferential-role varieties, that depend on ideal conditions or community consensus. I point out, however, that, in light of work developed too recently to be considered in Boghossian's paper, two new categories need to be taken into account. These categories are, first, non-ideal-conditions-based informational theories and, second, ability-based (and thus not purely conditional-behavior-based) inferential-role[19] theories. Fodor's asymmetric-dependence theory is of the first type, and Peacocke's possession-condition theory is of the second. Fodor's and Peacocke's theories are the best representatives of the two categories, and Boghossian's objections do not apply to these categories. So we have reasons, in addition to those mentioned at the beginning of section 1.6, for focusing on Fodor's and Peacocke's theories in chapters 4 and 5.

Further, I go on to point out that Boghossian's objections to dispositional theories of content, while correct as far as they go, do not identify the underlying problem. I offer an intuitive diagnosis. Boghossian's most important objection is that dispositional theories of content that rely on ideal conditions fail because of the holism of belief fixation. Boghossian points out that some dispositions – those that are caused by false background beliefs – are not among the content-determining dispositions. But there is no way to specify (in non-intentional terms) ideal epistemic conditions that will exclude the false background beliefs. I suggest that false background beliefs are just an instance of a broader problem. For example, dispositions (to make judgments, have thoughts, move between mental states) that are the result of the thinkers' desires or fears will not in general be the content-determining ones. Similarly, dispositions that result from non-inferential associations, however systematic, will not in general be relevant to content. I conjecture is that it is certain norms of thought, not dispositions, that determine content. This would explain why certain dispositions seem to be content-determining: dispositions to satisfy the content-determining norms will correspond to content. The conjecture also explains why dispositional theorists – *dispositionalists*, for short – have difficulty finding the resources to identify the apparently content-determining dispositions: those dispositions cannot be identified without employing normative notions.

[19] As we will see (section 5.2.3), by avoiding reliance on conditional behavior, Peacocke also avoids the need to appeal to ideal conditions.

In chapter 4, I criticize Fodor's theory. Standard informational views depend on ideal conditions: they hold that a mental representation's referring to a certain property is constituted by its covarying with that property under ideal conditions. Such theories therefore need there to be ideal conditions under which a mental representation can be triggered only by (or in the presence of) instances of the concept. As Boghossian has shown, however, there is a general problem with this approach: whether a person will token a mental representation depends not only on its content and on the situation, but also on the person's other mental states.

Fodor's theory avoids reliance on ideal conditions. Roughly, it holds that a mental representation refers to a given property if 1) there is a law connecting the occurrence of the representation with the property; and 2) any laws connecting the occurrence of the representation with other properties depend on the first law in a specified, asymmetric way. Fodor officially claims that he aims to provide only sufficient conditions for content. I show that Fodor fails to establish that his sufficient conditions succeed. Only by subtly assuming what it is at issue does he reach the result that if a mental representation satisfied his proposed conditions, it would have the appropriate content. Moreover, I show that he cannot restrict his aspirations to providing sufficient conditions. He needs it to be the case that mental representations that have a particular content will satisfy his proposed conditions. But this turns out to be false: the patterns of asymmetric dependencies between laws are not the ones that Fodor's account specifies.

In chapter 5, I criticize Peacocke's possession-condition theory of content. Peacocke's account makes an important advance over inferential-role accounts that appeal to bare dispositions to make certain inferences. He appeals not only to what mental transitions a thinker finds natural – "primitively compelling" – but also to the explanation – that the thinker finds them compelling "in virtue of their form." I argue, however, that Peacocke's theory is unsatisfactory because it lacks an adequate account of what makes it the case that a thinker finds an inference that is an instance of more than one form compelling in virtue of one form rather than in virtue of another. On Peacocke's account, the form of a mental transition is individuated in content-involving terms. The proposition that one content-involving form rather than another is causally explanatory is too close to what a theory of content aims to explain for it to be taken for granted.

Merely showing that specific accounts each fail for unrelated reasons would leave the possible response that the basic approach is correct but the details simply need to be worked out. My argument that there is a common thread running through the individual problems suggests, to the contrary, that the problems are endemic to the approaches. The accounts seem to lack the resources to account for the way in which content explains our judgments. We thus have reason to suspect that we need a type of account with richer resources at its disposal.

1.8 The second problem: normativity

A theory of content tries to say what makes it the case that thoughts have their particular contents. Leading theorists of content provide surprisingly little in the way of explicit argument that their proposed conditions are constitutive of content. They rely heavily on the intuitive plausibility of the achievement view. One way to support a theory of content is to show that there is an intuitive connection between the proposed condition for content and the normative features of content. The second of the Three Problems is that leading theories of content have little prospect of demonstrating such an intuitive connection and thus of gaining plausibility in this way.

Leading theorists of content typically argue, as I said in section 1.5, that their proposed conditions for content allow room for error. Establishing that a condition leaves room for error is plainly not enough to establish that a thinker's satisfying the condition is what makes it the case that the thinker's thought has a particular content. It is only because the achievement view is taken for granted – and because proposed conditions for content conform to the achievement view – that it could seem plausible even for a moment that solving the problem of error is sufficient to answer the content question.

A version of the silly full-moon example makes the point. Suppose it is proposed that the way in which a person would apply a particular mental symbol during the first full moon of the century determines that symbol's content. And suppose that the person's (possibly counterfactual) full-moon dispositions are in fact sufficient uniquely to determine a particular content for that symbol. We can even suppose that the content that the dispositions determine is the content that we would intuitively assign to the thinker's mental symbol. The proposed condition leaves room for error since it allows that thinkers can be systematically wrong except during the first full moon of the century. But it would be absurd to think that a mental symbol's satisfying the proposed condition is what makes it the case that the symbol has its particular content. We have no reason for thinking that if a symbol satisfied the condition it would have the appropriate content (nor the converse).

No one thinks, on reflection, that showing that a theory of content allows for error is sufficient to establish that the theory is correct. Theorists who focus on the problem of error of course believe that their proposals are otherwise plausible. But we should ask more explicitly what makes a proposed condition for content plausible.

In order to show that satisfying a proposed condition makes it the case that a thought has a particular target content, a theorist needs to make plausible that a thought that satisfied the proposed condition would have all properties that a thought with the target content necessarily has. A thought with a certain content necessarily has certain normative features – correctness conditions, for example. Thus, one thing that would make a proposed condition for content plausible would be an explanatory connection between the condition for content and the appropriate normative features. In section 6.8, I argue that because they presuppose the achievement view, leading theories of content lack such a connection. In principle, there may be other ways to show that any thought that satisfied a proposed condition would have the appropriate normative features – *e.g.*, by demonstrating that the proposal was so successful in accounting for other features of content. In practice, however, leading theories of content have poor prospects for showing that their proposals are plausibly sufficient for content.

1.9 The third problem: incomplete understanding

Even leaving aside normativity, the typical conception of the problem of providing a naturalistic account of content is too narrow. It neglects one part of the data almost entirely: thoughts involving incompletely understood contents. Theorists assume that the problem is to account for a thinker who fully understands a concept, and whose mistakes involving it are therefore a result of, for example, misperception or distraction. I will argue that standard views cannot account for thoughts involving incompletely understood contents, and that the phenomenon cannot be dismissed as insignificant and peripheral.

It is widely taken for granted that, with some kind of exception for contents that are dependent on deference, thinkers cannot have thoughts involving contents they only incompletely grasp. Call this *the demanding view*.

> **The demanding view** (**DV**): A necessary condition for a person's having thoughts involving a particular concept is for the person to have full grasp or mastery of the concept or to defer to someone who does.

DV is implicitly assumed by many who do not explicitly advocate it. I will consider the reasons for this implicit acceptance below. In light of Tyler Burge's well-known anti-individualist arguments,[20] most philosophers would probably recognize that having attitudes involving a concept does not require fully grasping it (and thus, *a fortiori*, that "having a concept" in the sense of being able to have attitudes involving it cannot simply be equated with fully grasping it). It is often supposed, however, that Burge's arguments do not require the distinction to be seen as a clear and fundamental one: an attitude's involving an incompletely understood concept is made possible by the thinker's deferring to others who fully grasp the concept. Such an incompletely understood content is thus derivative of and parasitic on the central case of fully understood contents. To the extent that theorists of content consider incompletely understood contents at all, they assume that they can be accounted for merely by tacking on a proviso about deference.

Although the notion of "fully grasping" a concept is intuitive and frequently used, it will require careful definition in order to avoid falling unwittingly into the assumption that being able to have attitudes involving a concept requires fully grasping it. I will use the terms "full grasp of a concept," "mastery of a concept," and "possession of a concept" interchangeably, as they are frequently used in contemporary philosophical discussions of concepts. I suspect that, on occasion, the distinction between possessing a concept and being able to have attitudes involving it is missed simply because of an ambiguity in the common philosophical use of the terms "possessing a concept" and "having a concept." "Possessing a concept" and "having a concept" are sometimes used to mean fully grasping or having mastery of a concept and sometimes used to mean being able to have attitudes involving a concept. In order to steer clear of this pitfall, I will not use the terms

[20] *E.g.*, Burge (1979, 1986a, 1986b).

"possessing a concept" and "having a concept" to mean being able to have attitudes involving a concept. Instead, I will use the term **having a concept*. I will scrupulously observe the distinction between "possessing a concept" – *i.e.*, having mastery of it – and *having a concept.[21] Thus, in my terminology, anyone who has or can have thoughts involving a given concept *has that concept. But such a person need not fully grasp or have mastery of the concept.

It is worth noting a related terminological danger. It is common to say that someone "uses a concept," when we mean that the person has attitudes involving the concept. If we think of *using* concepts on the model of using vacuum cleaners or lawn mowers, however, we will assume that in order to use a concept, one must possess it. Since we are using the term "possessing a concept" for having mastery of a concept, the result would be confusion. Obviously, without an account of what it is in virtue of which someone "uses" a concept – *i.e.*, has an attitude involving the concept – there is no justification for thinking of using a concept on the model of using a (physical) machine or a tool. We therefore cannot assume on this basis that in order to use a concept it is necessary to possess it.

Intuitively, the abilities of a thinker T with incomplete grasp of a concept *C* are insufficient by themselves to determine that her thought involves *C*. There are, at least superficially, two kinds of cases:[22] cases of partial and cases of incorrect understanding. Suppose that the incompleteness of T's grasp of *C* consists in her having no way to determine for some range of objects whether they fall under the concept.[23] This is partial understanding. Nothing T can do and nothing she knows can differentiate her concept from indefinitely many concepts that differ from *C* solely with respect to their application to the range of objects. Suppose, to take the other sort of case – incorrect understanding – that the incompleteness of T's grasp of *C* consists in her incorrectly taking a range of objects to fall under her concept (or not to fall under it). Again, nothing T can do and nothing she knows can differentiate her concept from the indefinitely many concepts that differ from *C* just with respect to their application to the range of objects.

[21] This introduction of a technical term seems the best option. Both idiomatic terms – "having a concept" and "possessing a concept" – have the same ambiguity. Since concept "possession" is something of a technical term for concept mastery, at least in certain theories, it can serve to mark the other side of the distinction.
[22] It is plausible that ultimately there is no important distinction between the two kinds of cases.
[23] For simplicity, I explain the distinction in terms of the thinker's understanding of what falls under the concept, but analogous points could be made for the thinker's understanding of the connections between concepts.

In the case in which T wrongly thinks that a certain range of objects falls under *C*, it would be odd to say that T fully grasps that concept *D* that differs from *C* only in the respect that it correctly applies to the range of objects. For example, in Burge's famous ARTHRITIS example, the thinker who believes he has arthritis in his thigh has an ability or understanding that is precisely the same[24] as the ability or understanding of one who fully grasps the concept THARTHRITIS, where that concept applies to degenerative conditions outside the joints as well as in the joints. Yet it may seem strange to say that the original thinker, T, fully grasps THARTHRITIS. The concept THARTHRITIS does not figure in his thoughts at all; he does not *have that concept, so it is odd to say that he fully grasps it.

In sum, a thinker who fully grasps a concept has something – what that something is will depend on one's theory of what constitutes full grasp of a concept – that uniquely determines that concept. Loosely speaking, the thinker has a sensitivity to whatever it is that individuates the concept. (The point is that this is a *necessary* condition for concept mastery, not that it is a sufficient one.) In contrast, the abilities of a thinker with incomplete grasp of a concept are intuitively insufficient by themselves to determine that her thought involves that concept.[25] Nothing the thinker can do – none of her practices or dispositions – and nothing she knows can by itself determine that she is using the incompletely grasped concept, as opposed to an indefinite number of concepts that differ from it in some aspect that she does not grasp.[26]

1.10 The Three Problems form a mutually supporting case

The Three Problems combine to form a powerful case against currently standard theories of content. The arguments concerning the Problems are mutually supporting in specific ways. For example, I

---

[24] It would obviously be question-begging to object that the abilities are different because one is an incorrect understanding of one concept, while the other is a correct understanding of a different concept. Also, I am here assuming a notion of ability according to which the thinker's ability is not socially individuated (and is independent of his deference to others).

[25] To say that a certain ability (or abilities) is insufficient to determine a particular concept is not, of course, to say that there is no function from abilities to concepts that maps the ability in question to that concept. Rather, the idea behind the achievement view is that when a thinker's exercise of her mastery of a concept *C* makes it the case that her thought involves *C*, her mastery of the concept determines *C* by a particular, straightforward mapping. See section 2.3.1 below. Thus, to say that a certain ability is insufficient to determine a particular concept is to say that the ability does not map to the concept by the relevant straightforward mapping.

[26] There is a loose use of the term "ability" on which it is trivial that someone who *has a concept has an ability that is sufficient to determine the concept. For by hypothesis he has an "ability" to have thoughts involving that concept. In order for the achievement view (as well as DV) to be a substantive claim, we need a restriction on what counts as an ability. Intuitively, an ability has to be a causal power; for example, in the relevant sense, being subject to a requirement and thus being "able" to violate it need not be an *ability*. In section 2.2, I introduce "the achievement restriction" to address this issue.

a is trivially not to be trivially false depends on understanding there is a way to trivialize the idea that that if will argue that the theories' failure on their own terms makes the normativity problem more threatening.[27] For a different reason, the theories' failure on their own terms bolsters my argument about the problem of incomplete understanding. An attempted dismissal of putative attributions of incompletely understood contents (that are not dependent on deference) as merely imprecise ways of talking is less compelling to the extent that it is not motivated by a successful theory.[28] So, by showing that leading theories of content are unsatisfactory on their own terms, we can reduce the force of theoretically driven objections to putative cases of incomplete understanding. Most importantly, the Three Problems in fact have a common source. In the next chapter, I suggest that all three stem from a very basic presupposition of leading theories of content about what it is in virtue of which an attitude's content involves a particular concept. Thus, the Three Problems are not just a miscellaneous collection, but symptoms of a deeper flaw.

one understands the notion of an ability too loosely, one might think that it is trivial that someone who *has C

[27] See section 6.7 below.

[28] See sections 2.7, 6.8 below.

Chapter 2: The Achievement View and the Responsibility View

This chapter introduces two very general views about what it is in virtue of which a thinker's attitude involves a particular concept. I argue that the first view – the achievement view – is presupposed by most currently standard theories of content, and that it is the source of their difficulties with the Three Problems. I show that the achievement view seems almost designed to produce the Three Problems. In contrast, the other view – the responsibility view – offers good prospects for avoiding them. The plot thickens, however. Theories based on the achievement view have a variety of options for addressing the Three Problems, and responsibility-based theories face difficulties of their own. After developing the distinction between the two views, I briefly set out the connections between the two views, on the one hand, and the Three Problems, on the other. I then consider refined versions of the achievement view and discuss how they are likely to fare in comparison with the responsibility view.

2.1 Introducing the achievement and responsibility views

2.1.1 The common source of the Three Problems: the achievement view

At a very general level, most currently standard theories of content take the same approach to the content question. They presuppose the achievement view. It can be difficult to perceive this fundamental similarity because of differences between the theories on other dimensions. Two dimensions, which we remarked on in chapter 1, are particularly relevant in this connection. Differences on these dimensions tend to produce theories that have very different overall structures and feels. First, theories of content take different positions on what individuates content. Second, theories can be pitched at a personal or subpersonal level. (In section 1.8, we noted that Peacocke's and Fodor's theories contrast on both dimensions.)[1]

[1] A third dimension is also relevant. Theories can presuppose different views about the nature of the attitudes. For example, on one such view, having a belief (say) is being in a certain relation (one that corresponds to believing, as opposed to fearing, intending, etc.) to a mental representation. On a contrasting view, having a belief is being in a certain unitary, perhaps functionally individuated state.

The potential for differences on these dimensions makes it difficult to state the achievement view in a way that is fully general, yet concrete and precise. A canonical characterization of the view would have to be stated in terms that are neutral with respect to positions on concept individuation and with respect to personal versus subpersonal levels. Terms that are neutral in these respects yet still precise are not available, however.

I will therefore begin with an intuitive, and very abstract, characterization of the achievement view and contrast it with an alternative, which I call the *responsibility view*. These intuitive characterizations will be cast at the personal level for clarity of exposition, but will be sufficiently abstract to be neutral on the question of how concepts are individuated. After offering some clarifications, I will make the discussion more concrete by providing examples of different forms that the achievement view can take, given different values on the two dimensions. To make the contrast as sharp as possible, I will begin with a very crude version of the achievement view. Later in the chapter (sections 2.7-2.9), after examining why the achievement view tends to predispose a theory of content to the Three Problems, I will consider refinements of the achievement view that mitigate that tendency.

At the simplest level, the achievement view holds that what it is in virtue of which a thinker *has a particular concept is the thinker's having a particular disposition. The responsibility view holds that it is the thinker's being subject to a standard. Roughly, the distinction is between being disposed to do something and being required to do something.

To see why the achievement view is a natural answer to the content question, let us step back and consider what an answer to the content question must accomplish. It must provide an account of the way in which content attributions characterize causally efficacious states of a flesh-and-blood human being in terms of abstract objects – concepts and contents. The question of what it is in virtue of which one's thought involves *that* concept asks for the feature common to different thoughts involving the same concept. Our ordinary talk encourages the idea that the thinker has something, which he then exercises or deploys in the particular instance. Someone who has thoughts involving a particular concept is said to "have" or "possess" that concept, and to "use" it on a given occasion. So we might ask what is it that a thinker has and then exercises? What is it to *have a concept?

A plausible approach is to take each concept to be associated with a condition on a thinker, *i.e.*, a condition specifying that a thinker do something. A concept's associated condition might be, for example, that a thinker discriminate a certain worldly property, classify objects in a certain way, make certain inferences, or have a state that is (causally) linked to certain other states. A very natural picture, which most current theories of content take for granted, is that *having a concept is having a disposition to satisfy the concept's associated condition, and having a thought involving the concept is deploying or exercising that disposition. This is the achievement view. Slightly more precisely:

> **The achievement view**: a thought involves one concept rather than others in virtue of the thinker's exercising a disposition to satisfy that concept's associated condition.[2]

From a certain perspective, the achievement view may seem inevitable, with the result that we can miss the fact that it is a substantive choice, a choice of only one from among several alternatives. When we consider how to answer the content question, the essence of the problem may seem to be how a person can "have" and "use" an abstract object. Suppose we hit on the plausible answer that concepts are associated with conditions on thinkers and thus with dispositions. From this idea, the answer to the content question may seem to follow immediately: a thinker has an attitude involving a particular concept in virtue of his having and deploying a disposition to satisfy the concept's associated condition. But this overlooks the fact that, even granted a concept's association with a condition, there is more than one possible view of the relation between a thinker who has an attitude involving a particular concept and that condition.

2.1.2 The responsibility view and the descriptive/normative contrast

The responsibility view offers an alternative picture of the relation between thinker and condition. The thinker could be *subject to*, or *responsible to*, the condition. In the sense I have in mind, to say someone is subject to a condition is to say that the condition is a standard of correctness that is applicable to the thinker. A thinker's being subject to a condition is a presupposition of the thinker's being correct or incorrect in conforming to or failing to conform to the condition.

> **The responsibility view**: a thought involves one concept rather than others in virtue of the thinker's being responsible or subject to standards distinctively imposed by that concept.

[2] If the achievement view is to avoid vacuity, the condition must not be such that it is trivial that a thinker's exercising the corresponding disposition is what it is in virtue of which a thinker has an attitude involving a particular concept. For example, the condition cannot be that the thinker have an attitude involving the concept in question. I discuss other restrictions on the condition below.

Thus, according to the responsibility view, attributing an attitude involving a particular concept is a way of typing a thought in terms of the standards to which the thinker is responsible, as opposed to the dispositions she activates.[3]

The basic contrast between the achievement and responsibility views is a contrast between descriptive and normative. The achievement view holds that a person's belief involves a particular concept in virtue of something descriptive; the responsibility view holds that it is in virtue of something normative. Two questions arise. First, given a minimal naturalistic requirement of supervenience of content on descriptive facts, how is the distinction sustainable? Second, what does the distinction amount to? In particular, what descriptive constraint must a position satisfy to count as an achievement view? I address each point in turn.[4]

2.1.3 The descriptive/normative contrast and multiple realizability

A naturalistic responsibility view could consistently maintain that the content of attitudes supervenes on descriptive facts but that what it is in virtue of which an attitude has a particular content is something normative. As functionalists pointed out against the type-type identity theory of mind, multiple realizability undermines the type-type identity of a mental property and a realizing physical property, but is consistent with supervenience. Analogously, if normative facts about responsibility to standards can be realized in indefinitely many ways, those normative facts may supervene on descriptive facts about the thinker and the world without its being the case that being responsible to a standard is identical to some descriptive property.

The point can be put another way. If the responsibility view is true and naturalistic, there will be sets of descriptive facts, perhaps indefinitely many, that realize thinkers' being responsible to a particular concept's standards. Suppose a theory that presupposes the achievement view succeeds in offering a specification, perhaps highly disjunctive, that captures one or more of these sets. It

[3] It is too undiscriminating to say merely that the thinker is subject to the concept's standards. For example, one thought may involve multiple concepts, and we have many thoughts. A qualification is needed to the effect that the thinker is subject to the relevant standards *with respect to the thought in question*. Roughly, he is required to conform that attitude to the relevant standards. This is still too crude, however. The thinker is subject to the relevant standards with respect to a particular aspect of the thought – intuitively, that which corresponds to the concept's role in the structure of the attitude's content. The details of a full account would depend on views about the nature of the attitudes and of their compositional structure, on the level (personal or sub-personal) at which the account is pitched, and so on. Analogous points apply to the achievement view. That is, a full account of the achievement view would have to specify that a thinker exercises the relevant ability not merely in general but in thinking the attitude in question, and so on.

[4] It is sometimes claimed that "ought implies can" – that is, that a person cannot be responsible to a standard without being able to satisfy it. In section 2.9, I respond to an objection based on this claim.

would then have specified one or more sufficient conditions for content (for an attitude's involving the concept in question). Does this show that the distinction between the two views cannot be maintained because a naturalistic responsibility view reduces to an achievement view? The answer is a familiar one. The hypothetical theory would not explain what the different sets of descriptive facts had in common that made them sufficient for an attitude's involving a particular concept, and it would not be able to predict what other sets would be sufficient. It would miss the crucial generalization (see section 6.5.1 below). If a responsibility-based theory can say what the different sets of descriptive facts have in common but an achievement-based theory cannot (because it lacks the resources to refer to the normative property of being responsible to a standard), the responsibility view has a potential for explanatory power that the achievement view lacks.

2.2 The achievement restriction

As we have so far defined it, the achievement view holds that a thinker has an attitude involving a particular concept in virtue of the thinker's exercising a disposition to satisfy the concept's associated condition. This definition is not sufficient to maintain the intuitive descriptive/normative contrast between the achievement and responsibility views. For example, suppose it is suggested that the relevant condition is that the thinker be responsible to the standards distinctively associated with the concept. This suggestion would collapse the distinction between achievement and responsibility views by encompassing the responsibility view within the achievement view.[5] In order to maintain the basic contrast between the achievement and responsibility views, we need to restrict what counts as a condition. Intuitively, the condition must be such that satisfying it is an achievement. In other words, something counts as a condition in the relevant sense only if it specifies that a thinker *do* something. It cannot specify that the thinker be *required* to do something. We can say that the condition must be descriptive in the sense that whether a thinker satisfies it is a descriptive, not a normative fact. Call this the *achievement restriction*. When we consider more sophisticated versions of the achievement view, the achievement restriction will need to be adapted appropriately.

[5] This collapse would be merely terminological. The substantive distinction would remain, but it could no longer be phrased in terms of exercising a disposition to satisfy a condition versus being responsible to a condition. Rather, it would be a distinction between types of dispositions (or types of conditions).

2.3 Fleshing out the achievement and responsibility views

Before turning to the relation of the achievement and responsibility views to the Three Problems, I have two items of business: 1) to develop the idea of a condition *associated* with a concept; 2) to clarify the achievement and responsibility views by examining some forms the achievement view can take.

2.3.1 The condition associated with a concept

I have thus far characterized the achievement view in a way that is vague about the relation between a concept and its associated condition. The starting point is that, by hypothesis, a thinker's exercising a disposition to satisfy the condition is what it is in virtue of which the thinker's attitude involves this particular concept as opposed to any other. To put it another way, the thinker's attitude involves the concept in virtue of the fact that the disposition she is exercising is a disposition to satisfy this condition. That is, it is not something else about the thinker – for example, the way in which she exercises the disposition – but the fact that it is a disposition to satisfy this condition that determines that the thinker's thought involves this concept.

It follows that the disposition must be sufficient uniquely to determine the concept in question. Since the disposition is defined in terms of its associated condition, the condition must be sufficient uniquely to determine the concept.

Of course, a concept's associated condition is not just any condition that is sufficient – by some mapping or other – to determine the concept. A concept's associated condition is based on whatever individuates the concept. Suppose concepts are individuated at the level of reference. Then a concept's associated condition specifies that a thinker must apply a mental representation to the reference of the concept. Suppose, alternatively, that concepts are individuated at the level of inferential role. Then a concept's associated condition specifies that the thinker must make the transitions between mental states that are canonical for the concept in question. Since a concept's associated condition is based on whatever it is that individuates the concept, and since the concept's associated condition must be sufficiently uniquely to determine the concept, it follows that the concept's associated condition will provide an account of what individuates the concept, or, at least, that such an account can be derived from that condition.

This notion of a concept's associated condition is needed in order to be true to the spirit of the achievement view. The idea behind the achievement view is that a concept is distinctively associated with a success in a particular mental task, and that *having the concept is being disposed to succeed in that task. So the achievement view is *not* the view that *having a particular concept is having a disposition that, by some mapping or other, corresponds to that concept's identity condition. Rather, it is the view that *having a concept is having a disposition that corresponds, via the concept's associated condition, to the concept's identity conditions.

According to the responsibility view, concepts impose requirements on thinkers; in particular, to say that a particular concept figures in the content of a given attitude is to say that the thinker is subject to standards distinctively associated with that concept. Thus, the responsibility view, like the achievement view, relies on a condition on a thinker that is sufficient uniquely to determine a concept. As I will illustrate in the next section, it too derives such a condition from whatever individuates the concept.[6]

2.3.2 Concrete instantiations of the achievement and responsibility views

We have so far discussed the achievement and responsibility views at a very abstract level in order to remain neutral with respect to various choices that a theory of content must make. As I remarked, I have assumed a personal-level account, but have remained neutral with respect to different positions on what individuates concepts. We can now consider more fleshed-out versions of the achievement and responsibility views, given different choices about reference-based versus conceptual-role-based individuation of concepts. This will help both to clarify the two views, and to support my suggestion that standard theories of content presuppose the achievement view. I will continue to frame the accounts at the personal level. At the subpersonal level, talk of what the thinker is disposed to achieve would translate into talk of the causal properties of subpersonal states of the thinker. And

[6] I am distinguishing between a "condition" that is merely a condition – in the case of the achievement view – and a condition that is also a "standard" – in the case of the responsibility view. Roughly, the point is to mark a distinction between a condition to which one is not responsible and a condition to which one is responsible. I use the term "standard" in the latter but not the former case. One who fails to satisfy a mere condition is not incorrect, in contrast to one who fails to satisfy a standard to which he is responsible. The use of the term "condition" serves as a reminder that we cannot assume that a thinker is responsible to a condition merely because he satisfies it. In these terms, someone who peels an apple in a way that produces an unbroken spiral of peel measuring over 10 meters does not thereby satisfy a standard (since he is not subject to any requirement that he produce a ten-meter peel) but he does satisfy a condition, *viz.* the condition that he produce a ten-meter peel.)

talk of a thinker's being responsible to a standard would translate into talk of a subpersonal state's being responsible to a standard,[7] or of the thinker's being responsible to a standard with respect to a given subpersonal state.

1) Let us begin with theories that hold that concepts are individuated by their reference. The natural condition for an achievement view is that the thinker apply[8] a mental representation to (or make appropriate judgments of) instances of the property that is the concept's reference. Thus, a typical version of the achievement view holds that an attitude involves a particular concept in virtue of the thinker's deploying a certain disposition – a disposition to apply a mental representation to instances of the appropriate property. Suppose a thinker judges that a particular table is C. A question arises as to what makes it the case that the content of the particular judgment is that the table is square, as opposed to that the table is rectangular. The achievement view we are considering would say that the thinker is disposed to apply *C* to square things but not to non-square things, including non-square rectangles. What makes it the case that the judgment involves the concept SQUARE is that it is that disposition, as opposed to, say, the thinker's disposition to apply a concept to rectangular things, that he exercises in making the judgment.[9]

In contrast, on the responsibility view, the judgment's content involves SQUARE in virtue of the fact that the thinker, in making the judgment, is responsible to a standard according to which it is correct to make the judgment of square objects and incorrect to make the judgment of non-square objects. More precisely, a concept *C* that is involved in a thinker's attitude is the concept SQUARE in virtue of the thinker's being subject to a standard according to which it is correct to apply *C* to square objects and incorrect to apply it to non-square objects.

2) Suppose now that concepts are individuated by inferential role. The natural way to derive a condition appeals to the idea of an isomorphism between the network of inferential relations between propositions and the network of causal relations between states of a person. Thus, a typical condition would be that the thinker move between states in a pattern that mirrors the abstract conceptual relations between propositions. Let P and Q stand for propositions. On the present

[7] The awkwardness of the notion of subpersonal states being responsible to standards suggests that the responsibility view will be more apt as a personal-level account; a further, subpersonal account could then be given of what realizes a thinker's responsibility to a standard.
[8] As I discuss in chapters 3 and 4, informational theorists tend to help themselves to the notion of an *application* of a concept, though they are, strictly speaking, not entitled to do so. I'll waive this problem at this point.
[9] Section 2.6 below considers moves that the achievement theorist can make to flesh out the idea that it is one

version of the achievement view, a concept *D* involved in a thinker's judgment that P D Q is the concept AND in virtue of the fact that the thinker, in making the judgment, exercises a disposition that corresponds to the inferential role of AND – a disposition to move from the judgment that P and the judgment that Q to the judgment that P D Q, to move from the judgment that P D Q to the judgment that Q, and so on.

The responsibility view takes the inferential role of a concept as a standard. Thus, the relevant standards require thinkers (if they consider the relevant questions) to make certain mental transitions. For example, a concept *D* involved in a thinker's judgment that P D Q is the concept AND in virtue of the fact that the thinker is subject to a standard according to which it is correct (if the thinker considers the question) to move from the judgment that P and the judgment that Q to the judgment that P D Q, to move from the judgment that P D Q to the judgment that Q, and so on.

These examples make plain that the achievement and responsibility views are not themselves accounts of how concepts are individuated, and are consistent with different positions on that issue. Given your favorite position on concept individuation, the achievement and responsibility views offer general positions on what it is to have an attitude involving a concept. Thus, the contrast between the achievement and responsibility views makes plain a distinction that is often neglected. An account of concept individuation is not an account of what it is for an attitude to involve a particular concept and does not imply such an account. Further, since *having a concept is simply being able to have an attitude involving the concept, an account of what a concept is does not imply an account of what it is to *have a concept.[10] Of course, an account of what it is in virtue of which an attitude involves a particular concept – an answer to the content question – cannot be independent of what makes that concept the concept it is. In particular, as we have seen, a concept's associated condition, which plays a central role in both achievement-based and responsibility-based answers to

---

disposition and not the other that is being deployed.

[10] Jerry Fodor suggests that "having" a concept is utterly straightforward:

> It's a general truth that if you know *what an X is*, then you also know *what it is to have an X*. And ditto the other way around. This applies to concepts in particular: the question what they are and the question what it is to have them are logically linked; if you commit yourself on one, you are *thereby* committed, willy nilly, on the other. Suppose, for example, that your theory is that concepts are pumpkins. Very well then, it will have to be a part of your theory that having a concept is having a pumpkin. And conversely . . . (1998, p. 2).

This reasoning is perilous, because it encourages the already tempting equivocation on the term "having." *Having a concept may not be having – in the sense in which one "has" a pumpkin – any thing at all; it may be being in some kind of complex state describable in terms of an abstract object –being in a relation, if you will, to an abstract object. Compare what it is to have thirty years of age or two extra inches on one's vertical leap.

the content question, must be sufficient to determine that concept, in the sense we discussed in the previous section.

One salutary effect of distinguishing an account of concept individuation from an account of the nature of content is that it allows objections to each kind of account to be treated separately. For example, I will suggest that an objection commonly raised against the view that concepts are individuated by inferential role is more plausibly taken as an objection to an achievement view.

2.4 The (crude) achievement view and the Three Problems

In this section, I will show that the achievement view seems almost designed to produce the Three Problems. Proponents of standard theories of content will no doubt be impatient with this discussion because it is based on such a crude version of the achievement view; it may seem that I am knocking down a straw man of my own creation. The point of sticking with the crude version of the achievement view at this stage is to make as plain as possible why the basic approach of the achievement view predisposes a theory that adopts it to encounter the Three Problems. In the following section, I will consider ways in which the achievement view can be refined to address those problems.

2.4.1 The problem of failure on their own terms

First, the crude version of the achievement view blatantly has a problem allowing for error. For the achievement view, *having a concept is having a disposition to satisfy a condition associated with the concept. As we have seen, the condition is typically that the thinker discriminate the concept's reference (given a reference-based position on concept individuation) or make the mental transitions that correspond to the concept's defining inferential role (given an inferential-role-based position). A disposition to satisfy such a condition is, loosely speaking, a disposition not to make mistakes of a certain kind – roughly, mistakes with respect to what is essential to the concept. Understood crudely, the achievement view therefore might seem to imply that a thinker cannot make such mistakes, at least not systematically.

Take a reference-based version of the achievement view. If a thinker judges that something is a penguin, he cannot be (systematically) wrong since his judgment would not involve the concept PENGUIN unless he were deploying a disposition to apply a mental representation to penguins and not to non-penguins. To put the point from the opposite direction, suppose the thinker systematically judged that gulls were penguins. His disposition would in that case not be a disposition to apply a mental representation to penguins, but to apply it to (say) penguins or gulls, as opposed to other birds. Therefore, according to the achievement-based view, the thinker's judgments would not involve the concept PENGUIN after all; they would not be mistaken judgments that gulls were penguins, but correct judgments that gulls were (say) penguins or gulls.

Consider instead a simple inferential-role-based achievement view. On such a view, it seems that a thinker cannot mistakenly infer from the fact that something is an overstuffed seat for one that it is a sofa. The judgment does not involve the concept SOFA unless the thinker has exercised a disposition to recognize that if something is a sofa, it is a seat for more than one person.

2.4.2 The normativity problem

The achievement restriction limits the resources of the achievement view in a way that makes it difficult for an achievement view to explain the normativity of content. Consequently, a proponent of an achievement-based theory – "an achievement theorist," for short – will encounter difficulty in establishing that her achievement-based proposal is constitutive of content. For we won't be convinced that a judgment is the judgment (say) *that's a bear* if we are not convinced that the judgment's correctness depends on whether the demonstratively presented object is a bear. In other words, as an epistemic matter, the achievement theorist encounters a difficulty in demonstrating that her theory is correct, if there is no explanatory connection between a thinker's having a disposition to satisfy a concept's associated condition and the thinker's being required to satisfy that condition.

The achievement view holds that what it is in virtue of which an attitude involves a particular concept is the thinker's exercising a disposition to satisfy a condition. The achievement restriction is that the condition must be descriptive, in the sense that it is a descriptive fact whether a thinker satisfies it. The restriction implies that whether a thinker is exercising a disposition to satisfy the condition is a descriptive fact. Thus, the achievement view attempts to account for content solely in terms of descriptive facts.

Content is normative in the sense that an judgment's involving a certain concept determines standards of correctness for the judgment. Since the achievement view is limited to descriptive facts, it faces a difficulty in explaining how the facts in virtue of which an attitude involves a particular concept can determine standards of correctness. The difficulty is particularly acute *because the descriptive facts to which the achievement view appeals are precisely the facts of the thinker's being disposed to satisfy the standards of correctness associated with the concept in question*. In general, being disposed to satisfy a standard does not make one subject to, or responsible to, that standard.

Suppose that an achievement theorist showed that a thinker's having a disposition to satisfy a concept's associated condition tended to correlate with the thinker's *having that concept, and that a thought's involving a particular concept tended to correlate with the thinker's exercising the disposition. Perhaps, given such a showing, the achievement theorist might be able to convince us that it was *in virtue of* the thinker's having the disposition that the thinker *had the concept in question. (Although, even then, as I discuss in chapter 6, we could think that the achievement theorist had the explanatory direction the wrong way around.) But, as I will argue in chapters 3, 4, and 5, achievement theorists have not established such a correlation; indeed, they have not even attempted as much. Instead, they argue merely that their proposed conditions are sufficient to determine content, and that their theories allow for the possibility of error. Even if they were successful in such attempts, however, they would fall short of convincing us that it is in virtue of *having a disposition to satisfy* the condition associated with a concept that a thinker *has a concept. For we would have no reason to think that a person who had such a disposition would therefore be subject to the condition. And if we were not convinced that the person would be subject to the condition, we would not be convinced that the person had a thought involving the concept in question. (Suppose it is objected that the person would be subject to the condition because the person's thought would involve the concept in question. This objection begs the question since what is at stake is precisely whether the person's thought would involve the concept in question.)

If an achievement-based theory specifies concepts' associated conditions in ways that ineliminably[11] presuppose thoughts with particular contents, it will not have the potential to provide

[11] Some conditions may presuppose thoughts with particular contents as long as those contents themselves (or others on which they depend, etc.) have conditions that do not presuppose conceptualized contents. That is, the chain of dependencies must end with concepts whose conditions do not presuppose conceptualized contents.

an explanatory account of what it is in virtue of which thoughts have particular contents. Non-circularity considerations exclude this possibility.

It might be argued that an achievement view could characterize the content-determining disposition in a way that presupposes propositional attitudes, but not their contents. Such a view would have to maintain that the nature of attitudes could be elucidated independently of their having any particular contents. For example, what it is to be a belief could be explained in terms of its relations to desires, intentions, and actions in a way that abstracted from the particular contents of those attitudes. The challenge for such an achievement view is to explain the normativity of content solely in terms of the normative aspects of attitudes.[12]

2.4.3 The problem of incomplete understanding

The source of the achievement view's problem in allowing for incomplete understanding is closely related to the source of its problem in allowing for error. The problem of incomplete understanding is even more intransigent, however.

On the achievement view, a thinker cannot have an attitude involving a particular concept unless he has a disposition to satisfy a condition associated with the concept. Further, that condition must be sufficient to yield an account of what individuates the concept; otherwise, the thinker's having and exercising that disposition could not be what it is in virtue of which the thinker's attitude involves that concept. Consequently, the typical condition is that the thinker discriminate the concept's reference (given a reference-based position on concept individuation) or make the mental transitions that correspond to the concept's defining inferential role (given an inferential-role-based position). A thinker with incomplete understanding of a concept will not be disposed, except perhaps by accident, to satisfy such a condition.

Moreover, in our discussion of incomplete understanding in section 1.9, we considered an intuitive argument that a thinker who incompletely grasps a concept will not in general have dispositions sufficient to determine that her thought involves that concept. We can briefly review the argument. A thinker who incompletely grasps a concept will lack an understanding of some aspect of the concept or will have an incorrect understanding of that aspect. (We called these two cases

[12] It is a difficult question whether presupposing thoughts but not their contents is a legitimate option for a theory that aims to answer the content question. Arguably, when we use the notion of a belief, we presuppose what it is for a belief to have a content. But perhaps we can understand what it is for a belief to have a content without having an account of what it is in virtue of which a belief has its content. So we can use the notion of a

partial and *incorrect* grasp of the concept respectively.) There will always be some concept *D*, however, that differs from *C* only in the aspect that the thinker does not grasp or incorrectly grasps. Suppose that concepts are individuated by their reference. In that case, a thinker who partially or incorrectly grasps a concept *C* will not be able reliably to determine for some range of objects whether they fall under the concept. By hypothesis, however, *C* and *D* differ solely in their application to that range of objects. It follows that the thinker's dispositions will not be able to determine that the thinker's attitude involves *C* rather than *D*. Similarly, if concepts are individuated by inferential role, a thinker who partially or incorrectly grasps a concept will not be reliably disposed to make certain of the concept's defining transitions. *C* and *D* again will differ precisely with respect to these transitions, with the consequence that the thinker's dispositions will not be able to determine that the thinker's attitude involves *C* rather than *D*.

If this argument is right, a thinker with incomplete understanding of a concept (whether partial or incorrect) will not have dispositions sufficient to determine that a thinker's attitude involves that concept. Therefore, the thinker's having and exercising a disposition could not make it the case that his attitude involves an incompletely understood concept. The achievement view thus encounters a problem accounting for attitudes involving incompletely understood contents. The problem is deeper than the problem of error because, as we will see, an achievement view may be refined to allow for a distinction between, on the one hand, what a person actually does or would do under counterfactual conditions (conditional behavior), and, on the other, what ability she has. The latter is the underlying ground that, along with interfering factors, is causally responsible for what she does or would do. In Chomsky's terms, that is, an achievement view can allow for a *performance/competence* distinction. Such a distinction will be of no help, however, in allowing for attitudes involving incomplete understanding. If a thinker genuinely lacks full grasp of a concept, the problem is not merely at the level of performance.

2.5 The responsibility view and the Three Problems

In sections 2.6-2.8, I will consider ways in which theories that are committed to the basic shape of the achievement view can attempt to address the Three Problems. I also compare the prospects of

belief without presupposing an answer to the content question.

such refined versions of the achievement view with those of the responsibility view. Before turning to that more sophisticated evaluation of the two views, I will sketch the outlines of the responsibility view's advantages in addressing the Three Problems.

2.5.1 The responsibility view and the problem of allowing for error

As we have noted, in order to have the potential to capture the data of attributions of content, a theory must be able to allow for error. (In fact, I have suggested that current theories of content do not do much more to show that they capture the data than attempt to account for error.) The responsibility view is well positioned in this respect. It holds that a thinker's attitude involves a particular content in virtue of the thinker's being responsible to standards derived from an account of what individuates the concept. It cannot, and need not, be assumed that a thinker who is responsible to a requirement will reliably satisfy that requirement (though some accounts of what makes a thinker responsible to a standard might have that implication).[13] The responsibility view therefore does not imply that a thinker who has an attitude involving a concept will be disposed to use it correctly.

The responsibility view is only a first step toward a full account of content. Without an account of what makes a thinker responsible to a concept's standards, it does not make sense to ask whether we have captured the actual data of content attributions. In chapter 7, I say something about the prospects for an account of responsibility to standards.

2.5.2 The responsibility view and normativity

The responsibility view is well-suited to account for the apparently normative aspect of content. It holds that a thinker has a thought involving a particular concept in virtue of the obtaining of certain normative facts: the thinker's being subject to certain standards. It might be objected that the view is *too* well-suited to the task: it accounts for the normative aspect of content by stipulating it or taking it for granted, and thus has no explanatory potential. There are really three different possible objections here. 1) The responsibility view is vacuous because it simply re-poses the question that theories of content are designed to answer. 2) The responsibility view is not explanatory because it takes normative facts for granted. 3) The responsibility view is circular because the normative facts it relies

[13] In section 2.9, I will consider an argument to the effect that being responsible to a standard necessitates having an ability to satisfy it.

on themselves depend on intentional facts. I respond to these objections in chapter 7.

2.5.3 The responsibility view and incomplete understanding

The source of the responsibility view's potential to account for incompletely understood contents is essentially the same as the source of its potential to account for error. That is, the responsibility view does not make an achievement or a potential achievement – satisfying a condition or being disposed to satisfy a condition – any part of having an attitude involving a concept. As I argued in discussing the achievement view's difficulties in allowing for incomplete understanding, a thinker who is disposed to satisfy a concept's associated condition will not, at least in general, have incomplete understanding of a concept. But the responsibility view, unlike the achievement view, does not imply that in order to have attitudes involving a concept, a thinker must be disposed to satisfy the concept's associated condition.

I mentioned in passing that the achievement view can attempt to make room for error by introducing a competence/performance distinction, but that it cannot take this tack with respect to incomplete understanding because a thinker with incomplete understanding falls short in competence as well as performance. The responsibility view has no such difficulty because being responsible to a standard does not imply even having an ability to satisfy that standard.[14] In sum, the responsibility view allows for incomplete understanding by holding that a thinker can be responsible to a standard that she lacks the ability to satisfy.

We have now seen in brief why the achievement view predisposes a theory to be afflicted by the Three Problems and why the responsibility view is *prima facie* well suited to avoid them. I now turn to a more in-depth discussion of the prospects for the two views. The achievement view, in more sophisticated form, has a number of options for addressing the Three Problems. And the responsibility view has some questions to answer. For each Problem, I compare the basic lines of argument available to proponents of the achievement and responsibility views.

2.6 Retreating from the crude achievement view to leave room for error

As we have seen, simple informational and inferential-role theories are notoriously afflicted with the

[14] See section 2.9 and chapter 7.

problem of error. Thinkers who are by all accounts competent users of a concept do not in fact do what the crude understanding of the achievement view would have them do. Informational theories have the problem that thinkers fail to apply a concept to objects in its extension (or do apply a concept to objects not in its extension), and inferential-role theories have the problem that thinkers fail to make the inferences that are (putatively) definitive of the concept.

Both kinds of theories retreat from the crude version of the achievement view by seeking characterizations of the thinker's dispositions that are consistent with actual failure to perform. On a very simple dispositional notion, having a disposition to do something implies that one reliably does so. This rules out the possibility of systematic error. But there may be dispositional or related notions that do not carry this implication. We want to be able to claim that a thinker is disposed to do something, but regularly fails to do so. What is needed is a principled way of distinguishing such a case from a case of a thinker's having a different disposition.

There are a number of options. We can see them as falling into two basic categories. First, we can restrict consideration to some specified class of occurrences. Second, we can shift the focus from occurrences to their ground or explanation. Each category has personal and subpersonal options. In the first category, a personal-level option will typically be framed in terms of a thinker's conditional behavior; a subpersonal option will be in terms of covariation between mental representations of some kind and worldly objects or properties. In the second category, a personal-level option will typically be in terms of an ability; a subpersonal option may use a notion such as that of a mechanism. We can name the categories in terms of their personal-level options: the *conditional behavior* category and the *ability* category.

In the conditional behavior category, the most straightforward option is to attempt to specify ideal[15] conditions under which a thinker cannot fail to perform correctly (or at least to be systematically disposed to do so). We can then restrict consideration to conditional behavior under the specified conditions. An ideal-conditions-based theory holds that for the thinker to *have a

---

[15] I will use the term "ideal conditions" to refer to any conditions under which the thinker cannot fail to perform correctly. This terminology glosses over a distinction between two different kinds of conditions. First, there are those that are ideal in a way that is independent of the fact that they are conditions in which the thinker cannot fail to perform correctly. For example, we could rely on epistemically ideal conditions or conditions that are ideal in a sense that is grounded in evolutionary theory. Second, there are conditions that are not independently ideal. Fred Dretske's learning conditions are an example (Dretske 1981, pp. 193-95). The idea in this second kind of case is not that the conditions are content-determining because they are optimum for perfect performance, but that they determine what counts as perfect performance because they are content-determining.

concept is for it to be the case that the thinker would satisfy the concept's associated condition under the specified conditions, not that the thinker does so under actual or typical conditions. Both informational and inferential-role theories can appeal to ideal conditions, but this option is at least superficially less attractive for an inferential-role theory because it is more difficult to make sense of the idea of conditions that are ideal for mental transitions than of the idea of conditions that are ideal for perception.

Another type of conditional-behavior strategy is to try to specify a property that is common to the content-determining dispositions, other than their being dispositions under ideal conditions. For example, the relevant dispositions might be stronger or more fundamental in some specifiable way. Jerry Fodor's asymmetric-dependence theory is a subpersonal, informational version of this strategy. At the subpersonal level, we need to substitute talk of covariation or nomic relations between mental representations and worldly objects or properties for talk of thinkers' conditional behavior. Fodor specifies a complex property of nomic relations between mental representations and properties. His theory retreats from pure performance – covariation – to underlying regularities in causal relations, but without moving "inside" to the abilities or mechanisms that ground performance.

Options in the ability category attempt to allow for error not by restricting content-determining conditional behavior (or covariation relations) to some specified class of actual and counterfactual behavior, but by focusing on the ground of that behavior. Such options substitute for the notion of a conditional behavior a notion such as an ability (in my technical sense). (As noted above, at the subpersonal level, the notion of a mechanism may be more apt). Thus, the achievement view can be modified to hold that a thinker *has a concept in virtue of his having an ability to satisfy its associated condition. A thinker's attitude involves a particular concept in virtue of his exercising the appropriate ability.

The ability option is not available to informational theories. As I am using the term, an "informational theory" holds that content is determined solely by covariation relations between bearers of content and worldly objects or properties. Such a theory cannot pursue an ability option since it cannot appeal to the abilities or mechanisms that ground such covariation.[16] An inferential-

[16] There is logical space for a theory that holds that concepts are individuated at the level of reference and that content is determined by the mechanisms that mediate between bearers of content and worldly objects or properties.

role theory need not hold that thinkers would make a concept's canonical inferences under specified conditions; it can hold that they need only have the ability to make those inferences.

The conditional-behavior- and ability-style refinements of the achievement view must of course respect the achievement restriction (sections 2.2 and 2.4.2). We have already addressed the application of the restriction to the specification of a concept's associated condition. That application has implications in the present context. For example, the relevant ability cannot be an "ability" to be responsible to a standard. We have just seen that achievement-based theories can specify that only certain of a thinker's dispositions must be dispositions to satisfy the concept's associated condition. So we need to adapt the achievement restriction to ensure that the way in which those content-determining dispositions are specified does not violate the spirit of the restriction. Consider, for example, a specification that the content-determining dispositions are those that the thinker would have if she were disposed to do what she was required to do. We need to rule out such a specification because it is just a complicated way of saying that content is determined not by what the thinker is disposed to do, but by what the thinker is required to do. We can solve the problem simply by providing that an achievement view may not rely on a normative characterization to specify which dispositions are content-determining.

We can summarize the contrast between the conditional-behavior and ability categories in the following intuitive way. Options of the first category seek to distinguish error-ridden performance from content-determining performance by seeking patterns within (actual and counterfactual) performance. They thus seek to construct something approaching a competence/performance distinction at the level of performance. Roughly speaking, the hope is that there will be a recognizable mark of correct performance, so that, given nothing but the facts about performance, we will be able to distinguish performance that is reflective of the disposition being exercised from performance that has gone astray. Options of the second category establish a genuine competence/performance distinction. Such theories seem to have better prospects for allowing for error since they appeal to an ability that is not constituted by any class of actual and counterfactual dispositions, but lies behind and explains them.

It is worth noting that an important epistemological consideration restricts how far a theory of content can go in relying on a competence that is independent of actual and counterfactual behavior. Attributions of content rely heavily on actual performance and surrounding context. Given the methodological goal of (largely) respecting those attributions, whatever constitutes content will have to be accessible in actual performance. Suppose a theory holds that a thinker's having an attitude with a particular content is constituted by his exercising a competence that is not (epistemically) accessible from his observable dispositions and their context. Such a theory faces a dilemma. Either the correspondence between the theory's assignments of content and our ordinary attributions is no better than random, or our ordinary assignments of content are miraculous. The latter possibility is unacceptable. The former undermines the theory's claim to be a theory *of content*.

We can now see why the achievement view runs squarely into the Kripkenstein and disjunction problems. The core insight of the achievement view is that a thinker's having an attitude involving a particular concept is constituted by his exercising a disposition that corresponds to the concept. It corresponds to the concept in the sense that it is a disposition to satisfy the concept's associated condition. At the same time, the achievement view must allow that having such a disposition is consistent with one's performance being systematically wrong. Thus, the achievement view needs to say what a disposition's being a disposition to satisfy a concept's associated condition amounts to, given the failure of the thinker to satisfy that condition. The two categories of options we have considered in this subsection are potential ways to answer this question. In each case, the same basic challenge arises.

Options in the conditional-behavior category need to be able to identify the content-determining conditional behaviors independently of the contents that they are supposed to determine. Thus, for example, since content determines correct performance, ideal conditions cannot be specified as the conditions under which the person will perform correctly. As we will see, there are fatal problems with specifying ideal conditions in a way that is independent of the contents of thinkers' attitudes. I will argue that Fodor's theory has an analogous difficulty in establishing, without appealing to the contents of a thinker's attitudes, that the appropriate patterns of nomic relations will be established.

Options in the ability category rely on a thinker's having an ability to satisfy a concept's associated condition. Having such an ability can be consistent with one's performance being systematically wrong, because faulty performance can be the consequence of interfering factors. In order for it to be true that the thinker has the ability to satisfy the relevant condition, all errors must be attributable to interfering factors. In other words, if the thinker's performance would still be faulty even when no interfering factors were present, it would be incorrect to attribute an attitude involving the concept in question. The challenge is therefore to provide a distinction between interfering factors and abilities that is independent of the contents of the thinker's attitudes. For example, if a person has apparently added two numbers incorrectly, we need to be able to say what determines that he exercised an ability to add, but was prevented from reaching the correct answer by some interfering factor, rather than an ability to quadd. We find ourselves in the territory of the Kripkenstein and disjunction problems. Whether any of an achievement view's options can be successful is addressed in chapters 3, 4, and 5.

In contrast, as we have seen, the responsibility view does not require that thinkers must have the ability to satisfy the concept's standards in order to have attitudes involving the concept. To have an attitude involving a particular concept is for a thinker to be subject to the standards that are distinctive of the concept. Thus, if a concept is individuated by its reference, it imposes a standard according to which it is correct to apply the concept to what falls under it and incorrect to apply the concept to what does not. The fact that thinkers often fail to meet this standard is simply a reflection of the fact that people can fail to meet the standards to which they are subject. There is therefore no pressure to provide an account of an ability to satisfy a concept's associated condition that is consistent with failing to perform correctly. Instead, we need an account of what makes a thinker responsible to a standard imposed by a concept. This is, of course, a tall order; I sketch a program for providing such an account in chapter 7.

2.7 The achievement view's Buddhist approach to explaining normativity

It is sometimes said that the Buddhist approach to happiness is to reduce one's desires to the point where they are satisfied in one's current circumstances. An achievement theorist can take either of two Buddhist approaches to the problem of explaining the normative aspect of content. First, she can attempt to reduce what needs to be explained. She can do this by denying that content is genuinely normative. In keeping with my project, I take the view that we should explore theories that do not deny that content is genuinely normative before we consider a deflationary approach that denies content's normativity and tries to explain only why we *think* content is normative.

Second, an achievement theorist can attempt to reduce what we demand from an explanation. She can argue that it is possible to provide a constitutive account of content without providing an intuitively satisfying explanation of the normativity of content. The argument would begin by claiming that a constitutive theory of content need only offer a scientific law connecting non-intentional properties to intentional ones. There is a way to establish that a correlation is a law, the argument continues, without showing that there is an *a priori* or intuitive connection. Sciences determine which generalizations are laws by determining which have the most explanatory power, not which strike us as intuitive. Suppose a scientist notices a certain correlation. The scientist can test whether the correlation holds generally, for example, by making and testing predictions. If a true generalization is simple and encompasses a broad range of data, we may become convinced that it is a law. Sometimes, the generalization may strike us as intuitive, and sometimes it may not. Whether it does so, however, may have little to do with science's ultimate assessment of the explanation's success.

Moreover, the argument continues, in some cases, it can be clear in advance that we will not be able to find a basic regularity that will be intuitively satisfying. In such cases especially, it is a mistake to hold that an account that fails to provide such intuitive satisfaction is defective. For example, one could take the following position with respect to consciousness. An account of the properties that constitute or realize consciousness in the brain will not include conscious properties on pain of circularity. Since we have no direct experience of non-conscious properties realizing conscious ones, and, perhaps as a consequence, no *a priori* intuitions about how they could do so, any account of what realizes consciousness will not be intuitively satisfying. On this view, science could successfully explain consciousness by finding a specification of a property that correlates perfectly with consciousness, and establishing, with ordinary scientific methods, that the correlation was lawful. Similarly, it might be argued, it is not possible for a naturalistic account of content to provide an intuitively satisfying explanation of the normativity of content. Since such an explanation is not possible, it would be a mistake to reject an account of content on the ground that it fails to provide such an explanation.

I will argue that, fortunately, we need not resolve this issue. Suppose we concede for the sake of argument that a constitutive account need only provide a scientific bridge law. And suppose further that science can establish that a correlation is a law by demonstrating its predictive power. Still, all this would avail an achievement-based theory nothing unless it could succeed in specifying a natural property that correlates with content. If my arguments concerning the first of the Three Problems are right, however, theories that presuppose the achievement view fail to do so. For example, if the standard theories of content cannot allow for the errors that thinkers make, they cannot claim to have discovered a correlation that would be a serious candidate to yield a constitutive account of content. (My argument that currently standard theories of content cannot account for attributions of content involving incompletely understood concepts – the third of the Three Problems – also undercuts the claim that such theories can offer a good correlation.) The fact that science can establish that a correlation is lawful without showing that it is *a priori* or intuitive is irrelevant if there is no serious candidate for a lawful regularity.

Finally, whatever force the analogous argument may have in the case of consciousness, we should reject the claim that a natural property that correlates with content is the most that we can hope for. In the case of content at least, we are not faced with a stark choice between no account and a reduction that at best provides an explained correlation. We have no idea how physiological properties could give rise to conscious ones. Further, it is often assumed that there can be no properties that are intermediate between the non-conscious and the fully conscious. In contrast, as I discuss in chapter 7, we have some familiarity with ways in which descriptive facts give rise to normative ones. Also, although it is not clear how to make the idea precise, it seems possible that normativity comes in degrees; indeed, it is not difficult to think of potential examples.

Moreover, we need not accept the reductionist claim that a natural account of content will have to leave out the normativity, just as an account of pain will leave out the painfulness. The naturalistic constraint that an account may appeal only to facts that supervene on descriptive facts does not imply that a naturalistic account of content must not appeal to normative facts. Thus, the responsibility view banks on the possibility of providing an account of intentional facts in stages. First, we could give an account of what it is in virtue of which a person has an attitude involving a particular concept that relies on non-intentional, normative facts. We could then provide an account of how descriptive facts can realize such normative facts.[17] This result would be that our account of what it is in virtue of which an attitude has a particular content (stage 1) would not leave out normativity. At the same time, however, stage 2 would ensure that the normative facts on which the account relied would be naturalistically legitimate. As noted above, stage 1 can offer explanatory value even without the completion of stage 2.

2.8 The achievement view and deference

In section 2.4.3, we saw that the basic form of an achievement view makes it problematic for achievement-based theories to account for incompletely understood contents. A thinker who incompletely understands a concept will not reliably have a disposition or ability to satisfy the concept's associated condition. Standard theories of content, when they do not ignore incomplete understanding altogether, tend to assume that it can be adequately accounted for simply by adding a qualification about "deference." In section 6.5, I will argue that a theory that presupposes the achievement view has difficulty explaining the role of deference. In contrast, the responsibility view nicely accounts for deference as simply one way in which a thinker can become responsible to a standard.

In brief, my argument will be that it is not plausible that deference enables a thinker who incompletely grasps a concept to acquire the disposition or ability that, according to an achievement view is constitutive of *having a concept. Thus, deference fails to enable a thinker to come within an achievement-based theory's standard explanation of what it is to *have a concept. (It is not the case that deference need merely play the role of uniquely picking out the concept. If *having a concept is having a disposition to satisfy the concept's associated condition, deference can enable a person to *have a concept only if it enables her to have the appropriate disposition.) An achievement-based

[17] As I discuss in section 7.5.2, stage 2 would not be a reduction of normative facts to non-normative ones.

theory can claim that deference is a second, different way of *having a concept. But I argue that a theory that makes this claim fails to answer the content question because it fails to explain what the two cases have in common that makes them both cases of *having a concept.

The responsibility view makes sense of the situation. According to the responsibility view, deference is a way for a thinker to become responsible to a standard that she cannot independently identify. This proposal 1) offers a natural view of the role of deference and 2) makes clear what it is that cases that depend on deference have in common with cases involving full understanding. First, deference is more plausibly a way of committing oneself to a standard than a way of acquiring a disposition or ability. As to the second point, the idea is roughly that a thinker who fully understands a concept can herself pick out the concept's associated standards and thus can commit herself to them;[18] a thinker who incompletely grasps a concept can use deference to accomplish the same thing. Having a disposition to satisfy a concept's standards and deferring to others who have that disposition are merely (elements of) different ways to realize one's being responsible to the concept's standards.

This line of thought raises the question whether there may be other ways in which a thinker who incompletely understands a concept (or indeed a thinker who fully understands it) may become responsible to its standards. When a thinker incompletely understands a concept, we have seen that there is a difficulty in explaining what could determine that her thought involves that concept, as opposed to others. The responsibility view has resources for providing such an explanation that are not available to an achievement view. In a nutshell, facts about the external world may play a broader role in determining what standards a thinker is responsible to than they play in determining what dispositions or abilities a thinker has.

2.9 Attempts to narrow the gap between achievement and responsibility views

There are a variety of ways in which the gap between the achievement and responsibility views could be narrowed. This section concerns two related tacks that an achievement theorist might take.

First, the theorist might seek to reinterpret an achievement view as a view about what makes a thinker responsible to a standard imposed by a concept. Thus, she might claim that the

[18] I should note that it is not my view that such explicit commitment is the typical way in which thinkers become responsible to conceptual standards. I suspect that typical thinkers do not have full grasp of many of their concepts. Even when a thinker does have full grasp of a concept, that full grasp may not be crucial to what makes the thinker responsible to the relevant standard. See chapter 7 for discussion.

achievement view is an attempt to answer a question that the responsibility view merely poses. On this reinterpretation of the achievement view, it would be in effect an achievement version of the responsibility view. It would accept that to have an attitude involving a concept is to be subject to the standards distinctively imposed by the concept. It would further hold, however, that what it is in virtue of which a thinker is responsible to a standard imposed by a concept is the thinker's having a disposition or ability to satisfy the standard.

It is difficult to see what the argument for this position could be. Absent special circumstances, whatever it is that makes it correct for someone to fulfill a given condition, it is not his being disposed or able to fulfill it. In general, having a disposition or ability to do something is not even *sufficient* for being required to do it. So having a disposition or ability to satisfy a standard is not constitutive of being responsible to the standard.[19] I have the abilities to peel an apple in one continuous piece ten meters long and to throw rubbish into the bin with pinpoint accuracy. Unfortunately, I am not required to perform either of these feats.[20]

Second, the achievement theorist might make the weaker claim that a thinker's having an ability[21] to satisfy a concept's standard is necessary for the thinker's being responsible to the standard (though it is not constitutive of it). The doctrine, derived from moral philosophy contexts, that "ought implies can" might be adduced as support for this claim. That is, it might be said that, since "ought implies can," a thinker must be able to satisfy a concept's standards in order to be subject to them. Though a thinker need not actually apply a concept to all and only the objects that fall under it (or make all the mental transitions required by the concept), a thinker must at least have the ability to satisfy those standards. As a result, the achievement theorist might continue, the implications of the responsibility view are in practice likely to be less different from those of the achievement view than the theoretical contrast might lead one to expect.

The doctrine that "ought implies can" (*the doctrine*, for short) does not imply that a thinker must have the ability to satisfy a concept's standards in order to be responsible to them. The basic

[19] It might be suggested that, nevertheless, having an ability to satisfy a standard is a necessary *part* of being responsible to a standard. In the next few paragraphs, I argue that having an ability to satisfy a standard is not necessary for being responsible to the standard. It follows that having an ability to satisfy a standard is not a necessary part of a constitutive account of responsibility to a standard.
[20] For brevity, I've made the point in terms of an ability to do T rather than an ability to satisfy a condition that one do T. The formulations are equivalent. If the argument of the achievement-view proponent is not to beg the question, a condition must not be understood as a requirement, but in a way that allows any ability to be cast in the form of an ability to satisfy a condition.
[21] I focus on abilities, as opposed to conditional behavior, because the corresponding claim for conditional

reason is that the notion of ability that figures in the doctrine is distinct from the notion of ability that figures in the achievement theorist's claim that an ability to satisfy a standard is necessary for responsibility to a standard. Thus, even if a strong version of the doctrine is true and can be validly imported from the moral theory context, it does not support the achievement theorist's argument.

In the sense in which the responsibility view employs the term, for a person to be *responsible* to a certain standard is for the correctness or incorrectness of his performance to be determined by that standard. This sense of responsibility must be distinguished from a second sense in which, roughly speaking, a person is (morally) responsible for an action, failure to act, or outcome just in case he freely chooses it. The first sense has to do with the wrongness of action; the second with the agent's blameworthiness. The achievement theorist cannot base her argument on the doctrine unless the doctrine is true for a use of "ought" that corresponds to the first sense of responsibility. Therefore, on the version of the doctrine that we need to consider, "ought" is used as follows: a person ought to do T just in case he is subject to, or responsible to, a standard specifying that he do T.

We can now turn to the crucial distinction between two notions of an abilit*y to satisfy a standard*. On the first notion, you have the ability to satisfy a standard if and only if you have an ability to perform the action that the standard in fact requires (though you need not know what the action is). We can call this the *mechanical* notion of an ability to satisfy a standard because, in order to have a mechanical ability, you need not be able to work out what the standard requires. Suppose that a certain tax law requires any person who meets a complicated condition to pay a tax of $1,000. The mechanical ability to satisfy the legal standard requires only that one have $1,000 (or be able to acquire it). On the second, *intelligent* notion, in order to have the ability to satisfy a standard it is not sufficient that you have the ability to perform the action that the standard in fact requires. You also need to be able to recognize what the standard requires. Thus, if you have a mechanical, but not an intelligent, ability to satisfy a given standard, you may try, yet fail, to fulfill the standard simply because you have an incorrect understanding of what it requires. In order to have the intelligent ability to satisfy the tax-law standard, one must be able to work out what it requires (as well as having $1,000).

---

behavior would be plainly false.

An example will make clear that "ought implies can" is plausible only with respect to a mechanical reading of "can." Suppose that you are faced with a moral dilemma concerning whether to return a gun to its owner. And suppose that the applicable moral standards require that you return the gun. If you are paralyzed or stranded on an island, you may lack even a mechanical ability to satisfy the standards. In such a case, it is at least arguable, in accordance with the doctrine, that you are not subject to a requirement that you return the gun.

In contrast, suppose that you have the mechanical ability to satisfy the applicable standards, *i.e.*, you are able to return the gun. In fact, however, you sincerely believe that the right thing to do is not to return the gun. As a result, despite your desire to do what is morally required (under that description), you do not return the gun. You do not return the gun because your moral understanding is inadequate. Although you have the mechanical ability to satisfy the moral standard, you lack the intelligent ability. You can do what is in fact required, but you can't do what is required under that description because your moral sense leads you astray. In such a case, it is clear that you are still subject to the moral requirement that you return the gun in the sense that your not returning the gun is wrong. (Again, whether you are morally responsible or blameworthy if you fail to return the gun is irrelevant to whether you are subject to, or responsible to, the moral requirement.)

If one were required to have an intelligent ability to satisfy a moral standard in order to be subject to that standard, poor moral judgment or weak character would tend to exempt one from the demands of morality. But that is not plausible. (It is a different question, which we can set aside for present purposes, whether poor moral judgment can reduce one's moral responsibility for the consequences of one's moral misjudgments). What is morally required does not depend on how good one's moral judgment is. A sincere belief that equality permits slavery does not negate the wrongness of keeping slaves. Distributions that would otherwise be unfair are still unfair even if the distributor misunderstands fairness. Thus, the doctrine that "ought implies can" would not be plausible if "can" were taken in an intelligent sense. One can be responsible to a standard even if one lacks the intelligent ability to satisfy it.

When an achievement view holds that *having a concept is having an ability to satisfy the concept's associated condition, it employs an intelligent notion of ability to satisfy a standard.

Suppose that the condition associated with the concept of a penguin is that a thinker apply the concept to all and only penguins. The mechanical ability to satisfy this condition is a rather meager ability. To put it loosely, it is the ability to apply the concept to all and only penguins, given an expert who tells you which are the birds that fall under the concept! (Compare the mechanical ability to satisfy the moral standard in the gun example: in order to have that ability, you must not be paralyzed, stranded on an island, and so on.) In order for someone to have the ability to satisfy the condition associated with the concept of a penguin (in the achievement view's sense), it is not sufficient to have the mechanical ability. The ability that the achievement view requires is an ability to recognize when to apply the concept. Thus, it is consistent for a thinker to lack the ability that the achievement view requires and to have the (mechanical) ability. For example, a person who believes that *C* is correctly applied to all water birds will not have the ability to satisfy the condition in the relevant sense. He will lack the intelligent ability. He may have the mechanical ability, however; if an expert tells him which birds fall under the concept, he may succeed in applying the mental representation to penguins and only penguins.

Turning to an inferential-role-style example, consider the intuitive condition for AND (the concept of conjunction) mentioned in section 2.4.2. In order to have the ability to satisfy the condition, a thinker must have more than the mere (mechanical) ability to make the mental transitions specified in the condition. That is, it is not sufficient that if he tries to make each of the specified transitions, he will succeed. The thinker must be able to recognize the appropriate mental transitions. In other words, one has the ability just in case when one tries to satisfy the standard, one makes the transitions one makes because they are the ones specified in the condition. Thus, a person may lack the ability to satisfy the condition for AND, despite the fact that she has the mechanical ability to make each of the specified mental representations. It is thus clear that the relevant ability to satisfy a conceptual standard is an intelligent ability.

Since the doctrine that "ought implies can" is not plausible for intelligent abilities, the doctrine provides no support for the achievement theorist's claim that a person must have the ability to satisfy a concept's standards in order to be responsible to them. (It may be true, of course, that a person must have the *mechanical* ability to satisfy a concept's standards in order to be responsible to them. But nothing in my argument will depend on denying that claim.)

Part II: Theories of Content

Chapter 3: The Rorschach Problem

3.1 Introduction

This chapter begins my argument that currently standard theories of content are inadequate even on their own terms. Since I am evaluating them on their own terms, I do not raise issues of incomplete understanding or normativity.[1] I obviously cannot evaluate all theories of content or even all types of theory. Instead, I have two aims in the present chapter.

First, I set up a classification of theories, and eliminate from consideration categories that can be shown to have irremediable defects. This will establish which categories need further attention, thus motivating my critiques of Fodor's and Peacocke's theories in chapters 4 and 5 respectively.

And second, based on the categories considered and rejected in this chapter, I put forth a tentative, intuitive diagnosis of the underlying problem with achievement-based theories of content. The extensive critiques of Fodor's and Peacocke's theories as well as the discussion of normativity and incomplete understanding in chapter 6 will provide further confirmation of that diagnosis. I thus lay the groundwork for an argument that specific objections to particular theories reflect a deeper underlying problem for achievement-based theories of content.

The roadmap for the chapter is as follows. First, drawing on Boghossian's (1989) excellent survey,[2] I classify theories of content and review reasons for rejecting two major categories of achievement-based theories – those that rely on ideal conditions and those that appeal to community consensus. I expand Boghossian's classificatory scheme to include two types of theories he does not consider (they were not well represented at the time). Next, I suggest that Boghossian's objection to ideal-conditions-based accounts reflects a broader problem for achievement-based theories of content. I make the suggestion not as an

[1] The one exception is that normativity plays a role in my tentative diagnosis of the underlying problem with standard theories of content (section 3.4).

objection to any specific theory, but as an attempt at a unifying diagnosis of what lies behind such diverse objections to achievement-based theories. The test of the diagnosis will lie in its success in explaining the specific problems that particular theories encounter.

3.2 Classification

3.2.1 Preliminaries

In this section, I lay out Boghossian's classificatory scheme, summarize his treatment of the various categories, and identify two types of theories that need to be considered in chapters 4 and 5. Not until the next section (3.3) do I consider Boghossian's arguments in detail.

Boghossian divides theories of content into reductionist and non-reductionist theories.[3] On the reductionist side, he considers dispositional theories – theories that hold that what it is in virtue of which a thought has a certain content is the thinker's having a certain disposition. According to Boghossian's scheme, dispositional theories include both informational and inferential-role theories. He considers two versions of such theories – those that rely on ideal conditions and those that appeal to community consensus. On the non-reductionist side, Boghossian thinks that nothing substantive can be said in answer to the content question – that the only possibility is "quietism." In sum, Boghossian's scheme includes two types of reductionist theories (informational and inferential-role), each of which comes in two flavors (ideal-conditions and communitarian) and one type of non-reductionist theory (quietist).

As a preliminary matter, I want to address Boghossian's use of the terms "reductionism" and "naturalism." Boghossian effectively equates reductionism and naturalism.[4] Though he does not further define reductionism explicitly, it is fair to say that he takes a reductionist account of what determines content to be one that is limited to non-normative, non-intentional (and non-semantic) terms.[5] This is a reasonable understanding of reductionism. (It is not as extreme as a version that

---

[2] In this chapter, all citations to page numbers alone are to Boghossian (1989).
[3] Boghossian begins by arguing that "irrealist" conceptions of content – ones that hold that there are no truths about content – are incoherent (pp. 522-527). I agree with his conclusion, and leave such views outside the scope of this essay.
[4] He defines a reductionist view as one that gives an account of what determines content in terms of "naturalistic facts and properties" (pp. 527, 540-541). Further, he holds that non-reductionism "concedes the failure of naturalism" (p. 527).
[5] As quoted in the previous note, he defines reductionism in terms of naturalism. At one point, he characterizes non-reductionism in a way that implies that a reductionist account is one that accounts for content properties in terms of "physical/functional" properties (p. 541). "Physical properties" here must mean something much looser than the properties of basic physics or even those of the physical sciences; otherwise, a restriction to the

would require the intentional to be reduced to terms employed in the physical sciences, but is stronger than a version that would allow the use of normative, but not intentional, terms (which would perhaps not ordinarily be considered fully reductionist).

As for naturalism, I don't intend in this essay to enter the debate over how best to understand it. It is worth noting, however, that we might want to be cautious about following Boghossian on this question. He evidently holds that in order to give a naturalistically acceptable account of mental content, we need to give an account in non-intentional, non-normative terms.[6] A naturalism that requires reductionism plainly cannot allow appeal to intentional notions, nor (at least given a typical reading of reductionism) to normative ones.

We might prefer a more relaxed understanding of naturalism, such as one that allows an account of content to use terms (properties, facts) that satisfy an appropriate supervenience requirement (and that do so for ontological reasons, as opposed to, for example, reasons deriving from linguistic rules).[7] (We also might want to impose additional requirements.) The supervenience base could be restricted to the non-intentional and non-normative, or could be specified in other ways. As noted above (sections 1.3.1, 2.1.3), such an understanding of naturalism might allow a theory to employ normative, or even intentional, terms, if those terms could be shown to satisfy the supervenience requirement. (Apart from naturalism, there might of course be other reasons, such as non-circularity ones, that would restrict the use of normative or intentional terms.) As we will see, Boghossian does not consider the possibility of a non-reductionist theory that tries to show that an appropriate supervenience requirement is satisfied (and perhaps to explain why it is satisfied). He assumes that there is no option lying between, on the one hand, reductionist theories and, on the other, theories that can say nothing about the relation between intentional and non-intentional facts or properties.

"physical/functional" would rule out the theories Boghossian takes as his central cases of reductionism. Those theories are cast in terms of a person's dispositions to apply mental symbols to properties (pp. 528-540). Thus, I take his "physical/functional" category to be roughly equivalent to my "non-normative and non-intentional:" it rules out normative, evaluative, and intentional properties, terms, or facts, but does not imply a restriction to those of physics.

[6] Read literally, Boghossian's definition of a reductionist view as one that gives an account of content in terms of "naturalistic facts and properties" might seem to imply a use of the term "naturalistic" according to which even a successful reduction of intentional facts to the terms of physical science would not show that intentional facts were naturalistic. This usage would be highly confusing. Presumably, Boghossian means simply that intentional facts and properties are not naturalistic "in their own right" but only if they can be reduced to the non-intentional. As discussed in the text, we might not want to require reduction.

[7] Peacocke gives a characterization of naturalism that includes a supervenience requirement (1992, pp. 127-128). His version would also hold "that any explanation of an event or temporal state of affairs is a causal explanation" (1992, p. 127).

3.2.2 Reductionism and dispositional theories

Boghossian takes dispositional accounts to be the only serious reductionist options (p. 527). Dispositional accounts, as I noted above, are those that try to give a constitutive account of content in terms of the thinker's dispositions. Although Boghossian does not define "disposition" or distinguish the sense in which he uses it from other accepted senses, his use of the term makes reasonably clear that, in his sense, a person has a disposition to do X in circumstances C just in case it is true that, *ceteris paribus*, the person would do X if C were the case.[8] In other words, Boghossian uses the term "disposition" for actual and counterfactual behavior, rather than for some more "full-blooded"[9] notion. (As I stipulated in section 1.4, I use the term *conditional behavior* in the same way, reserving the term *ability* for the causal basis or underlying mechanism of the disposition.) Boghossian does not consider the possibility of a theory that appeals to abilities rather than dispositions.[10]

As mentioned above, under Boghossian's classificatory scheme, dispositional accounts include both informational and inferential-role theories of content (pp. 527-528). In the case of informational theories, content is accounted for in terms of dispositions to apply[11] (syntactically characterized) symbols to objects; in the case of inferential-role theories, in terms of dispositions to move between (functionally characterized) mental states.[12]

Boghossian notes that two general types of dispositional accounts have been proposed: 1) those that rely on some kind of idealization of actual dispositions; and 2) those that rely on community dispositions. As we will discuss in section 3.3, he gives general arguments for rejecting both types (pp. 534-540). Since, as mentioned above, he finds "no merit" in the suggestion that there are viable non-dispositional reductionist accounts (p. 527), he reluctantly concludes that we are forced to accept non-reductionism (pp. 548-549).

How do Boghossian's dispositional theories relate to the achievement view? In principle, his dispositional category includes more than just achievement-based theories. An achievement-based

---

[8] See pp. 528-529.
[9] See Evans (1982, p. 329).
[10] As noted above, he asserts that dispositions are the only "available naturalistic facts" for an account of content (p. 527).
[11] In sections 3.4.1 and 4.5.1, I argue that a reductive dispositional theory cannot take for granted that an occurrence of a symbol is an *application* of a symbol. For simplicity, I elsewhere allow the dispositionalist to talk of "applications" of symbols.
[12] Boghossian's classification scheme does not distinguish between subpersonal- and personal-level theories. Strictly speaking, for example, a subpersonal-level informational theory appeals not to a *thinker*'s dispositions to token a mental representation but to a *mental representation*'s dispositions covariation with instances of a property.

theory holds that the relevant dispositions are dispositions to satisfy a concept's associated condition.[13] (Of course, an achievement-based theory can also appeal to abilities rather than the conditional behavior that Boghossian considers, but we can leave that qualification aside for the moment.) That is, according to an achievement-based theory, to *have a concept is to have a disposition to satisfy that concept's associated condition. Boghossian does not officially require that a dispositional theory rely on dispositions to satisfy a concept's associated condition. But, given the naturalness of the achievement view (section 2.1), it is unsurprising that all the theories he considers are in fact achievement-based. Ideal conditions and community consensus are ways of picking out the dispositions that are supposed to satisfy a concept's associated condition. Boghossian does not consider the possibility of a dispositional theory that holds that *having a concept is constituted by dispositions other than dispositions to satisfy the concept's associated condition.

3.2.3 Non-reductionism and quietism

Boghossian maintains that once we have ruled out reductionism ("naturalism," in his terms,) we have lost any possibility of answering the content question – of saying what it is in virtue of which a thought has one content rather than another. He seems to take largely for granted this view that non-reductionism must be quietist, and gives very little argument for it. For example, he equates holding that there is "no interesting reduction *of mental content properties* to *physical/functional properties*" with holding that "at some appropriate level mental content properties must simply be taken for granted" (p. 541) (emphasis in original). Similarly, he uses interchangeably the phrases "rejecting substantive *reductive* answers to the constitutive question" (the content question) and "reject[ing] substantive answers to the constitutive question" (pp. 543-544) (emphasis added).[14]

Boghossian considers two non-reductionist positions that do try to say something about the nature of, or conditions for, content. He criticizes and rejects those positions,[15] but give no general grounds for thinking that non-reductionist positions can have nothing to say in answer to the content

[13] See, *e.g.*, sections 2.1, 2.3 above.

[14] The quoted phrases are used interchangeably in a discussion of McDowell's position (pp. 543-544). I do not mean to defend McDowell against Boghossian's specific criticism (see following note). My point is simply that, throughout his discussion of McDowell, as elsewhere, Boghossian assumes that a non-reductionist can say nothing substantive about the content question.

[15] First, he argues that McDowell fails to justify his conclusion that a communal practice is necessary for meaning (pp. 543-544). Second, Boghossian considers Crispin Wright's attempt to derive a "judgment-dependent" view of content – one on which facts about content are constituted by our judgments about content – from epistemological problems facing the position that we cannot answer the content question. Boghossian shows that a judgment-dependent view collapses because it presupposes judgment-independent facts about content (pp. 544-547).

question. Then, having already rejected all the reductive categories, he reluctantly embraces a non-reductive position. In his discussion of this position, he leaves no doubt that he assumes that the position must be quietist (pp. 547-549). For example, he raises the worry that a non-reductionist view would run out of things to do (p. 548). In response, he lists several projects that non-reductionist theories of content could legitimately pursue, but makes it clear that those projects, even if successful, would not answer the content question (p. 549). (One such project, for example, is to provide an account of the composition of structured mental contents out of their component contents (p. 549). This project, however, has no potential to offer insight into the content question itself since it takes the contents of the components for granted.)

Pace Boghossian, we cannot assume that no non-reductionist theory can provide an answer to the content question. As I noted in section 1.3, normativity and quietism tend in practice to be linked: theorists who insist on the normativity of content often hold that it bars an attempt to answer the content question.[16] Since positions that appeal to normative notions are prominent among non-reductionist approaches to content, it is understandable that Boghossian would associate non-reductionism with theories that do not attempt to answer the content question. He gives no general reasons, however, for thinking that all theories that his scheme categorizes as non-reductionist – or even all those that employ normative notions – cannot hope to answer the content question. We mentioned some potential non-reductionist approaches in sections 1.3 and 2.7.2 and will return to the issue in chapter 7. For example, Boghossian does not address the option of allowing appeal to normative properties and facts in answering the content question, with the further possibility of showing that normative properties and facts are multiply realized by non-normative ones.[17] Indeed, he ignores the possibility that content facts might supervene on, without being reducible to, non-normative, non-intentional facts, and seems to take for granted that if reduction is not feasible, nothing can be said about the relation between the intentional and non-intentional. And he does not consider the possibility that a large enough explanatory circle might not be vicious. I do not mean to

[16] As Boghossian recognizes, however, accepting that content is normative need not be tantamount to accepting that standard kinds of explanation of content are ruled out. Otherwise, given his position on the normativity of content, he would not need even to address reductive accounts of content. And he is explicit that a dispositional account might be able to "captur[e] the normative force of an ascription of meaning" (p. 533).

[17] It is not obvious that taking normativity for granted is the same as taking intentionality for granted. For one thing, many theorists of content deny that it is normative at all or, at least, that normativity is even partially constitutive of content (see section 6.7). For another, as I discuss in chapter 7, there are at least arguable examples of normativity that are not dependent on the content of attitudes. And if normativity does in some ways depend on mental content, the circle may be large enough that appealing to normative notions in

suggest, of course, that the non-reductionist who wants to say something substantive has smooth sailing; the point is simply that it is not evident that quietism is forced upon the non-reductionist.

3.2.4 Other theories to consider

We need to broaden our net to encompass (at least) two other types of theories. The first additional category includes reductive theories that do not depend on ideal conditions or community dispositions. Jerry Fodor's (1990) asymmetric-dependence theory, which Boghossian does not discuss (the earliest version was first published only in 1987), is the best-known contender in this category.[18]

The second category includes theories that are not reductionist (at least not in Boghossian's sense), but not quietist: theories that employ normative or intentional notions but that do so in an attempt to answer the content question, rather than in the service of a view according to which the content question cannot be answered. Christopher Peacocke's (1992) possession-condition theory is the best example of such a non-quietist non-reductionist theory.

In the next section, I turn to a substantive discussion of Boghossian's arguments against ideal-conditions and community versions of dispositional theories. I will agree with his basic conclusions that such theories are not viable contenders. I will nevertheless discuss his argument against ideal-conditions versions of dispositional theories at some length in order to develop my suggestion about the underlying problem.

3.3 Boghossian on reductive dispositional accounts

Boghossian uses an informational account as a paradigm dispositional account throughout his discussion. Since he takes inferential-role and informational theories to have the same very general structure, and since (he asserts) his arguments do not depend on specifically informational aspects of his example, he takes his arguments to apply to inferential-role as well as informational theories of content (p. 528). In this chapter, I will for the most part follow Boghossian in not treating inferential-role theories separately, though I will

---

answering the content question may provide explanatory gain. We cannot assume that, once we have appealed to normative notions, we have ruled out the possibility of explanation. See sections 7.5, 7.7 below.

[18] Boghossian's ideal-conditions category includes dispositional theories that appeal to biology to specify ideal conditions (p. 537). We may want to distinguish a theory that holds that the content of a representation is determined not by that to which the thinker is disposed to apply it under certain conditions (which we look to biology to specify), but by what it is the representation's biological function to represent. Ruth Millikan's

occasionally point out why the arguments apply to those theories.

The basic task that an informational account sets itself is to specify which applications[19] of a symbol determine its reference. Since not all dispositions to apply a symbol correlate with the property that is the symbol's reference, an informational account must specify – and, since it is supposed to be reductionist, specify in non-intentional terms – which dispositions are the crucial reference-determining ones. That is, an informational account must specify a class of symbol-tokening dispositions such that a property is the symbol's reference if and only if there is a correlation between that property and dispositions in the specified class.

3.3.1 The futility of an appeal to community dispositions

As mentioned above, Boghossian considers two ways of specifying the relevant class of dispositions, in terms of community agreement and in terms of ideal conditions. Let us begin with community-based accounts. A pure community account, one that holds an expression's content to be determined by the community's actual dispositions to apply it, has the implausible consequence that communal consensus constitutes truth.[20]

It is implausible enough to hold that the community as a whole cannot be wrong, but matters are worse. As Boghossian points out, many of our mistakes are "systematic" in the sense that "they arise because of the presence of features . . . that have a generalizable and predictable effect on creatures with similar cognitive endowments;" such mistakes "are bound to be duplicated at the level of the community" (p. 536). The community may, for example, be systematically disposed to apply "horse" to cows under certain conditions, and the problem is entirely general: "[t]here are countless possible impostors under countless possible conditions; and there is nothing special about the term 'horse'" (p. 536). The consequence is that a communitarian account will inevitably get the reference of our concepts wrong. The account is thus faced with the very problem that the appeal to community dispositions was intended to solve. Since community dispositions are no better off than individual dispositions, I will consider communitarian dispositional accounts out of the running and discuss them no further.[21]

(1987) theory is closer to this latter model.

[19] See note 11 above.

[20] It is worth noting that such a communitarian account seems especially poorly suited to the case of mental, as opposed to linguistic, content.

[21] As Martin Davies has pointed out to me, the futility of appeal to community consensus seems much clearer with respect to the disjunction problem than it has seemed to some with respect to the Kripkenstein problem.

3.3.2 The belief-holism objection to ideal-conditions-based accounts

Turning to ideal conditions, Boghossian sketches a general argument against the possibility of non-intentionally "specifiable conditions under which a subject will be disposed to apply an expression only to what it means" (p. 537). In brief, the argument begins from the holism of belief fixation: "what contents a thinker is prepared to judge will depend upon what other contents he is prepared to judge. . . . just about any stimulus can cause just about any belief, given a suitably mediating set of background assumptions" (p. 539). Further, he argues, plausibly, that because beliefs are fixed in this holistic way, there are indefinitely many sets of background beliefs that could lead a thinker to a wrong belief. In his example, "Neil may come to believe LO, A MAGPIE, as a result of seeing a currawong, because of his further belief that that is just what magpies look like; or because of his belief that the only birds in the immediate vicinity are magpies; or because of his belief that whatever the Pope says goes and his belief that the Pope says that this presented currawong is a magpie" (p. 539).

Therefore, Boghossian argues, in order to specify conditions under which a thinker could not make an error, it would be necessary to rule out all of the indefinitely many sets of background beliefs that could lead to an erroneous belief. Thus, Boghossian thinks, what is needed is a non-intentional way to specify that no such beliefs are present. At this point, he claims that a fatal circularity has been exposed. According to Boghossian, in order to specify in non-intentional terms a situation in which none of a "potential infinity of background clusters of belief is present," "[w]hat is needed is precisely what a dispositional theory was supposed to provide: namely, a set of naturalistic necessary and sufficient conditions for being a belief with a certain content" (p. 540). In light of his arguments against community- and ideal-conditions-based theories, Boghossian concludes that a dispositional theory cannot provide a reductionist account of content.

It is not entirely clear why Boghossian thinks that the problem that he has identified for ideal-conditions-based theories is one of *circularity*.[22] Boghossian has shown, however, that a

In light of my arguments that the two problems are fundamentally the same (sections 1.3.2-1.3.3), it would be interesting to pin down the reasons for this difference. The influence of Wittgenstein is undoubtedly part of the explanation. Also, the disjunction problem, more often than the Kripkenstein problem, is posed in terms of the content of thought rather than the meaning of linguistic expressions. As mentioned in the previous note, it is difficult to see how community consensus could be a solution in the case of mental content.

[22] For one thing, it is not obvious that an answer to the content question, even if we had it, would provide a way of ruling out false background beliefs (and if it would not, then Boghossian has not shown, as he would need to in order to support his circularity charge, that what is needed in order to specify ideal conditions is an answer to

specification of ideal conditions of the sort the dispositionalist needs must be a specification of conditions under which a thinker would have no false background beliefs. Since it is exceedingly implausible that there could be such a specification, we should accept Boghossian's conclusion that no account that requires that there be (non-intentionally specifiable) ideal conditions can succeed.

Before turning to my diagnosis, we should address how the argument would go in the case of inferential-role accounts (Boghossian assumes that the argument goes through for inferential-role accounts but does not discuss the point). Consider an inferential-role theory that holds that whether a thought has a particular content depends on whether the thinker is disposed to make certain canonical transitions to and from that thought. A thinker whose thought has a given content can fail to be disposed to make the transitions that are canonical for that content and can be disposed to make the transitions that are canonical for a different content. The inferential-role theorist needs some way of identifying the content-determining transitions.

Even if ideal conditions could rule out sources of error such as distraction or failure of short-term memory, background beliefs can lead a thinker to fail to make the appropriate canonical transitions or to make the wrong ones. The point is not that which inferences a thinker will make will depend on which premises he believes (which is unproblematic for the inferential-role theorist). Rather, whether the thinker who has a certain belief will make a canonical transition from that belief will depend on background beliefs. Suppose that the transition from *X is a sofa* to *X is a piece of furniture made for sitting on* is among the canonical transitions for the concept of a sofa. Believing that the Pope has said that sofas are religious artifacts could prevent someone from moving from the belief that something is a sofa to the belief that it is made for sitting in. Also, propositional attitudes other than beliefs, such as fears, desires, and hopes can affect a thinker's disposition to make canonical transitions. Various desires and fears, for instance, might prevent someone from inferring from his belief that he talks a lot to the conclusion that he is a chatterbox or from his belief

---

the content question). And the attempt to specify extra-strength ideal conditions that would rule out false background beliefs, though it may be doomed to failure, is no more circular than an attempt to specify the ordinary, milder form of ideal conditions (that do not aim to rule out false beliefs). In a later article (1991), Boghossian seems to see the point that an answer to the content question would not help with false beliefs (though he does not mention its impact on the circularity charge). He recognizes that there is no finite way to state the set of beliefs whose absence must be ensured. Therefore, even given naturalistic necessary and sufficient conditions for meaning, "there is no way to certify" that we have ruled out every member of the set (1991, p. 81). In that work, Boghossian develops the belief-holism argument differently from the original version, omitting the accusation of circularity in favor of the claim that it "appears utterly incredible that there should be" a non-semantically specified situation that guarantees the absence of all the potentially interfering beliefs" (1991, pp. 79-80).

that he is taller than other people to the conclusion that he is tall.

3.4 The diagnosis

Boghossian's belief-holism argument, which was an objection to ideal-conditions-based versions of both informational and inferential-role theories, is limited in two respects. First, it does not purport to cut against theories that do not rely on ideal conditions.[23] Second, it offers no diagnosis of the underlying problem with ideal-conditions-based theories or with dispositional theories in general.[24] Boghossian draws no positive suggestions from the belief-holism argument about what does determine content.

3.4.1 The Rorschach problem

Boghossian's objection is that some dispositions – those that are caused by false background beliefs – are not among the content-determining dispositions. And there is no way to specify (in non-intentional terms) ideal epistemic conditions that will exclude false background beliefs. I suggest that false background beliefs are just an instance of a broader problem. I have already suggested (section 3.3.2) that dispositions (to make judgments, have thoughts, move between mental states) that are caused by propositional attitudes other than beliefs, such as desires or fears, also need to be excluded. I will now argue that there is another important class of dispositions that are not content-determining.

We can begin by considering a problem for informational theories that tends to be overlooked. Boghossian's discussion itself overlooks the problem. Boghossian takes the informationalist's starting point to be the set of *applications* of symbols to objects. He thus frames the informationalist's task as finding conditions under which a subject will be disposed to apply an expression only to what it means (*e.g.*, pp. 528, 531-532, 538). This formulation understates the difficulty of the informationalist's task. An *application of a symbol to an object* is a use of a symbol to characterize a presently perceived object as falling within the symbol's extension. To apply "horse" to a horse is to *call* a horse "a horse," not merely to token "horse" in the presence of a horse

[23] In an exchange with Fodor, Boghossian claims (and Fodor denies) that, despite appearances, Fodor's asymmetric-dependence account of content depends on such conditions. The two agree that the issue of whether Fodor's theory relies on ideal conditions is determinative of whether Boghossian's general objection to dispositional theories applies to Fodor's account (Boghossian, 1991, pp. 71-73, 78-81; Fodor, 1991a , pp. 272-274, 276).
[24] The closest Boghossian comes to attempting a diagnosis is the circularity idea, which I have disputed. See note 22 above.

(pp. 529, 531; Fodor 1990, p. 97).

Contrast examples of two kinds of cases. In the first, the smell of my house reliably causes me to think HOME. In the second, the smell of cardamom reliably causes me to think HOME because, let us suppose, of some childhood association. Call the latter kind of case a *Rorschach* case because (like the Rorschach ink-blot test) it involves some kind of non-inferential association.

Rorschach cases present a problem for the informationalist analogous to the problem of error. The informationalist claims, for example, that a thinker's mental representation, "home," expresses the concept HOME in virtue of the representation's covariation relations with home. Rorschach cases raise the possibility that "home" may have the same covariation relations with cardamom. The informationalist needs to distinguish Rorschach correlations from content-determining ones. Intuitively, what the informationalist wants to say is that Rorschach occurrences of symbols are not applications of symbols. But how is the informationalist to draw this distinction between occurrences of symbols that are applications of the symbols and occurrences that are not?

The problem of Rorschach cases is arguably more fundamental than that of mistakes. In cases of systematic error, something has gone wrong. It is therefore natural to think that we can distinguish correct (and content-determining) from mistaken (and non-content-determining) applications by identifying the interfering factors or finding conditions that filter out those factors. In contrast, Rorschach cases seems to show that there is a problem for dispositional accounts even in cases in which nothing has gone wrong.

This difference is reflected in the fact that ideal conditions do not seem to be even the right category of remedy for the failures of correlation in Rorschach cases. We have no conception of the sort of conditions under which associations cannot lead from a perceptual experience to a Rorschach occurrence of a token of a mental representation. They would have to be conditions that would, for instance, guarantee that the visual experience of an American flag would not – without first causing her to think FLAG – lead a patriotic American to think AMERICA. We do not know how to specify such conditions even in (non-vacuous) *intentional* terms since we do not have a notion of what such conditions would consist in. And it seems plausible that there are no non-intentionally specifiable conditions under which a thinker is guaranteed not to make Rorschach associations.

What the informationalist needs, in order to solve the Rorschach problem, is a way of distinguishing applications of symbols from other occurrences of symbols; the distinction must be drawn in a way that uses only non-intentional properties of symbol-world correlations. (Similarly, the inferential-role theorist needs a way of distinguishing content-determining transitions from associative or Rorschach transitions.) There are a variety of approaches the informationalist could take, but none seems promising.

For example, the informationalist might suggest that the home/"home" correlation is not mediated by other mental representations, while genuine applications are always so mediated. The idea might be that cardamom causes occurrences of "home" by first causing occurrences of N, a mental representation that (intuitively) expresses the concept CARDAMOM (and then N reliably causes occurrences of "home"). But the informationalist is not warranted in assuming that Rorschach cases must be mediated by occurrences of other mental representations. (Nor is she warranted in assuming that *applications* of concepts cannot be mediated by occurrences of other mental representations.) So the informationalist cannot distinguish the cardamom/"home" correlation from the home/"home" correlation on the basis that the former is mediated by a mental representation other than "home."[25]

Similarly, suppose the informationalist tries to distinguish the cardamom/"home" correlation from the home/"home" correlation on the ground that, in the latter case, the occurrence of "home" is not embedded in a phrase in the language of thought. Again, the informationalist's assumption is unwarranted. There is no reason to assume that there cannot be cases in which cardamom produces unembedded occurrences of "home."

The informationalist may try to argue that applications can be distinguished from other occurrences of symbols by the reliability or strength of the symbol-world relations. The idea would be that, even if cardamom reliably causes thoughts of home, perhaps home even more reliably does so. Again, however, there is no justification for assuming that the correlations in Rorschach cases

[25] Actually, we need not rely on unmediated cases in order to head off the informationalist's approach. It is enough that there be cases in which, though the cardamom/"home" correlation is mediated by the occurrence of other mental representations, the mediation does not include any application of a concept to the cardamom. (Even those who do not accept the possibility of unconceptualized perceptual experience need not be committed to the position that the *object* that is perceived, as opposed to the experience, is conceptualized.) Thus, a Rorschach case could involve either a direct transition from a perceptual experience to an occurrence of a mental symbol, or a transition through a chain of symbol tokenings to the Rorschach occurrence – just as an application of a symbol to an object may or may not be based on inference.

cannot be as reliable or strong as the correlations between correct applications of mental representations and the objects they are applied to.[26] I do not claim to have made an unanswerable objection to dispositional theories. I have not made a serious attempt to rule out possible replies of the dispositional theorist. The point is rather that we will be more likely to reach an accurate diagnosis of the problem underlying Boghossian's belief-holism objection if we recognize that the basic difficulty is broader than false background beliefs.

3.4.2 The achievement view and the order of explanation

Let's take stock of the situation. Boghossian has pointed out that the dispositionalist cannot find a non-intentional way of distinguishing correct from incorrect applications of symbols because she cannot find a way of distinguishing applications that are based on false beliefs from those based on true beliefs. I have suggested that the dispositionalist will have a similar problem in filtering out erroneous applications of symbols (or transitions between mental states) that are the result of desires, fears, hopes, or the like. And I have further argued that the dispositionalist encounters a difficulty in separating occurrences of symbols that are applications from occurrences that are not. More generally, the existence of Rorschach occurrences of mental symbols and (Rorschach transitions between mental states) makes it difficult for the dispositionalist to identify the content-determining occurrences (or the canonical transitions).

At an intuitive level, the relevance to content of a person's uses of symbols seems to depend on the reasons behind those uses. Call this apparent dependence *reason dependence*. For example, an occurrence of a symbol that is the result of a false background belief, a desire or a fear, or a free association is not content-determining. In contrast, what seems to make an occurrence of a symbol an application – the informationalist's paradigm content-determining occurrence – is that the occurrence is the result of an intention to classify an object. This way of drawing the distinction relies on intentional notions, however. It also does not offer a general principle for distinguishing the dispositions that are relevant to content from those that are not. Dispositional theories of content need there to be a non-intentional way of distinguishing content-determining from non-content-determining uses of symbols. Thus, reason dependence is a *prima facie* problem for such theories.

[26] Fodor's asymmetric-dependence theory could be used as a sophisticated version of this strategy for replying to the Rorschach problem. I critique Fodor's theory in chapter 4.

Here is a conjecture as to a general principle that captures the intuitive distinction in non-intentional terms: what the content-determining dispositions seem to have in common and the other dispositions to lack is that the content-determining dispositions are dispositions to perform the actions that the thinker is required to perform, or dispositions to do what satisfies an applicable standard. For example, incorrect applications of symbols – whether they result from bad light, distractions, false background beliefs, desires, or fears – violate standards of correctness – that is what it is to be incorrect. What about Rorschach cases? As suggested above, they are not cases in which anything goes wrong. They are neither correct nor incorrect, but seem not to be subject to norms at all. Thus, I suggest that what distinguishes content-determining applications of concepts from Rorschach occurrences is that the former satisfy correctness standards and the latter do not. (There are two ways to fail to satisfy correctness standards – to violate them and not to be subject to them in the first place.)

The same kinds of points can be made with respect to inferential-role versions of dispositional theories. The content-determining transitions between mental states seem to be the ones that the thinker is required to make. An inferential-role theory might well hold that the transition from the judgment that something is a sofa to the judgment that it is a piece of furniture for sitting on is content-determining. Consider a thinker who judges that something is a sofa but, because of his beliefs about the Pope's pronouncements on sofas, is not disposed to judge (when he considers the question) that it is a piece of furniture for sitting on. This thinker plausibly violates a requirement on thought. (Of course, a theory that appeals to requirements on a thinker to explain the content of the thinker's thought will not be able to appeal to the content of the thinker's thought to explain the requirements to which the thinker is subject. That is, such a theory cannot hold that it is because the thinker's thought involves the concept SOFA that she is subject to the inferential requirement involved in the example. See sections 7.2, 7.5-7.7 below.) Similarly, transitions between mental states that are merely free associations are not content-determining.

Here is a first try at a diagnosis of the problem with dispositional theories (of which Boghossian's belief-holism problem is a manifestation). If the content-determining dispositions are simply the dispositions to accomplish the tasks that are required of the thinker, the tasks that satisfy the standards to which the thinker is subject, perhaps the dispositions are doing no explanatory work. The dispositionalist has things the wrong way around. She proposes various dispositions as the determinants of content, but in fact, they are at best rough proxies or substitutes for the norms that govern the dispositions.

For example, perhaps what a person is disposed to judge under ideal conditions gains whatever plausibility it has as a determinant of content because it approximates what the person should judge – what applicable norms of thought require.[27] So, the suggestion continues, Boghossian has shown that what a person is disposed to do under the best conditions we can specify is an imperfect approximation for what the person should do. But there's a more fundamental problem: the order of explanation seems to be the wrong way around for an ideal-conditions-based dispositional account to provide a constitutive account of content. What makes the conditions *ideal* is precisely that they are conditions under which I am able to satisfy applicable norms. It is what I should do – the norms to which I am subject – that are explanatorily prior, not what I would do under ideal conditions.

It might be objected that my suggestion gets us nowhere. The *starting point* for dispositional theories, the objector might say, is that the content of thought determines correctness conditions (and, if one is an inferential-role theorist, requirements governing transitions between mental states). According to the objector, dispositional theorists are trying to explain, in purely dispositional terms, the content of thought and, consequently, the associated correctness conditions.

This objection misses the gist of my diagnosis. A preliminary point (discussed further in section 7.5.2) is that although Boghossian himself regards dispositional theories as attempts to give an account of what it is in virtue of which thinkers are subject to conceptual norms (pp. 533-534), that is not the way in which dispositional theorists typically view their project. Dispositional theorists tend to deny that content is normative, or, at least, to deny that its normativity is something that a theory of content must account for (see, *e.g.*, Fodor 1990, pp. 128-130, 135 n. 35). They typically see their proposals as seeking to identify dispositions that correspond to or pick out a concept (by picking out a property or a set of canonical transitions that individuate the concept), while leaving room for the possibility of error (see section 1.5).

[27] In other contexts, a person could intentionally fail to do what he should do. But content is special in this respect. A person could on a particular occasion deliberately use a *linguistic expression* wrongly, and still use the expression with its ordinary meaning. But mental representations are not within our intentional control. So

But let us waive this point. Let us grant that dispositional theories attempt to give a constitutive explanation of the normativity of content. How do they go about this task? As the imagined interlocutor suggests, in practice, dispositional theorists try to specify dispositions the possession of which has the consequence that thinkers satisfy the requirements to which thinkers are intuitively subject. The point of my diagnosis is that this approach does not take the task of giving a constitutive account seriously. A constitutive account must not merely find a dispositional way of picking out the appropriate norms, it must say what it is in virtue of which the thinker is subject to the norms. That is, a constitutive account needs to have the order of explanation the right way around: the proposed dispositions must make it the case that the thinker should make certain judgments or transitions.

If we are trying to find dispositions that merely correspond, by some mapping, to the requirements to which the thinker is intuitively subject, there are indefinitely many possible candidates. But the most natural candidates are the dispositions whose possession makes it the case that a thinker satisfies the requirements. The achievement view is precisely the view that what it is to have a certain content is to have a disposition to satisfy the appropriate requirements. Thus, given a conception of the goal as finding dispositions that correspond to the appropriate requirements, the achievement view is very natural.

As I argue in chapter 6, however, if we are trying to find the dispositions that make it the case that a thinker is subject to the appropriate requirements, the achievement view seems much less apt. Being disposed to do something is not a good candidate for what makes it the case that you should do it.

This is not to claim that dispositional accounts could not find resources to account for norms. It might be that a thinker is subject to certain norms in virtue of his having certain dispositions. But we must not overlook the distinction between, on the one hand, the dispositions to satisfy the norms to which the thinker is subject and, on the other, the dispositions in virtue of which the thinker is subject to those norms. If my diagnosis is right, dispositional accounts should be seeking not the dispositions to satisfy the norms, but the dispositions that make the thinker subject to them.

My suggestion, in sum, is that achievement-based accounts have the order of explanation the

---

the issue of a deliberate failure to use a mental representation correctly does not arise.

wrong way around. They are seeking a way of picking out the norms in question – by picking out dispositions to satisfy those norms. Dispositions under ideal conditions will tend to be dispositions to satisfy governing norms. But it is not *because* thinkers have such dispositions that they are subject to the norms. (Indeed, as noted, it is plausible that conditions count as ideal at least in part because thinkers are disposed to satisfy governing norms under those conditions. It may also be that the norms are part of the explanation of why the thinkers, under ideal conditions, behave as they do.) A constitutive account of meaning must identify what it is – dispositions or something else – in virtue of which the thinker is subject to the appropriate norms.

It bears repeating that this diagnosis is not itself intended as an objection to the achievement view, but merely as a suggestion as to what underlies various objections to different achievement-based theories. We have reviewed Boghossian's arguments against two major types of achievement-based theories. In the next two chapters, I give extended critiques of two more prominent achievement-based theories. The point of the foregoing discussion is to lay the groundwork for the idea that there is a common underlying problem, and, in particular, that the achievement view is the culprit.

Chapter 4: Fodor's Asymmetric-Dependence Theory of Content

4.1 Introduction

In chapter 3, we saw that there are compelling reasons for rejecting reductionist theories of content that rely on specifying ideal conditions. As promised, we now turn to Fodor's asymmetric-dependence theory, which is an informational theory that avoids the need to specify ideal conditions.[1] The theory is cast in terms of laws covering the occurrence of an individual thinker's mental symbols. Rather than attempting to identify the content-determining applications of a symbol by specifying situations in which a thinker would apply the symbol only to instances of its reference property, Fodor maintains, roughly speaking, that the content-determining laws are those on which the other laws asymmetrically depend (in a sense to be elaborated).

My central argument (section 4.5) is that, even if the laws that the theory requires obtain, the laws do not in fact exhibit the appropriate asymmetric-dependence relations.[2] In a nutshell, I show that, in general, part of the mechanism for the crucial, supposedly content-determining law for a mental symbol is not shared by the mechanisms for the other laws covering the occurrence of the same mental symbol. As a result, the former law can be eliminated (by eliminating the non-overlapping part of the mechanism) without eliminating the latter laws. The latter laws do not asymmetrically depend on the former law.

Before I can reach my main arguments, I need to do some stage setting. In the next section (4.2), I outline the asymmetric-dependence theory. In section 4.3, I discuss Fodor's views about laws, which are crucial to an understanding of his theory. Next, in section 4.4, I examine and dispute Fodor's claims about what work he needs to do to support his theory and about what kinds of objections are legitimate. My substantive critique is contained in section 4.5.

---

[1] As Loewer and Rey (1991, pp. xxvi-xxvii) point out, the motivation of the asymmetric-dependence theory can be understood as very similar to that of ideal-conditions-based theories.
[2] Note that Fodor holds that the asymmetric-dependence conditions are sufficient, but not necessary, for content. Consequently, he insists that he need consider only cases in which it is stipulated that the conditions are satisfied. In section 4.4, I show that this position cannot be maintained.

4.2 The Asymmetric-Dependence Conditions

Fodor's conditions for a linguistic or mental symbol "S" to mean X are (pp. 93,

121): [3]

1) It is a law that Xs cause the tokening of "S"s[4] (information condition); and

2) if it is a law that any Ys (that are not Xs) cause "S"s, then that law is dependent on the law that Xs cause "S"s, but not the other way around (asymmetric-dependence condition).[5]

There are two further conditions that have an uncertain status in Fodor's account:

3) Some "S"s are actually caused by Xs (actual-history condition).

4) For some property Y (not = X), it is a law that Ys cause "S"s – *i.e.*, the asymmetric-dependence condition is not satisfied merely vacuously (robustness condition).[6]

The first two conditions form the core of the theory. They are Fodor's way of distinguishing a symbol's reference from other properties that reliably cause the symbol's occurrence – in other words, they provide Fodor's solution to the disjunction problem. Specifically, the two conditions implement Fodor's basic notion that of all the properties that are nomically related to (the property of causing occurrences of) a symbol, the symbol's reference is the property that figures in the nomic relation on which all the other nomic relations depend.

We will not need to be much concerned with the actual-history condition. Fodor tentatively proposed it as a way of preventing the theory from implying implausibly strong verificationist consequences (pp. 119-24), [7] but, as described below, he has subsequently repudiated it.

The idea behind the robustness condition is that it distinguishes semantic from (merely) informational nomic relations (pp. 92-93). True symbols are "robust," which means that they retain their meanings despite being reliably caused by properties other than those to which they refer (1990,

[3] In this chapter, all citations to page numbers alone are to Fodor 1990.
[4] For convenience, I will use Fodor's loose way of talking according to which a letter or word in quotation marks can stand for either a linguistic or mental symbol, or can be short for an occurrence or tokening of the symbol, as in *horses cause "horse"s* (shorthand for *horses cause occurrences of the symbol "horse").*
[5] It should be noted that Fodor's asymmetric-dependence account assumes (at least some elements of) the language-of-thought hypothesis, for which he argues elsewhere (*e.g.*, 1987, pp. 135-154; 1975). I will assume the language of thought hypothesis throughout this chapter in order to address Fodor's theory on its own ground.
[6] See Fodor 1990, p. 118; 1991a , p. 261.
[7] Roughly, the actual-history condition is intended to make ineligible for a symbol's reference any property that in fact has no causal impact on the thinker. This is designed to avoid the consequence (to which the theory apparently otherwise would be committed) that in order for "X" to mean *X*, it must be nomologically possible for the thinker to distinguish X "from any property that *would* cause "X"s *if it were instantiated*" (p. 122).

p. 91). If, to put it loosely, it is a law that a property causes an event of a certain type,[8] then events of that type are natural indicators and carry information about the property. (*E.g.*, it is a law that fire causes smoke, so smoke carries information about fire.) Thus, a natural indicator is not robust – it means whatever reliably causes it. In contrast, Fodor thinks, it can be a law that a property causes (occurrences of) a symbol, despite the fact that the symbol does not refer to that property. Symbols are robust – they retain their meaning, despite their occurring (both mistakenly and non-mistakenly) without being caused by their reference. (For example, "horse" means HORSE even if it is a law that cows on dark nights cause "horse" tokens and a law that thoughts of horses do.)

Fodor's theory presents a moving target.[9] The original formulation of the theory (1987, pp. 106-111) and its initial presentation in *A Theory of Content* (*TOC*) (1990, p. 89-100) include only the information and asymmetric-dependence conditions. Robustness enters the discussion in TOC as a feature of content that a theory must be able to explain (1990, pp. 90-93, 97-100), but in the course of the essay Fodor responds to an objection by treating the robustness condition as (a necessary) part of his sufficient conditions for a symbol's having content (p. 118). At the end of the essay, the robustness condition is not mentioned in the canonical specification of the theory (pp. 119-124). In subsequent work, however, Fodor continues to appeal to the robustness condition (1991a , p. 261). As for the actual-history condition, Fodor introduces it tentatively, then includes it in the canonical statement of the theory and treats it as a firm part of the theory (pp. 119-124), and, as noted above, finally rejects it (1994, pp. 115-119). My discussion focuses on the two principal conditions, but I will point out where appropriate that Fodor's account is not saved from my objections by either the robustness or actual-history condition.

4.3 Fodor on laws

Fodor thinks that appealing to laws, as opposed to counterfactuals, will enable him to avoid problems that are fatal to other informational accounts. It will be useful to examine Fodor's views about laws in order to clarify the framework within which his theory of content is situated. I will accept this framework, at least for the sake of argument.

[8] For expository ease, I will follow Fodor in using terms such as "law," "property," "cause," and "event" in a relaxed way, writing, *e.g.*, of properties causing events, or of laws that Xs cause Ys.
[9] In this paragraph, I am leaving aside significant shifts in Fodor's position from his 1987 to his 1990, see, *e.g.*, 1990, pp. 131-132, fn. 3, as well as further changes that are implicit in his 1998, pp. 120-145.

4.3.1 Preliminaries

We can begin with a few preliminary clarifications and terminological points. On Fodor's view, reliable causal relations are subsumed by laws. Laws hold of properties, not individuals. As Fodor does, I will often use "it is a law that As cause Bs" as a shorthand for "there is a nomic relation between the property of being an A and the property of causing Bs."[10]

Suppose Qs reliably cause Rs – that it is a counterfactual-supporting generalization that Qs cause Rs. It follows that there is a law that subsumes the causal transactions between Qs and Rs, but it is a further question whether the property of being a Q figures in that law. For example, if Qs are a subkind of Ts, Qs may cause Rs in virtue of being Ts rather than in virtue of being Qs. That is, the covering law may be that Ts cause Rs. Fodor gives the example that, though small horses reliably cause occurrences of the mental symbol "horse" (in a particular thinker), the covering law may be that horses cause "horse"s, not that small horses do (pp. 101-102).

The laws that figure in Fodor's theory are not paradigms of laws of nature. A law of the sort Fodor relies on is a nomic relation between a worldly property and a property of causing the occurrence of a particular mental symbol in an individual thinker's head. (A mental symbol here must be understood as a type, tokens of which can occur or be triggered on particular occasions.) An example is a nomic relation between the property of being a star and the property of causing the occurrence of a certain mental symbol – call it "star" – in my head (in Fodor's shorthand, the law that stars cause "star"s). Such a law is specific to an individual thinker's mental symbol; there is no suggestion in the theory that there are community- or language-wide laws. One notable way in which Fodor's laws are peculiar is their specificity to particular thinkers. A related oddity is that the laws are easily changeable. Learning, aging, and injury can create and eliminate laws.

A number of further terminological shorthands will be convenient. If there is a law covering the occurrence of some mental symbol, such that all other laws covering the occurrence of that symbol asymmetrically depend on that law, I will say that the law is *primary*. Suppose we are considering the laws covering the triggering of my mental symbol "S." We can refer to the law that As cause "S"s as "the A law," the law that Bs cause "S"s as "the B law," and so on. (In this terminology, Fodor's theory of content holds, very roughly, that "S" means *X* in virtue of the X law's

[10] See generally, Fodor 1987, p. 164 fn. 6; Fodor 1990, pp. 93-95.

being primary.)

When the reference of "S" is A, we can also call the A law "the reference law" or "the content-determining law" and the other laws "wild laws." Wild laws include laws covering occurrences of "S" that are not *applications* of "S" (in the sense defined in section 3.4.1), for example, occurrences of "S" that are caused by other thoughts. As discussed in chapter 3, such occurrences of a symbol need not be in any way mistakes. I will for the most part set aside this qualification until section 4.5, where laws covering non-application occurrences of symbols will become important to my argument.[11]

4.3.2 Ceteris-paribus laws as a response to Kripke's doubts about dispositions

The laws that figure in Fodor's asymmetric-dependence theory are obviously high-level laws, not the basic laws of physics. Such laws are not, in general, exceptionless;[12] they have implicit *ceteris-paribus* clauses.[13]

Fodor explains his appeal to *ceteris-paribus* laws by considering Kripke's (1982) argument that ideal conditions will not help to solve the Kripkenstein problem. After pointing out that our actual dispositions are imperfect, Kripke considers the possibility of appealing to *ceteris-paribus* dispositions. Even if a thinker would in fact make mistakes when apparently adding, perhaps what is relevant is what he would do other things being equal. Kripke rejects this idea on the ground that we are not justified in making assumptions about what would happen under all the varying circumstances encompassed by the *ceteris-paribus* clause. For example, he argues, we "have no idea" what a thinker would do if given enough extra brain capacity to handle very large numbers; he might "go insane" (Kripke 1982, p. 27). In effect, Kripke is pointing out that the ideal conditions to which dispositional accounts appeal are impossible conditions – conditions under which we have infinite memory and time, no distractions, and so on – and that evaluating the truth of counterfactuals with impossible antecedents is problematic. In possible-worlds terms, the closest possible worlds in

[11] See also note 23 below.
[12] I use the term "exception" in the most obvious sense, to refer to any case in which the antecedent of a law or putative law is satisfied, but the consequent is not, even if that case is consistent with the law's obtaining (for example, because the case is accounted for by interfering factors). This contrasts with the way in which Pietroski and Rey (1995), discussed below, use the term. In their terminology, an apparent exception is not an exception if it is consistent with the law's obtaining. Thus, in their terminology, even *ceteris-paribus* laws are exceptionless (1995, pp. 87-88).
[13] See Fodor 1991b.

which the physically or nomologically impossible conditions held might well be worlds that were wildly different from our world.

Reasonably enough, Fodor thinks that this objection is too swift. In his view, Kripke's mistake is to assume "that we can't have reason to accept that a generalization defined for ideal conditions is lawful unless we can specify the counterfactuals which would be true if the idealized conditions were to obtain" (p. 92). Fodor points out that in general it cannot be true that we have to be able to specify all that would happen if ideal conditions were to obtain in order to be justified in accepting that a generalization holds, *ceteris paribus*. In other words, we can be justified in relying on a *ceteris-paribus* generalization without being able to cash the *ceteris-paribus* clause into counterfactuals. The idea is that we can know how a system would behave in the absence of external interference, without knowing what all the possible interfering factors are – and, *a fortiori*, without being able to replace the *ceteris-paribus* clause with fleshed-out counterfactuals.[14] We can have reason to accept that you would add very large numbers correctly if the limitations of memory did not interfere, without having to worry about what would actually happen if you had extra brain matter.

4.3.3 Laws and mechanisms

Fodor thinks we can avoid the problematic counterfactuals by appealing directly to laws. He prefaces his account with an affirmation of faith that it is ontological "bedrock that the world contains properties and their nomic relations": nomic relations between properties are deeper than and therefore need not be analyzed in terms of counterfactual relations among individuals (p. 93). He is even more sure of the epistemological point: we can know that there is a nomic relation without knowing which corresponding counterfactuals hold. We can know that the ideal gas laws hold even if we have no idea what would happen under the physically impossible conditions the laws specify (pp. 93-95). Fodor's strategy is therefore to frame his account in terms of nomic relations, and then to resist, on these theoretical grounds, the challenge to specify all the corresponding counterfactuals.

[14] See pp. 93-95 & fn. 10. Fodor follows an unpublished manuscript of Georges Rey, which was a predecessor of Pietroski and Rey (1995), in maintaining that an ineliminable *ceteris-paribus* clause does not make a law vacuous. As Pietroski and Rey spell out the proposal, a *ceteris-paribus* law is true only if all cases in which the law apparently fails to hold can be explained by the interference of *independent* factors. In order to be independent, a factor "has to explain not only the apparent counter-example to the [*ceteris-paribus* law], but also something other than that failure" (1995, p. 90). As discussed in note 16 below, Pietroski and Rey's proposal does not address the conditions for a something to be a law; rather, it concerns only a sufficient condition for non-vacuity (pp. 91-93, 98-100).

For example, he thinks he can maintain that the property of being a horse is nomically related to the property of causing "horse"s (in a particular thinker), without establishing what the thinker would do and say under various specified conditions.

Suppose we grant that *ceteris-paribus* laws need not be vacuous and are (ontologically and epistemologically) prior to counterfactuals. There remains the question of when a regularity counts as a law. We saw that Fodor's appeal to *ceteris-paribus* laws is a response to the problem that the correlations between instantiations of properties and occurrences of symbols are imperfect. When faced with an apparent counterexample to a candidate law, Fodor need not find some way of specifying ideal conditions under which the counterexample will not occur. As long as mistakes can be attributed to other things not being equal, they do not undermine the existence of the law.

Of course, Fodor does not attribute every apparent counterexample to other things not being equal. He implicitly assumes a distinction between situations in which "other things" interfere with a law's operation and situations in which there is no law. Fodor's laws, as non-basic, are mediated by mechanisms. To say that other things are not equal is to say that something prevents the mechanism from operating in the way that it otherwise would. Thus, the distinction Fodor needs is between exceptions to a law – situations in which a law's mechanism is intact but conditions prevent its operation – and situations in which part or all of a law's mechanism is lacking and thus the law does not obtain.

In an article that Fodor (1990 p. 132, fn. 8) cites approvingly, Paul Pietroski and Georges Rey (1995)[15] express the role of *ceteris-paribus* laws in scientific explanation: "scientists state [*ceteris-paribus* laws] in an attempt to focus on particular factors . . . and thereby 'carve' complex phenomena . . . in a theoretically important way" (1995, p. 92). On this picture, most phenomena in the real world are the result of the complex interaction of many different systems. Science proceeds by considering isolated systems in abstraction from other external factors. It explains the phenomena as the result of the interaction of these different systems. Thus, the question whether a generalization that is subject to apparent counterexamples is nevertheless a law is the question whether the best

[15] As noted above, Fodor cites what was then an unpublished manuscript of Georges Rey, which was a predecessor of Pietroski and Rey (1995).

scientific explanation of the phenomena – of the cases in which the generalization holds and the cases in which it does not – holds that the apparent counterexamples are the consequence of interactions between an underlying system whose behavior corresponds to the generalization and factors external to that system. In other words, whether a regularity is a law depends on whether, according to the best science, there is an intact underlying mechanism.

I don't know how to give a rigorous account of the distinction between absence of a mechanism and interfering factors, and Fodor offers no help.[16] But it is clear that Fodor's appeal to *ceteris paribus* laws requires such a distinction. Also, as we will see, Fodor implicitly relies on the idea that eliminating a law's mechanism eliminates the law.

4.3.4 Asymmetric dependence

Central to Fodor's theory is the notion of asymmetric dependence. For law B to be asymmetrically dependent on law A is for law B to depend (ontologically) on law A but not *vice versa*. For law B ontologically to depend on law A is for the obtaining of law A to be part of what makes law B obtain.

The relevant kind of dependence is not historical or across time – "diachronic" – but "synchronic" (p. 134 fn. 18; 1987, p. 109).[17] In other words, the question is not whether the existence (or elimination) of one law would, over time, lead to the existence (or elimination) of another, but whether, at a particular point in time, one law ontologically depends on another – in other words, whether the latter law's obtaining is part of what now makes the former law obtain.

Just as laws are more basic than counterfactuals, and thus are not analyzable in terms of them, the same is true of the relations between laws (pp. 93-95). Thus, strictly speaking, counterfactual formulations do not capture what dependence (or asymmetric dependence) is – rather,

---

[16] Fodor (p. 132, fn. 8) quotes a passage from Rey that, taken out of context, might sound as if it offers a way of spelling out sufficient conditions for a generalization to be a *ceteris-paribus* law: "The viability of a *ceteris-paribus* clause depends not upon the actual specification or realizability of the idealization, but rather upon whether the apparent exceptions to the law can be explained as due to independently specifiable interference." Pietroski and Rey emphasize, however, that they do not purport to offer sufficient conditions for a generalization's counting as a *ceteris-paribus* law, but only to offer sufficient conditions for the non-vacuity of a *ceteris-paribus* clause, on a narrow reading of vacuity. See, *e.g.*, 1995, pp. 98-99. Accordingly, they define an "independent" factor as one that explains something other than the failure of the generalization (in a sense that they specify more precisely) (pp. 89-90). Given this definition, the condition that any exception to a generalization be explained by an independent factor, cannot be (and does not purport to be) a sufficient condition for the generalization to count as a law. As Pietroski and Rey write, as far as their proposal goes, "every singular causal claim of the form A caused B is a candidate for an interferable [*ceteris-paribus* law]" (1995, p. 99). They make no attempt to distinguish situations in which a putative law lacks a mechanism (and thus the law does not obtain) from situations in which the law's mechanism exists but is interfered with.

[17] See text accompanying note 37 below.

the dependence is what makes the counterfactuals true.[18] In practice, however, Fodor often treats the claim that law B is asymmetrically dependent on law A as equivalent to the counterfactual (indeed, counterlegal) claim that, *ceteris paribus*, if law A did not exist, law B would not exist (or, as he often puts it, that, *ceteris paribus*, eliminating law A would eliminate law B), but not the converse.[19] Subsequently, in response to an objection from Paul Boghossian, Fodor offers a different formulation according to which law B's asymmetric dependence on law A amounts to the following: other things being equal, you would have to change more to eliminate law A without eliminating law B than to eliminate law B without eliminating law A (1991a , p. 276). (Also, although he resists the idea that possible-worlds translations are needed (p. 95; 1991a , p. 271), Fodor offers such translations and often relies on them (113-116; 1991a , p. 276-277).[20])

It may be harmless for most purposes to follow Fodor in treating the dependence of law B on law A as tantamount to the counterfactual truth that, *ceteris paribus*, law B would not obtain were it not for law A's obtaining. Nevertheless, I want to pause to point out that my argument does not depend on a conflation of asymmetric dependence with the corresponding counterfactuals. First, as Fodor makes clear, the claim that one law is asymmetrically dependent on another *entails* the corresponding counterfactual claim (p. 95). Since I will be arguing that the asymmetric dependences to which Fodor's theory is committed do not hold, it would be sufficient for my purposes to show that the corresponding counterfactuals do not hold. Second, it is difficult to evaluate the truth of the relevant counterfactuals in the abstract. Thus, in order to evaluate the claim that one law depends on another (or the corresponding counterfactual claims), I will examine the mechanisms that underwrite the laws, and the relations between those mechanisms – which, as we will see, is the method that

[18] In his 1987 (pp.108-109), Fodor offered the counterfactual formulation as an analysis of asymmetric dependence. In his 1990, he resists the idea that it is an analysis, but, as explained in the text, he nevertheless works with it in practice.

[19] See, *e.g.*, pp. 91, 93, 107, 112-113, 115-116. ("*but that* "cow" *tokens carry information about cows, they wouldn't carry information about anything*;" "No noncow-caused "cow"s without cow-caused "cow"s;" p. 91 (italics in original).)

[20] In earlier discussions, Fodor often works with possible-worlds formulations of asymmetric dependence – there is some variation among them – that hold, roughly, that law B asymmetrically depends on law A if and only if there are (nearby) possible worlds in which A obtains and B does not, and they are nearer to our world than any world in which B obtains and A does not (1987, p. 109; 1990, pp. 113; 1991a , p. 276). When Fodor introduces the later understanding of asymmetric dependence described in the text, he explains that, in possible worlds terms, the claim that law B asymmetrically depends on law A should be understood as the claim that, other things being equal, worlds in which A holds but B does not "are closer to us than *corresponding* worlds" in which B holds but A does not (1991a , p. 276, emphasis added). As we will see, there may be diverse ways of changing the world so as to eliminate law A or law B. The point is that we are to compare each way of eliminating law A with the corresponding way of eliminating B – the way of eliminating B that is most similar to it. As I make clear in section 4.5, my arguments do not depend on the precise possible-worlds treatment of asymmetric dependence.

Fodor generally employs. In other words, rather than determining dependences between laws by evaluating counterfactuals, I will determine dependences (and indeed evaluate the corresponding counterfactuals) by considering the relations between the mechanisms that sustain the laws. Thus, the counterfactual formulations will not play an essential role in the argument.

4.3.5 The lawfulness of systematic mistakes

Since Fodor's theory addresses the disjunction problem by holding that wild laws asymmetrically depend on reference laws, Fodor holds that thinkers' dispositions to make mistakes can be lawful. According to his view, if a thinker is systematically disposed to mistake cows on a dark night for horses, there will be a law under which cows on a dark night cause "horse"s.[21] (Of course, the relevant law might not be that cows on a dark night cause "horse"s; the kind that figures in the law might be, for example, horsey-appearing objects. See notes 32-33 below.) This is just an application of his more general position that "it is necessary and sufficient for such reliable causation that there be a nomological – lawful – relation between certain (higher-order) properties of events" (1987, p. 99).[22]

In fact, it is easy to see that Fodor's asymmetric-dependence theory would be superfluous (and, because of the robustness condition, its conditions for content would never be instantiated) if properties other than a symbol's reference were not nomically related to the causing of the symbol. The problem at which the theory is directed is that there will be multiple laws governing the causation of a given symbol. Fodor's solution is that the law on which the other laws asymmetrically depend is the content-determining law. "Horse" means HORSE because all laws covering the

[21] Many philosophers seem to have assumed that mistakes are not supposed to be lawful on Fodor's account. A careful reading leaves no doubt that this is incorrect (though perhaps Fodor is not quite as explicit about this point as he should have been) (*e.g.*, pp. 93, 121-122, 126). Indeed, as explained in the text, the asymmetric-dependence theory would not be needed to address mistaken applications if they were not lawful. In *A Theory of Content*, Fodor manages to confuse even himself with respect to an immediate implication of the nomic status of mistakes. As discussed below (section 4.6.3), he maintains there that in order for a mistake to be a mistake and not content-determining, it must be nomologically possible for the thinker to avoid the mistake: "An organism can't have a kind of symbol which it necessarily misapplies, *i.e.*, which it misapplies in every world consonant with its psychology" (p. 108; see also p. 132 fn. 10). For, Fodor argues, if the putatively mistaken application were necessary given the organism's psychology, the application would be correct (p. 108). Fodor later (1991a, p. 263) recognizes that since the mistaken connections are nomic, it is nomologically (but not physically) impossible for the thinker to avoid them.

[22] Fodor thus eschews the implausible approaches 1) of denying the possibility of systematic mistake (see, *e.g.*, pp. 90-91) and 2) of maintaining that a symbol's connections to its reference are nomic, but that its systematic connections to other properties are not. If laws are the kind of exception-ridden generalizations that Fodor needs them to be, systematic mistakes will be nomic.

causing of "horse"s by non-horses are asymmetrically dependent on the law that horses cause "horse"s.

Fodor goes further than merely holding that systematic error is a problem that a theory of content must address. As we saw, he conjectures that robustness is what distinguishes a symbol from a mere natural sign. Thus, on his view, the existence of multiple laws governing a symbol's causation is partially constitutive of content (pp. 90-93, 99-100, 117-119, 128-130). Accordingly, Fodor imposes the robustness condition, which requires that, in order for a symbol to have content, there be at least one nomic relation between a symbol and a property other than its reference.[23]

4.3.6 High-level laws and unorthodox "natural" kinds

As noted, the laws that figure in Fodor's theory are high-level – "special science" – laws. Fodor cannot and does not take the view that only (natural) kinds of basic, or even relatively low-level, sciences can figure in laws. On his view of content, after all, any property that cannot figure in a law cannot be the content of a (primitive) symbol.[24] In order for "red," "messy," "fuzzy," "vase," "rage," and so on to mean what they do, it has to be the case that it is a law that redness causes occurrences of "red," that messiness causes occurrences of "messy," and so on.[25] Thus, Fodor's account requires the possibility that properties that are gerrymandered from the point of view of relatively low-level sciences can figure in the relevant laws. For it would be absurd to rule out the possibility of having concepts with references that are not physical-, chemical-, or biological-kind properties. On Fodor's account, then, highly non-natural, interest-relative properties will qualify as "natural kinds" – kinds that figure in scientific laws.[26]

---

[23] The reliable causal relations between properties other than a symbol's reference and the causing of a symbol's occurrence include not only those relations that are intuitively mistaken, but also those that involve non-mistaken associations or chains of thought (pp. 80-81; 90-91). See section 3.4.1 above. As noted above, for convenience, I often omit this qualification in the present discussion.

[24] The qualification "primitive" does not make much difference to the point. Fodor holds that most mental representations for lexical concepts are primitive. And, anyway, it is not plausible that most ordinary concepts are definable in terms of concepts of relatively low-level sciences.

[25] As he must, Fodor has explicitly accepted that properties like that of being a shirt or being crumpled can figure in laws (1991a , pp. 256-257). He thus rejects his earlier position (1986) that such properties were anomic. More precisely, according to his 1986, to be an intentional system is to be able to respond selectively to a property "even though no law connects the property it responds to and the selective property of its response" (1986, p. 14). See Antony and Levine (1991).

[26] In his 1998, Fodor takes the view that such unorthodox laws (and kinds) are laws about our minds, not genuinely laws about, for example, doorknobs, coat-hangers, or messiness.

4.4 Mere sufficient conditions and who gets to stipulate the counterfactuals

Fodor emphasizes that he is making only the limited claim that he provides *sufficient* conditions for a state to have particular content. Although his discussion often hints at larger aspirations, Fodor at least officially does not claim that any actual mental states ever satisfy his conditions. He explicitly leaves it open that all or some of our mental states have their content because they satisfy different conditions (p. 131).

Fodor maintains that since his conditions purport only to be sufficient for content, he can restrict consideration to hypothetical cases in which he stipulates that his conditions are satisfied: "Don't forget, this stuff is supposed to be philosophy. . . . I get to *stipulate* the counterfactuals. It's enough if I can make good the claim that 'X' *would mean* such and such if so and so *were to be* the case" (p. 96). Accordingly, in replying to objections, he protests repeatedly that attempts to show that his conditions do not hold in actual cases are irrelevant (*e.g.*, p. 102; 1991a, pp. 258-59, 263). I have several grounds for rejecting Fodor's view of the dialectical position.

1) It is false that one who offers (mere) sufficient conditions for content need only defend the view in hypothetical cases in which a symbol is stipulated to satisfy the conditions. Fodor seems to fail to recognize that his sufficient conditions for content imply necessary conditions. In order for a thinker to have an attitude involving a certain content, it is *necessary* that the thinker *not* satisfy Fodor's sufficient conditions for having a different content. If actual thinkers' mental symbols do satisfy his conditions for meaning something other than what they intuitively mean, it would threaten even his limited official claim that he has provided conditions such that, if a symbol ever satisfied them, the symbol would have the indicated content. Thus, it is not enough for Fodor to show that hypothetical thinkers who satisfied his conditions would have the appropriate contents; he also needs to show that his conditions do not assign intuitively incorrect contents to actual thinkers.

2) We have seen that Fodor takes the position that all he needs to do is to show that a symbol that satisfied his conditions for meaning *X* would mean *X*. He combines this position with a specific view about what is required in order to show that a symbol means *X*. The view is that it is necessary to show only that the symbol would carry information about Xs and that occurrences of the symbol can be (systematically) caused by non-Xs. I will now argue that this view – given Fodor's position

that he need show only that a symbol that is stipulated to satisfy his conditions would have the appropriate meaning – illegitimately skews the dialectical position in Fodor's favor.

Fodor assumes that if he can give an informational account that allows for symbol tokens that are reliably caused by objects not within the symbol's extension, he has succeeded in naturalizing content (p. 89). He is explicit on this point: "I've tacitly assumed throughout this paper that if you can get a theory of content that squares the intuition that "X" means *X* only if "X" tokens carry information about X-instantiation with the intuition that "X" means *X* only if you can have X-tokens that aren't caused by Xs, then you've done all that a solution to Brentano's problem is required to do" (p. 128).[27] To put it another way, he assumes that to have both information and robustness is to have content (p. 128).[28] Accordingly, the bulk of his argument is concerned with showing that the conditions would allow for error (pp. 89-92, 122, 127-128).[29] (He also offers some intuitive considerations in support of his theory, which I discuss immediately below.)

We can now see why Fodor's restricting consideration to cases in which he stipulates the counterfactuals has a peculiar effect when combined with his assumption that showing that his conditions allow for error is all that he needs to do. If we restrict attention to hypothetical cases in which we stipulate the laws (or dispositions or causal relations), allowing for error is too easy. The problem is that we can simply stipulate counterfactual situations in which the person's erroneous applications differ in some crucial way from the non-erroneous applications. For example, Fodor stipulates counterfactuals in which the person's erroneous applications asymmetrically depend on his non-erroneous ones.

Suppose, by way of comparison, that my theory is that possessing property Z is sufficient for laws (or dispositions or causal relations) to be content determining. By Fodor's rules, I get to stipulate the counterfactuals. So I will of course stipulate counterfactuals in which all the wild laws lack property Z and all the (intuitively) content-determining ones possess property Z. Consequently,

[27] Fodor uses the term "disjunction problem" for the problem of providing such an account, but, as the quotation in the text makes clear, he is using that term to mean only the problem of avoiding the consequence that systematic error is not possible – that is, of finding a theory that makes it possible to have occurrences of a symbol that are caused by Ys without it following that Ys are instances of the symbol's reference.
[28] Fodor goes on to toy inconclusively and very briefly with the idea that perhaps we need "to throw in" some consciousness or normativity. He also suggests that the psychological might be a superset of the intentional and thus that conditions for intentional content might not be conditions for being in any psychological state (pp. 128-130).
[29] He also tries to show that he can handle some philosophical problems for dispositional accounts of meaning. For example, in addition to addressing variations on the disjunction problem (pp. 101-110), he tries to show that his account can account for contents that refer to uninstantiated properties (pp. 100-101, 123-124) and does not have unduly verificationist implications (pp. 119-123).

by Fodor's lights, my theory solves the problem of error. Thus, Fodor's rules make it trivial to solve – or more accurately, to avoid facing – the problem of error. To put it another way, the problem of accounting for error is a problem that arises in real cases – it is the problem of finding in actual cases a property that both distinguishes the wild laws from the (intuitively) content-determining ones and is a plausible candidate for a constitutive element of content. If we limit ourselves to stipulated hypothetical cases, we do not address the real problem of accounting for error, but stipulate it away. In sum, in order to evaluate whether Fodor has solved the problem of allowing for error, we need to ask whether wild laws asymmetrically depend on reference laws in actual cases. (Or else we need to recognize that allowing for error in stipulated cases does not show that one's proposed conditions are sufficient for content.)

3) It is revealing to consider Fodor's attempt to explain the intuitions that lie behind his theory. Fodor's discussion has the following structure. First, Fodor considers our actual linguistic practices: "Some of our linguistic practices presuppose some of our others, and it's plausible that practices of applying terms … are at the bottom of the pile" (p. 97). He goes on to suggest that non-information-carrying symbol tokens asymmetrically depend on the information-carrying ones (p. 98). Next, Fodor points out that the mechanisms that sustain these asymmetric dependences are our linguistic intentions and policies (pp. 98-99). Finally, Fodor speculates that the asymmetric-dependence relations, rather than the intentions or other mechanisms that mediate them, might constitute content (pp. 98-100).

My point here is not to criticize this argument. Rather, I want to note that Fodor is attempting to support his theory by showing that, in actual cases, asymmetric dependences plausibly fall the way his theory suggests. With some reason, he thinks that this claim lends intuitive support to his speculation that the asymmetric dependences are sufficient for content. (In other places, Fodor similarly tries to show that the asymmetric dependences would fall in the way his theory specifies in actual cases (*e.g.*, p. 115).[30] But if he can support his theory by arguing that actual cases exhibit a suggestive pattern of asymmetric dependences, it is surely fair to respond that, in fact, actual cases do not exhibit the appropriate asymmetric dependences. The point is not merely *ad hominem*. There is surely some plausibility to Fodor's idea that the existence of asymmetric dependences in actual cases

[30] See also section 4.5 below.

bears on whether asymmetric dependence is constitutive of content. Why should we think asymmetric dependence has anything to do with content, if we find no asymmetric dependence in the only cases of content with which we are familiar? (My arguments suggest that in the actual world, it is in general not the case that, among the laws covering the occurrence of a symbol, one law is primary. If this conclusion were correct, it would at least weaken the case that asymmetric dependence is a plausible candidate for a constitutive account of content.)

In light of the foregoing discussion, I therefore consider it fair game, in evaluating Fodor's theory, to ask whether he has plausibly accounted for actual cases. Given Fodor's insistence that he purports to provide only sufficient conditions for content, I want to emphasize this point. Thus, once again: *I will not limit the discussion to cases in which the appropriate nomic relations (and dependences between nomic relations) obtain, but will examine whether it is plausible that they do obtain in actual cases (at least typically).*

4.5 Why the star law is not the primary law

In this section, I argue that the asymmetric dependences do not fall the way Fodor needs them to. For expository purposes, I will use the example of my (or any person's) linguistic or mental symbol "star." As we have seen, on Fodor's theory, "star" has the content STAR because the following conditions are satisfied.[31] First, it is a law that stars cause "star" to be triggered. Second, there are other, wild laws covering the occurrence of "star." Third, all wild laws are asymmetrically dependent on the star law.

For the purpose of challenging Fodor's assumptions about which laws asymmetrically depend on which, we can grant Fodor the *existence* of the laws. Thus, let us assume that it is a law that stars cause "star"s and that it is a law that, say, relatively slow-moving comets cause "star"s. (Suppose I systematically mistake such slow comets for stars.)[32][33]

---

[31] Once again, as argued in section 4.4, although Fodor's official position is not that his theory explains the contents of any actual person's words or thoughts, it is warranted to explore his theory's plausibility as an account of actual cases.

[32] As we've seen (sections 4.3.5-4.3.6), Fodor's account depends on there being wild laws for every meaningful symbol. Nothing will depend on the exact details of the example; we could equally use a law that electrical storms, catastrophic cosmic collisions, planets, or moons of certain kinds (or under certain conditions) cause "star"s. My examples, like Fodor's, are somewhat unrealistic for purposes of expository simplicity. In light of two qualifications, however, the examples are less unrealistic than may at first appear. First, as noted, Fodor's theory concerns the contents of symbols of individual thinkers, so our assumption is only that a single thinker systematically makes the relevant mistakes. Second, to say that a connection, whether involving a mistake by the thinker or not, is nomic is not to say that it occurs in an exceptionless way. So there may be a slow-comet law even though slow comets do not cause "star"s when they are close to the horizon or seen through a telescope (say). In fact, it is consistent that it is a law that slow comets cause "star"s

As we saw, Fodor's rough test for whether one law asymmetrically depends on another is whether, other things being equal, eliminating the latter law would eliminate the former, but not *vice versa*. Since he thinks that the laws and their dependence relations are the consequence of mechanisms (pp. 99-100; 1991a , p. 313 fn. 8), he answers the test question by examining the mechanisms that he assumes underwrite the relevant laws (pp. 113, 115-116, 117; 1991a , pp. 276-277). The example that Fodor develops the most concerns why the law that XYZ causes occurrences of "water" asymmetrically depends on the law that water causes occurrences of "water."

> My point is that the intention to use "water" only of stuff of the same kind as the local samples has the effect of making its applications to XYZ asymmetrically dependent on its applications to H2O *ceteris paribus*. Given that people are disposed to treat "water" as a kind term . . . it follows that – all else equal – they would apply it to XYZ only when they would apply it to H2O; specifically, they would apply it to XYZ only when they *mistake* XYZ for H2O; only when (and only because) they can't tell XYZ and H2O apart. Whereas, given a world in which they *can* tell XYZ and H2O apart (and in which their intentions with respect to "water" are the same as they are in *this* world), they will continue to apply "water" to H2O and refrain from applying it to XYZ (p. 115).

Fodor uses the same basic form of argument in other kinds of cases (*e.g.*, pp. 112-114, 116). When Fodor introduces his later formulation of asymmetric dependence,[34] he gives a useful clarification of the argument: "to get to the [kind of world in which only H2O causes 'water' tokens], you'd only have to make XYZ and H2O distinguishable; to get to the [kind of world in which only XYZ causes 'water' tokens] you'd have to both do that *and* alter the mechanisms that underlie our intentions to use 'water' only of stuff that's of the same kind as our local samples." (1991a , p. 277).

---

and that it is not a law that slow comets close to the horizon cause "star"s. The argument of this section will offer something toward a more precise understanding of what counts as a (*ceteris paribus*) law and an interfering factor.

[33] In my view, it is more plausible that the relevant wild law is not that slow comets cause "star"s but that, very roughly, things with a starry appearance – the appearance that stars have in the *actual* world – do. I don't have space to argue for this point here, but it would only strengthen my argument; the law that starry-appearing things cause "star"s, unlike the slow-comet law, can be eliminated only by making changes in the thinker's semantic dispositions or intentions. (See the discussion in text below.) Moreover, in other work, I argue that stars themselves cause "star"s not in virtue of their being stars but in virtue of their having a starry appearance. If that is right, Fodor's theory does not get off the ground since it is not a law that stars cause "star"s. Thus, in order to grant Fodor the strongest case, I begin with the assumptions in the text.

[34] See section 4.3.4 above.

We can apply Fodor's argument to the slow-comet/star example. In order for one law to hold, but not the other, I (the thinker) have to be able to tell slow comets and stars apart. (As Fodor does, I'll consider the case of linguistic symbols for simplicity.) If we change things so as to make slow comets and stars distinguishable, but change nothing else, my intentions with respect to the use of "star" (or the mechanisms that underlie those intentions) will ensure that I continue to apply "star"s to stars. But I will no longer apply "star"s to slow comets. In contrast, Fodor thinks, in order to have the slow-comet law but not the star law, we would need both to change things so as to make slow comets and stars distinguishable and to change my intentions with respect to the use of "star."

Fodor's reasoning assumes that my intentions about how to use "star" (and the mechanisms that implement those intentions) are the crucial mechanism underwriting the star law. His view is that the only way to eliminate the star/"star" connection is to change my intentions with respect to the use of "star;" in other words, to change the meaning of "star." Now it is uncontroversial that the various uses of a term depend on its having the meaning that it does. That is, if I changed my intentions with respect to the use of "star," thus changing its meaning, the wild laws, including the slow-comet law, would, as a rule (*i.e.*, barring coincidence), be eliminated along with the star law.

But – and this is the crucial point – there is another way to eliminate the star law, one that does not involve changing our intentions, or anything about our psychology. We can change the stars themselves so that, even when there are no interfering factors, even when conditions are ideal, the stars do not have the effects that they dependably have in the actual world. For example, we can cool them down or change their composition or surround them with opaque clouds of dust. Since I (the thinker) am unchanged – it is not legitimate to take into account the changes in me that would, over time, come about as I took into account the changes in the stars – such a change in the stars will prevent me from applying "star"s to stars, even when there are no interfering factors. But such a change need not alter the way in which slow comets (and other starry-appearing non-stars) affect us. All the talk of "changes" may be misleading. The point is that there must be worlds in which a non-psychological part of the (actual world) mechanism for the star law does not exist.

Let us take the argument step by step. The crucial first step is that part of the mechanism underwriting the star law (or any nomic relation between a worldly property and the property of causing occurrences of a symbol) must be outside the thinker. Here is the argument for this first step. As we have already seen, (non-basic) laws require mechanisms. The thinker's intentions cannot be the entire mechanism for the star law. In order for there to be a real regularity connecting a worldly property to a symbol, the world must do its part. The star and the rest of the world around the thinker have to have certain dependable effects. And the thinker must have certain sensory equipment, beliefs (for example, about how stars look and how they affect various instruments), and intentions about how to use the word "star."

Fodor recognizes that the world must do its part. First, he emphasizes that nothing purely mental is sufficient for content – there must be real patterns of causal dependences (pp. 98-99).[35] Second, he insists that in order for there to be a law that Xs cause "X"s, there must be a world consistent with the thinker's psychology in which the thinker can discriminate Xs from all non-Xs that cause "X"s (at least all non-Xs that do so in the actual world) (pp. 122-123). That is, it must be consistent with the thinker's psychology that she can distinguish Xs from all nonXs (that cause "X"s in the actual world). To say that the thinker must be able to discriminate stars from all other causes of "star"s is to say something about the world as well as about the thinker. The world must be such that stars and non-stars have effects on thinkers that are different enough for the thinker to be able reliably to distinguish them. If intentions were enough for there to be a star law, it would not be necessary for the thinker to be able to discriminate stars from other things.

We have established that part of the mechanism that underwrites the star law must be located outside the thinker. The conclusion that the star law can be eliminated without changing our intentions or anything about our psychology follows immediately. If laws depend on mechanisms, and a (non-redundant) part of the mechanism for the star law is out in the world, then the world can be changed in such as way as to eliminate the mechanism, without changing the thinker.

It might be objected that whatever "change" we make will just be the addition of an interfering factor. The objection cannot be correct. We can see this without having an account of how to draw the distinction between the presence of an interfering factor and the absence of the mechanism. There must be such a distinction, and that is enough for our purposes. Suppose, for *reductio*, that every change counts as an interfering factor. In that case, removing the mechanism for

[35] Fodor is explicit on this point: "words can't have their meanings *just* because their users undertake to pursue some or other linguistic policies; or indeed; just because of any purely *mental* phenomenon, anything that happens purely 'in your head'…. For there to be a relation between "John" and John, something has to happen

a law counts as introducing an interfering factor. It follows that in a world in which there is no mechanism for a given law, the law still holds. In other words, if every change counts as an interfering factor, laws do not require mechanisms. Every law obtains in every world; the absence of a mechanism is just an interfering factor. But this is absurd. Thus, for every part of a mechanism, there must be changes that do not introduce interfering factors but eliminate the part of the mechanism.

Since part of the mechanism for the star law is in the world outside the thinker, there must be changes in the world outside the thinker that constitute the elimination of part of the star law's mechanism. Once again, the basic point is just this: since part of the mechanism underwriting the star law must be located outside the thinker, it must be possible to eliminate the law without changing the thinker.

Now we can spell out the conclusion of the argument. If the world were different in such a way that stars did not have the dependable effects that they actually have, it would not be a law that stars cause "stars." Suppose, for example, that, while holding everything else constant, including everything internal to the thinker, we drastically lowered the temperature of the stars or changed the facts of chemistry or optics in a way that prevented light from escaping from the stars.[36] As noted, Fodor is clear that such counterfactuals must be understood *synchronically* rather than diachronically (p. 134 fn. 18; 1987, p. 109).[37] That is, the point is not that if stars had looked different, we would not have applied "star" to them – presumably we would have done so because our beliefs and theories would have developed differently. (Similarly, it is beside the point that if stars changed their appearance we would eventually develop new beliefs and theories that would enable us to apply "star" to stars.) Rather, the point is that we, as we actually are, would not apply "star" to stars if stars were changed in such a way that, even with no interfering factors, they would not have any of their characteristic effects on us or on our instruments.[38] The same goes for scientific experts. Given the

in the *world*" (pp. 98-99). See also Fodor 1991a , pp. 289-290.

[36] Compare Kripke's (1980, p. 118) example in which it turns out that gold is actually blue, and that an optical illusion has caused us to think it is yellow.

[37] See also Loewer and Rey (1991, p. xxxvi fn. 51). Admitting diachronic effects would undermine Fodor's theory since, as Fodor points out, diachronic dependences cannot be expected to be the right way around (1987, p. 109).

[38] It might be objected that if either stars or slow comets were different in a way that affected their appearance, thus making them look different to us, then, given our linguistic intentions or dispositions, we would apply the term only to the stars. Thus, the objection continues, other things being equal, it is easier to eliminate the slow-comet law. The problem with this line of reasoning is that it evaluates the dependence diachronically rather than synchronically. It is true that if stars and comets had always looked different, whether because stars were different or because comets were different, we would have applied the term "star" only to stars. But given how

theories and methods they use in the actual world to detect stars, they would no longer apply "star" to stars if stars were changed in the way I have described. (It would of course beg the question to say that we would apply "star" to (the changed) stars if we were told that they were stars. The informational theorist cannot simply grant himself a reliable star detector and the knowledge that it is one. For in that case, his position would amount to the view that to mean STAR by "star" is to intend to apply "star" to whatever a reliable star detector tells us is a "star" – not an informational theory at all.)

If we make the minimal change in the world that is necessary for stars not to have starry appearances, slow comets will still have starry appearances. Since the thinker is unchanged, the thinker will still apply "star"s to objects with starry appearances. So the law that slow comets cause "star"s still obtains. Thus, *prima facie*, eliminating the star law does not eliminate the slow-comet law, so the latter law does not depend on the former. What about the other way around? Just as we can eliminate the star law by changing the appearance of stars, we can eliminate the slow-comet law by changing the appearance of slow comets (suppose for example, that they are surrounded by clouds of dust). And that change will leave the star law intact. At least with respect to this kind of case, it seems that the two laws' situations are symmetrical; neither law depends on the other. In sum, it is false that, *ceteris paribus*, eliminating the star law eliminates the slow-comet law. Therefore, the slow-comet law is not asymmetrically dependent on the star law.

So far I have been arguing that, by changing the stars, the star law can be eliminated without eliminating the slow-comet law (and that, in a parallel way, the slow-comet law can be eliminated without eliminating the star law). It might be objected that there are other ways of eliminating the star law that would eliminate the slow-comet law. As we have seen, Fodor's preferred way of eliminating the star law is to change my intentions with respect to the word "star". Eliminating the star law *in this way* will in general eliminate the slow-comet law as well.

Now the mere fact that *some* ways of eliminating the star law will eliminate the slow-comet law is irrelevant. An asymmetry is required. But the objector's basic idea is to question whether the counterfactual situations that I have considered are the relevant ones. Thus, the objection raises the issue of the proper way of understanding asymmetric dependence.

---

we now apply "star," if stars were made to look different, we would now – before we had time to adjust – apply the term to the things that looked the way stars used to look.

As Fodor emphasizes, the fundamental issue is not about counterfactuals (or possible worlds interpretations thereof), but about the ontological dependence of one law on another. I will now argue that we can resolve the issue of whether the slow-comet law asymmetrically depends on the star law by directly examining the relations between the laws' mechanisms. We need not enter into the subsidiary, technical, and tricky, issues about which counterfactuals are relevant.[39]

The fundamental issue is whether the slow-comet law's existence depends on the star law's existence, but not *vice versa*. The existence of the laws in question, as we have seen, depends on mechanisms. Thus, in general, one law's existence depends on another law's existence if, and only if, the former law's existence depends on the existence of the mechanisms that sustain the latter law. It is this kind of thought that drives Fodor's intuitions about which laws depend on each other (as well as his method of arguing for the dependence of one law on another). For example, as we have seen, Fodor's very natural idea is that an XYZ law obtains because 1) the mechanism for the water law is in place, and 2) the thinker can't distinguish XYZ from water. Thus, the reason it seems plausible, at first blush, that the XYZ law asymmetrically depends on the water law is that the XYZ law's existence seems to depend on the existence of the mechanisms that sustain the water law. Notice also that questions about laws' mechanisms are more fundamental than questions about whether a law can be eliminated without eliminating another law, or about whether eliminating one law requires more change than eliminating another law.

Precisely what relations between the mechanisms of laws A and B are necessary in order for law B to depend asymmetrically on law A? If no part of law A's mechanism is part of law B's mechanism, law B obviously does not depend on law A. If only part of law A's mechanism is part of law B's mechanism, law B's existence does not depend on law A's existence, but only on the existence of part of law A's mechanism. (That part of law A's mechanism could exist and law B could obtain without law A's obtaining.) Thus, in order for law A's obtaining to be part of what makes law B obtain, all of law A's mechanism has to be part of law B's mechanism. Moreover, in order for the dependence to be asymmetric, at least part of law B's mechanism cannot be part of law A's mechanism. Therefore, in order for law B to depend asymmetrically on law A, law A's mechanism must be a proper part of law B's mechanism.

[39] In other work, I show that both Fodor's original, looser counterfactual formulation and his later one support the conclusion that the star law is not primary. Since the issues about counterfactuals are secondary I omit the details here in order to save space.

I have showed, however, that the slow-comet law's mechanism and the star law's mechanism will have non-overlapping parts – *i.e.*, neither is a proper part of the other. Thus, the slow-comet law's obtaining does not depend on the star law's obtaining, so the slow-comet law does not asymmetrically depend on the star law. It is only by ignoring the part of the star law's mechanism that is out in the world that it can seem plausible that all of the star law's mechanism is a proper part of the slow-comet law's mechanism. That we can eliminate either law without eliminating the other is just a consequence of the more basic fact that neither law's mechanism is a proper part of the other. And it is because of this more basic fact that the slow-comet law does not depend on the star law.

Where did things go wrong? Fodor's argument assumes that if we make stars and slow comets distinguishable, we will eliminate the slow-comet law but not the star law. In order to eliminate the star law, he thinks, it is necessary to change our intentions about the meaning of "star," which will eliminate the slow-comet law as well. But we should have expected something was wrong at this point. Even if it is kosher for Fodor to appeal to our intentions to use "star" with a certain meaning, those intentions plainly cannot constitute the entire mechanism of the star law. There has to be a component that is in the world outside the thinker. In follows that the law can be eliminated not only by changing the meaning of "star" but by interfering with the part of the mechanism that is outside the thinker.[40]

The argument so far has concerned wild laws, such as the slow-comet law, that cover systematically erroneous applications of symbols. It is worth noting that the argument is even stronger for a different kind of wild laws. Fodor emphasizes that robustness includes not just erroneous occurrences of symbols but, for example, occurrences of symbols caused by other thoughts (1990, pp. 80-81; 1991a , p. 261). It might be that *thoughts* about nuclear fusion or about Van Gogh (because of his famous painting of a night sky) reliably cause "star"s. Now if you assumed, as Fodor does, that eliminating the star law requires changing the meaning of "star," then it would seem to follow that eliminating the star law would eliminate the law that, say, thoughts of Van Gogh cause

[40] How could Fodor think that my intentions ensure that the slow comet law asymmetrically depends on the star law? One possibility is that he simply overlooked the possibility of eliminating the star law by making a change in the world. He invariably characterizes the change as making As and Bs distinguishable, without analyzing what that would involve, thus making it possible to overlook the possibility and implications of changing the stars. Another is that he unwittingly slipped into a diachronic interpretation of the relevant changes. That is, if we changed the appearance of stars, our intentions would, over time, produce changed linguistic dispositions so that stars but not slow comets would cause "star"s. See note 38 above.

“star”s. As we have seen, however, changing the stars can eliminate the star law. Changing the stars would not in general (synchronically) affect the various nomic relations between thoughts and the causing of “star” tokens, however. Even if stars were different – even if they had outer crusts that blocked all light – it would still be a law that *thoughts* of nuclear fusion would cause “star”s; the mechanisms that (now) mediate that law do not depend on how the stars are. The explanation is that the thoughts-of-fusion law is mediated by beliefs, for example, the belief that fusion occurs at the core of stars. And, synchronically, those beliefs do not depend on the nature of the stars. Similarly, Van Gogh thoughts would still cause “star”s because (synchronically) my association of Van Gogh with stars is not mediated by actual stars, but by beliefs, memories, and other mental states.

There are at least two ways in which the argument against the star law’s primacy is stronger for wild laws such as the thoughts-of-fusion law and the thoughts-of-Van-Gogh law than for wild laws covering erroneous tokenings of “star.” First, wild laws of the first kind cannot be eliminated except by making changes in the thinker’s mind. The reason is that the mechanism connecting thoughts with tokenings of “star” is entirely mental. A consequence is that there can be no question whether it requires less change in the world to eliminate the star law than to eliminate the relevant wild laws (since no amount of change in the external world will eliminate the relevant wild laws).

Second, and more importantly, there are psychological changes in the thinker of just the sort that Fodor focuses on, that will eliminate the star law without eliminating the thoughts-of-fusion and thoughts-of-Van-Gogh laws. Suppose we change the thinker’s beliefs about how stars look or we change his visual acuity so that he no longer applies “star” to dots of light in the sky. The thinker continues to believe that nuclear fusion takes place in the core of stars, so the thoughts-of-fusion law still obtains. And assuming the thinker associates Van Gogh with stars directly (rather than via mental images of Van Gogh’s painting of the night sky), the thoughts-of-Van-Gogh law still obtains. (Once again, it is irrelevant to synchronic dependence that, as a historical matter, the thinker’s disposition to apply “star” to dots of light in the sky played a role in the thinker’s acquiring the Van Gogh-star association.) Thus, the argument against asymmetric dependence is even stronger for these laws than for wild laws covering erroneous tokenings.[41]

---

[41] I noted above the possibility, for which I argue elsewhere, that slow comets and stars cause “star”s in virtue of being, very roughly, starry-appearing objects (*i.e.*, there is no slow-comet law or star law). That might lead one to think that, on Fodor’s theory, “star” would mean starry-appearing object. For essentially the reasons discussed in the text, however, the thoughts-of-fusion law and the thoughts-of-Van-Gogh law do not asymmetrically depend on the starry-appearing-objects law (nor the converse).

In this section, we have seen that it is false that, in general, laws covering the causing of "star"s depend – let alone asymmetrically depend – on the star law. And, as we have noted, there is nothing special about "star." In general, it is not plausible that the reference law for a mental or linguistic symbol will be the primary law. Given that a thinker means something by a symbol, she will develop dispositions, beliefs, mental associations, and so on that mediate between objects (including thoughts) and occurrences of the symbol. Once those mediating mechanisms are in place, they are not synchronically dependent on the mechanism that mediates the connection between the symbol and objects in its extension. Thus, wild laws do not in general depend on the reference law.

4.6 Conclusion

In chapter 3, we saw that it is intuitively plausible that the reasons for a symbol's uses are what separate the content-determining from the non-content-determining uses. To return to Boghossian's example, suppose someone applies "magpie" to a currawong because the person believes that the Pope is always right and that the Pope has said that there are no magpies around. Intuitively, this disposition to apply "magpie" is based on the wrong kind of reason for it to be among the dispositions that determine the meaning of "magpie." Similarly, my disposition to token my mental symbol "star" in response to thoughts of Van Gogh is based on the wrong kind of reason for it to be among the dispositions that determine the content of "star." These are examples of what we called the "reason-dependence" problem for theories of content.

The asymmetric-dependence theory gains initial plausibility from its apparent potential to address the reason-dependence problem. When Fodor attempts to bring out the intuitive plausibility of his theory, he appeals to the notion that applications of symbols to objects that fall under them (as opposed to mistaken applications and non-application occurrences of symbols) are the most basic uses of symbols, the uses on which all others depend (pp. 96-99). If we find this notion plausible, we will find it plausible that at least roughly the right kind of asymmetric dependences will tend to be established. What makes the notion plausible in the first place is the thought that, for example, the *reason* we apply "horse" to cows on dark nights is that "horse" applies to horses (and cows on dark nights look like horses). In contrast, the reason we apply "horse" to horses does not involve cows on dark nights. Thus, there is at least an initial temptation to think that the distinction between primary laws and other laws might mirror the relevant structure of reasons for uses of symbols. In particular, the uses of symbols that are content-determining might be among the reasons for the uses of symbols that are not content-determining, but not *vice versa*. Accordingly, there is some plausibility to the idea that reference laws will be primary.

There is a serious worry about this putative intuitive motivation for the asymmetric-dependence theory. In assessing the suggestion that non-application uses of symbols depend on (correct) applications, we must be careful not to confuse it with the proposition that non-application uses of symbols depend on the symbol's having the reference that it does. It is uncontroversial that all (relevant) uses of a symbol depend on its having the reference that it does. (In fact, since we are granting for purposes of argument the informationalist's assumption that meaning is reference, the proposition that (meaningful) uses of symbols depend on the symbol's having the reference that it does is just the platitude that (meaningful) uses of symbols depend on the symbols' meanings.) This is not equivalent to the claim that all other uses of symbols depend on applications; in fact, an application of a symbol to an object depends on the symbol's reference as much as any other use: we would not apply the symbol "slab" to a slab if "slab" did not refer to slabs.

In sum, it is uncontroversial that we apply, say, "star" to thoughts of Van Gogh only because "star" means *star*. It is highly controversial – and, in effect, the basic claim of Fodor's theory – that we apply "star" to thoughts of Van Gogh only because "star" has a nomic connection with stars. It is crucial not to slip from the plausibility of the uncontroversial claim to the plausibility of Fodor's theory. Fodor in fact seems to equivocate in just this way (1987, p. 108; 1990, p. 97). But my present purpose is not to argue against the asymmetric-dependence theory or Fodor's attempt to motivate it.

The point is that the initial plausibility of asymmetric dependence stems from its apparent potential to track or mirror the reasons behind uses of symbols. Informational theories, including the asymmetric-dependence theory, cannot appeal directly to the reasons behind a thinker's use of a symbol (though they can try to find properties of symbol-world correlations that track the relevant structure of those reasons). In the next chapter, we turn to a theory of content that explicitly appeals to the reasons behind a thinker’s use of a symbol.

Chapter 5: Peacocke's Possession-Condition Theory of Content

5.1 Introduction

In chapter 3, we built on Boghossian's (1989) classificatory scheme for theories of content. Boghossian first divides the field into reductionist and non-reductionist theories. On the reductionist side, he thinks that dispositional theories are the only serious candidates. He takes dispositional theories to include both informational and inferential-role theories. On the non-reductionist side, Boghossian considers only "quietist" theories, ones with very limited explanatory ambitions. Boghossian produces convincing general reasons, we saw, for rejecting dispositional theories of the sort he considers and for thinking that the quietist theories cannot provide the kind of explanatory account that a solution to the Kripkenstein problem would require.

Let us briefly review Boghossian's objections to dispositional theories. Dispositional theories cannot identify the meaning of a linguistic or mental symbol with the property to which a person is disposed to apply the symbol. They cannot do so for two closely related reasons. First, this proposal would assign intuitively incorrect extensions to our symbols because it would include within a symbol's extension everything to which the thinker was disposed to apply the symbol. Second, it would rule out the possibility of systematic error since it would define the concept's extension, and thus its correct application, in terms of what the thinker is disposed to do. Because of these problems, dispositional theories must try to pick out a select class of meaning-determining dispositions. More precisely, a dispositional theory must identify a property such that a disposition to apply a symbol has the property if and only if the disposition is a disposition to apply the symbol correctly. And it must identify that property of dispositions in non-intentional terms.

There are two main kinds of proposals for picking out the select class of meaning-determining dispositions. First, communitarian proposals identify the meaning-determining dispositions as those that are in accord with community consensus. Second, ideal-conditions-based proposals identify the meaning-determining dispositions as those that the thinker would have under non-intentionally specified ideal conditions. We agreed with Boghossian that communitarian proposals are not serious contenders because, like individual dispositions, community dispositions can be mistaken.

We also agreed with Boghossian's basic argument against ideal-conditions-based dispositional theories. The argument, it will be remembered, is based on the holistic character of belief fixation. What beliefs a person will form depends not only on the person's surroundings and the good functioning of his or her perceptual mechanisms, but also on what other beliefs the person has. So in order for ideal conditions to rule out the possibility of systematic error, they would have to rule out any of the indefinitely many background beliefs that could mislead the thinker. But it is not plausible that non-intentionally specified ideal conditions could accomplish this task. Further, I have suggested that Boghossian's objection to ideal-conditions-based theories might be a symptom of reason-dependence: the relevance to content of a person's symbol-using dispositions depends on the reasons behind the dispositions.

Boghossian does not consider Fodor's or Peacocke's theory (neither was fully developed at the time). Both theories fall outside of Boghossian's categories. First, Fodor's asymmetric-dependence theory is a reductionist theory, but relies on neither ideal conditions nor community consensus. Moreover, that theory is reductionist without being strictly dispositional: it is cast in terms of laws. Second, Peacocke's theory is not fully reductionist since it makes use of content-involving states. We could classify it as non-reductionist, but non-quietist. It is also unlike the dispositional theories Boghossian considers in two other respects. It does not rely on ideal conditions or communitarian consensus, and it goes below conditional behavior (see section 1.4 above) to its causal explanation.

We saw that Fodor's theory might at first blush be thought to have a good chance of responding to the reason-dependence problem. The intuitive idea is that the applications of symbols that depend on misleading background beliefs or mental associations (or other irrelevant reasons) will be less fundamental than the meaning-determining applications. As argued in chapter 4, however, the appropriate asymmetric dependencies do not obtain.

In this chapter, I turn to Christopher Peacocke's possession-condition theory of content. Peacocke assumes that the primary task of a theory of content is to say what determines contents involving only fully grasped concepts. The core of his theory of content thus takes for granted that thinkers fully grasp the concepts that figure in their thoughts (see section 5.3 below). As discussed in section 6.5.4, however, Peacocke adds a proviso allowing that thinkers can, through deference to others, have thoughts involving concepts they incompletely understand. My discussion in this chapter will be restricted to thoughts involving fully grasped concepts, thus treating Peacocke's theory on its own terms. We will consider incomplete understanding in chapter 6. The next section sets out the basic shape of Peacocke's theory, and explains why the theory might be thought to avoid the reason-dependence objection.

5.2 Peacocke's possession-condition theory of concepts

5.2.1 Individuating concepts

Peacocke's possession-condition theory is a theory of concept individuation, of what concepts are. As we will see, he assumes that a theory of content can be derived almost immediately from the theory of concept individuation. Peacocke holds a kind of conceptual-role theory of the nature of concepts: he thinks that each concept is defined by certain canonical patterns of inference, or belief-forming practices more generally (pp. 5-6). [1] More specifically, Peacocke maintains that each concept is individuated by the condition for fully grasping or mastering it, which he calls a "possession condition." The possession condition specifies the concept's canonical or defining transitions.

Unsurprisingly, logical concepts are the ones for which it is easiest to provide immediately plausible examples of canonical inferences. For example, Peacocke's possession condition for the concept of conjunction includes the standard introduction and elimination rules for conjunction (p. 6). To take a different kind of case, Peacocke thinks that among the inferences that are plausibly canonical for a certain sort of natural-kind concept will be the inference from the premise that something falls under the concept to the conclusion "that it is of the same underlying kind as certain samples he has encountered" (p. 26). A canonical transition need not be an inference. It may be a transition involving mental states other than beliefs, such as a transition from a non-conceptual content to a belief. For example, for an observational concept such as the shape concept SQUARE,

[1] In this chapter, all citations to page numbers alone are to Peacocke 1992.

the transitions that Peacocke specifies include a transition from a perceptual state with non-conceptual content to a belief (p. 108).

The underlying idea, though Peacocke does not explain it in this way, is that a concept's constitutive role in attitude psychology is to impose certain requirements or standards on thought – "the concept's requirements," for short. To have *mastery* of a concept is to have an ability to fulfill the requirements it imposes – a capacity to conform one's attitudes involving the concept, and their relations to other aspects of one's mental life, to those requirements. Having this ability does not guarantee that one will fulfill the concept's requirements in making particular judgments. For example, one may have the ability to infer X from Y yet not do so in a particular case because of abnormal conditions (p. 144) or defects of another faculty, such as memory or attention. By specifying what is necessary for a thinker to have mastery of a concept, Peacocke maintains, one simultaneously specifies what a concept fundamentally is. In order to give an account of content, one could then go on to give an account of how a thought has to be linked to a concept's possession condition in order for the thought to involve that concept.[2]

5.2.2 Non-reductionism and circularity

Since Peacocke's possession-condition theory of concepts is the engine of his theory of content, we need to examine the former theory. As mentioned above, a distinctive feature of the possession-condition theory, and thus of the theory of content it yields, is that it is not fully reductionist. One way in which Peacocke's possession conditions are not reductionist is this: since a possession condition specifies a concept's canonical transitions, it makes reference to mental states, including mental states with conceptual content.

In order to avoid circularity, Peacocke imposes two restrictions on possession conditions. First, the concept that is the subject of the possession condition may not be mentioned as such within the scope of a thinker's propositional attitudes. For in that case, Peacocke maintains, the possession condition would presuppose possession of the concept (p. 9).[3] To avoid mentioning the concept

---

[2] As mentioned above, Peacocke's primary account of content relies on the assumption that thinkers satisfy the possession conditions for the concepts that figure in their thoughts (see section 5.3 below). For his treatment of deference, see 6.5.4 below.

[3] This position might, at first blush, seem inconsistent with Peacocke's official view that a thinker can have a thought involving a concept without possessing a concept. But since Peacocke assumes that an account of what it is to have a thought involving a concept is parasitic on an account of possession of the concept, it is arguably reasonable for him to hold that an account of possession of a concept cannot mention thoughts involving the concept. As I discuss below (see section 6.5.4), despite an official disclaimer, Peacocke often

within the scope of the thinker's attitudes, Peacocke uses a variable over concepts: a possession condition provides that the concept *C* is that concept *X* to possess which a thinker must meet a particular condition, where "X" is a variable over concepts. In the specification of the possession condition, "X" may appear within the scope of propositional attitudes.

Second, if a possession condition for a particular concept *C* presupposes that the thinker has another concept *D*, the possession condition for *D* must not presuppose that the thinker has *C*. And there may not be finite circles of this kind (pp. 12, 100-101). Thus, a concept may sit at the top of a hierarchy of concepts such that possession of the concepts at the bottom level does not presuppose possession of any other concepts. It is not plausible, however, and Peacocke does not claim that every concept is part of such a hierarchy. Rather, Peacocke holds that there are families of concepts that form "local holisms" (pp. 10, 12). His strategy with respect to such a family of concepts is to attempt to provide one joint possession condition for all concepts in the family.

5.2.3 Advances over dispositional theories

Peacocke's possession conditions have several other noteworthy features. I will first elaborate them, and then (in section 5.2.4) comment on ways in which they might seem to help with Boghossian's belief holism problem and with the intuitive problem of reason dependence.

First, a possession condition does not require that a thinker actually be disposed to make the canonical transitions. Rather, it specifies that the thinker must find those transitions "primitively compelling." The notion of primitive compulsion can be seen as an attempt to develop and make more precise Wittgenstein's notion of finding certain steps or ways of going on "natural" (p. 13). Peacocke does not elaborate on the intuitive idea of what it is to find a transition compelling. A transition is *primitively* compelling if and only if the thinker finds it compelling and does not find it compelling because he has inferred it from something else (p. 6).[4] For ease of writing, I will often write simply "compelling," rather than "primitively compelling."

---

relies on the assumption that a thinker cannot have a thought involving a concept without possessing that concept.

[4] There is a third component to Peacocke's definition: "for possession of the concept *C* in question . . . he does not need to take the correctness of the transitions as answerable to anything else" (p. 6; see also p. 135). This is an odd aspect of the definition because it concerns not the relation of the thinker to the transition, but the role of the transition in individuating a particular concept. It has the consequence that a transition is primitively compelling only if it plays a certain role in a condition for mastering a concept. It thus seems to relativize the notion of primitive compulsion to a given concept. That is, it raises the possibility that a transition might be primitively compelling with respect to one concept – because that transition needn't be taken as answerable to

Second, a possession condition requires not only that a thinker find a canonical transition compelling, but that the thinker does so *in virtue of the form* of the transition. This requirement needs some explication. As Peacocke uses the term, to say that a thinker does so "in virtue of" the form of the transition is to say at least that the thinker's doing so is *causally explained* by the form of the transition (pp. 135, 137-138, 189, 195, 197). Also, the notion of the "form" of the transition is not meant to be purely logical or syntactic form. Rather, the form of a transition is what remains of a transition-type after expressions for the concept (or concepts) in question – the one for which the transition is canonical – have been replaced by variables. The form of a typical transition will be defined in terms of conceptual contents. In the example of the natural-kind concept given above, the transition was from a judgment that something falls under the concept to a judgment that it is of the same underlying kind as certain other encountered objects. Thus, part of the "form" of this transition involves the concept of an underlying kind. When I say that form is not purely logical or syntactic, I mean that what makes it the case that an instance is an instance of a particular form may depend on the content of the instance, not just on its logical structure or syntactic properties. "Type" might have been a better term than "form" for this reason.[5]

The general point is that the thinker's finding instances of the type of transition compelling must be causally explained by the appropriate feature of those instances – the feature that makes them instances of the relevant type of transition. This rules out both that the thinker's finding instances of the transition compelling is explained by anything peculiar to the particular instances and that it is explained by those instances being instances of a "bent" form.

Third, Peacocke does not attempt to identify a class of dispositions such that being a manifestation of a disposition in the class is necessary and sufficient for being a correct application of the concept. He does not directly appeal to dispositions to determine correctness. Rather, a concept's possession condition, with the world, determines the concept's reference, or, more generally, its semantic value. Roughly, a concept's reference is that property (or relation, object, or function) that would make the concept's canonical inferences truth preserving, or that would make beliefs formed in accordance with the specified belief-formation practices true (pp. 17-19). Thus,

anything else for possession of that concept – but not primitively compelling with respect to another concept.

[5] The use of the term "form" is less apt when a transition is not an inference, but a transition from a non-conceptual state to a belief. For example, Peacocke's possession condition for the concept RED requires that the thinker "be disposed to form the belief [roughly that that object is red] *for the reason that* the object is so presented" (p. 7, emphasis added). In this example, the object's being *so presented* is the analogue of an

Peacocke's account has the following structure. For any given application of a concept, Peacocke needs to link the application back to the person's satisfaction of a possession condition. (In section 5.5, I address the crucial question of how Peacocke links a particular use of a concept back to a possession condition – the question of what makes it the case that a person's thought involves one concept rather than another.) The possession condition, along with the world, then determines the concept's semantic value, which determines whether the application was correct. The important point for now is that Peacocke manages to avoid the problem of saying how dispositions can themselves determine the standard of correctness.

5.2.4 Peacocke's theory and the problems of belief holism and reason dependence

Peacocke does not need there to be ideal conditions such that a thinker, under those conditions, will be disposed to apply a concept correctly. As explained, he does not appeal to the thinker's actual dispositions to determine the concept's extension or reference. And he makes no appeal to ideal conditions.

Peacocke does need it to be the case, however, that a thinker is compelled to make a concept's canonical inferences. Requiring that thinkers find the relevant transitions compelling rather than that they be disposed to make the transitions may go some way toward solving the kind of belief-holism problem that Boghossian identified in the case of ideal-conditions-based theories. First, the mere fact that a thinker is disposed to apply "magpie" to currawongs does not imply that the thinker finds the transition from the perceptual state to the judgment primitively compelling. In fact, if the thinker is so disposed because of background beliefs, he will not be *primitively* compelled to make the transition (since being primitively compelled rules out the possibility that the thinker finds the transitions compelling because of background beliefs). Second, a thinker may find a judgment compelling, but not be disposed to make it. There may be interfering factors, for example. It is not clear whether background beliefs could count as interfering factors. That is, Peacocke does not make clear whether his notion of "compelling" is such that a thinker can count as continuing to find a transition compelling, if other reasons convince him that, all things considered, the transition is not correct.[6] If this were possible, a thinker might find the transition from the perception of the

---

inference's having a particular form. I am grateful to Martin Davies for this point.

[6] At one point, Peacocke seems to suggest that if a person rejects a transition, *a fortiori* the person doesn't find it primitively compelling (pp. 136-137).

currawong to the judgment that it is a currawong primitively compelling, but other beliefs might prevent the thinker from being disposed to make the transition to the judgment that it is a currawong.

It also seems plausible that Peacocke's theory is in a much better position than strictly dispositional theories to deal with the reason-dependence issue. First, transitions that the thinker finds primitively compelling are transitions that she finds compelling not because of inference from other beliefs. Thus, requiring that a thinker find a transition primitively compelling is a way of pinpointing transitions that are based on the right kinds of reasons. Second, the requirement that the thinker find the canonical transitions primitively compelling because of their form is another way of focusing on transitions made for the right reasons. Third, since possession conditions make use of content-involving states, Peacocke can state explicitly the content-involving reasons for making a canonical judgment. Thus, at least *prima facie*, Peacocke's theory has more resources than purely dispositional theories for addressing reason dependence.

5.3 The causal explanation solution to the Kripkenstein problem

Peacocke takes his possession-condition theory to provide an answer to the Kripkenstein question of what determines that a particular concept, as opposed to others, is involved in the content of a thinker's attitude (pp. 134-146, 190-197). The first point is that a theory of what is required to possess a concept is not yet a theory of what makes it the case that a particular thought involves that concept. I am not here raising the issue that thinkers can have thoughts involving concepts they do not fully grasp. Rather, the point is simply that a thinker will simultaneously possess many concepts. A theory of content, even one limited to cases of fully grasped contents, needs to say what determines which of those concepts figures in a particular attitude. For example, I may possess both the concept of addition and some variant concept of quaddition. My possession of those concepts – my satisfaction of both their possession conditions – plainly doesn't settle the question of which concept figures in a given thought on a given occasion.

Peacocke's proposed answer is none too explicit (pp. 107-111, 134-145; 177-197). It rests on the possession-condition requirement that a thinker not only be disposed to make certain judgments in circumstances defined in terms of content-involving states of the thinker, but that "the relevant judgements [be] made *because* the subject is in these other content-involving states" (p. 189). Here is Peacocke's most explicit statement of his view: "what makes it the case that someone is employing one concept rather than another is not constituted by his impressions of whether he is. Rather, it is constituted by complex facts about the explanation of certain of his primitively compelling judgements" (p. 195). Thus, the view seems to be that a given thought involves a particular concept because 1) the thinker finds compelling certain transitions to and from the thought – those specified by the possession conditions of the concept and 2) the causal explanation of the thinker's finding these judgments compelling is the one specified in the concept's possession condition.[7]

For example, my belief that Socrates is human and I am mortal involves the concept of conjunction not only in virtue of my finding certain inferences (the canonical ones for the concept) to and from that belief compelling, but also in virtue of my finding them compelling *because* they instantiate the appropriate form. A particular inference, say the inference from *Socrates is human and I am mortal* to *I am mortal* instantiates many forms in addition to the form *P C Q*/ *Q*. For example, it instantiates the form *R*/ *I am mortal*. On Peacocke's theory, I must find the relevant inference compelling because it is an instance of the form *P C Q*/ *Q*, rather than because it is an instance of some other form. In sum, what determines that a thinker is using a particular concept is that the thinker's willingness to make inferences of the form specified by the concept's possession condition is causally explained by the inferences' having that form.

[7] The transitions involving a given thought that a thinker must find primitively compelling in order for the thought to have a certain propositional content must be determined by combining the transitions specified in the different clauses of the possession conditions of the component concepts of the propositional content. Peacocke outlines a procedure for "multiplying out" the clauses of the possession conditions of the constituent concepts to obtain the requirements for a state to be a belief with that content (pp. 108-111).

Consider Kripke's (1982) example of the concept *plus*. Peacocke's proposed possession condition for *plus* specifies the following form of inference, which I call the *plus-form* (pp. 134-138):[8]

$m\ C\ k = n$

----

m C (the successor of k) = the successor of n

Thus, the plus-form is a property of a transition from one mental state to another that is defined in terms of the *contents* of those mental states. Take QUUS to be the concept with the same possession condition as PLUS except that the *quus-form* differs from the plus-form in the following way: when k is larger than 10^{70}, the value on the right side of the "=" sign in the conclusion of the inference is the successor of the successor of n, rather than the successor of n. Consider a thinker, T, who has never confronted the question of what inference to make from $m + k = n$ in a case where k is larger than 10^{70}. She regularly has made and makes inferences of the plus-form. Of course, all those inferences also have the quus-form since, for all cases she has encountered, any instance of either form is also an instance of the other form.

Although Peacocke does not spell it out, according to his account, what is supposed to determine that T's attitudes involve PLUS rather than QUUS is that her disposition to make inferences of the plus-form is causally explained by their having the plus-form – and not, for example, by their having the quus-form. The account requires a notion of causal explanation on which it is possible that, even if all the relevant inferences that a person is disposed to make – even all she could make given her psychological make-up – are instances of both the plus-form and the quus-form, she is disposed to make the inferences *because* they have the plus-form, not because they have the quus-form (or *vice versa*). Peacocke does not develop the critical notion that the form of a transition can causally explain a thinker's disposition to make the transition, where the form of the transition between mental states is individuated in terms of the contents of the mental states, though he does make clear that he thinks causal explanation is not normative or interest-relative (pp. 138, 184). After suggesting that a response to Kripke's skeptic should rely on the notion that the form of the transitions can causally explanatory, Peacocke immediately refers the reader to a section of a

[8] The same is required for the principle $m\ C\ 0 = m$, but that principle can be ignored for present purposes. In these forms m and n are variables ranging over natural numbers, and C is a concept variable.

later chapter "[o]n the question of how the form can be explanatory" (p. 145).

That section (7.1), however, is devoted to the question of how a form could causally explain anything, in the sense of how a form's causally explaining something could be *implemented*. The section discusses a psychological explanation of a thinker's satisfying a possession condition – specifically of the fact that the thinker is disposed to make the specified transitions between mental states.[9] Peacocke emphasizes that an explanation of a thinker's satisfying the possession condition must not mention states individuated in content-involving terms: the explanation must "cut below the level of states with reason-giving characteristics" since those characteristics are among what is to be explained (pp. 179-180).

In this context, as a response to the problem that "it is far from clear" how the "form of a transition between contents . . . can causally influence a thinker," Peacocke gives an example of a hypothetical explanation (pp. 183-184). The example is a subpersonal psychological theory that accepts the language of thought (LOT) hypothesis, holding that there are mental representations of concepts that can be combined to form mental representations of complex contents. The theory would also give a subpersonal account of the property that realizes a thinker's finding something primitively compelling (pp. 184-185). (Call this "the realizing property.") The theory could then account for a thinker's finding inferences primitively compelling because of their form in the following way:

> Such a theory can postulate a mechanism ensuring that the mental representation of an inference is assigned the subpersonal property that realizes the property of being found primitively compelling if the mental representations of premises and conclusion are themselves of certain forms. Since the latter forms are forms of token representations, rather than forms of abstract objects or something in Frege's third realm, there is no principled problem of how they can causally influence a thinker's mental states (1992, p. 184).

It seems clear that Peacocke is here trying to explain how his story about a form's causally explaining an inference could be implemented. That would explain, for example, why he makes the

[9] Strictly speaking, that a thinker's willingness to make the specified inferences has the right explanation is itself a possession-condition requirement, and therefore is one of the things to be explained by a psychological account of why a thinker satisfies the possession condition. So the explanation of a thinker's having the required dispositions is not an explanation of why the thinker satisfies the whole possession condition, but only a part of it. But since

quoted remark that the explanation must cut below the level of states with reason-giving characteristics. On the face of it, however, the question of what makes it the case that one form, rather than, say, a "bent" one, is causally influential should be distinguished from questions about implementation. The former question is the one that goes to the heart of Peacocke's theory of content. And, as we have noted, Peacocke's general approach is to give a non-reductionist account, one that does not cut below the level of states with reason-giving characteristics. Unfortunately, Peacocke does not clearly distinguish the question of what makes it the case that one form, rather than another, is causally influential from the question of how a form's being causally influential could be implemented. Moreover, he sometimes seems to suggest that the implementation account is an essential part of the account of what makes it the case that a thought involves one concept rather than another.

Peacocke provides no further account of the notion of an inference's being found primitively compelling because of its form. But let us distinguish two ways in which we could develop the notion, corresponding to two levels of explanation – personal and subpersonal. In particular, the two ways of developing the notion take the form of two ways in which the LOT example could be related to Peacocke's solution to Kripke's problem. First, it may be that the example plays no constitutive part in the solution. According to this line of thought, Peacocke's solution is wholly at the personal level: T's attitudes involve PLUS rather than QUUS because her disposition to make the right inferences is explained by their having the plus-form. Further, this account holds that no more needs to be said to solve Kripke's problem: the idea of the plus-form's causally explaining the disposition is one that we understand and can see to be naturalistic without need of an account at a lower level. The example of the subpersonal explanation addresses an independent question, though the existence of a subpersonal mechanism of a certain kind may be necessary for the truth of the personal-level causal explanation. Call this first version of Peacocke's solution "the personal-level solution."[10]

According to the second account of the notion of an inferential form's being causally explanatory, the hypothesized subpersonal explanation of how a form can causally influence a thinker's mental states is a constitutive part of the solution to Kripke's problem. On this understanding, for a form causally to explain a thinker's willingness to make certain inferences, that

Peacocke does not note the qualification, I put it aside.

[10] Peacocke has informed me in conversation that he intends the personal-level solution.

willingness must have the right subpersonal explanation, where the right subpersonal explanation is one that makes the form in question, rather than some other form, the causally explanatory one.[11] Thus, if the example is right, a certain (content-involving) inferential form, F, causally explains the thinker's finding the inference primitively compelling just in case the subpersonal property that is the causally explanatory one is the syntactic representation of F in the language of thought. Call this version of Peacocke's solution "the subpersonal-level solution." I will consider the two versions in turn, beginning with the personal-level one.

5.4 The personal-level solution

On the personal-level way of developing the account, Peacocke does not attempt to go down to the syntactic level but claims that the appeal to a causally explanatory property can discriminate between the plus-form and the quus-form. He simply insists that our understanding of the notion of a causally explanatory property includes an understanding of how a thinker's willingness to make a transition can be explained by its being of a particular content-involving form (*cf*. pp. 138, 146).

The idea that the reason T means PLUS is that her willingness to make certain inferences is explained by their having the plus-form has strong intuitive appeal. It avoids many of the problems of a bare dispositional account, bringing us closer than such an account can to what we intuitively think it is to be using a given concept.[12] Even if Peacocke's account is right as far as it goes, however, does it go far enough to solve Kripke's problem?

To say that T's willingness to make certain inferences is causally explained by their having the plus-form is to say that T's willingness is causally explained by certain mental transitions' being instances of one content-involving form or type rather than instances of another. Although the transitions are in fact instances of both forms or types, T's finding the transitions primitively

---

[11] Martin Davies (1991, p. 246) proposes such a philosophical elaboration of the idea of an inference's being found primitively compelling because of its form:

> At least part of what is involved in the idea of particular inferences being found primitively compelling by virtue of their form is the following requirement. Mirroring the commonality in the inferences that are found primitively compelling – namely, their form – there should be a commonality in the causal processes that explain their being found so.

Indeed, Davies argues that the Peacockean account of concept mastery works only if the LOT hypothesis is true. See especially pp. 245-247.

[12] We understand the idea that something can have the causal explanation that corresponds to following a given rule, though, in practice, its dispositions do not always respect the rule because of failure of background conditions. There are difficult questions about how precisely to develop the distinction between causally explanatory factors and background conditions and about whether it can be developed in a way that is not interest relative.

compelling is causally explained by their being instances of the plus-form, and not by their being instances of the quus-form.

Remember that the plus-form is a property of a transition from one mental state to another that is defined in terms of the contents of those mental states. So, according to Peacocke, a thinker means PLUS because her dispositions are explained by one content-involving property – the plus-form – rather than another – the quus-form. This claim is not trivially circular because the plus-form does not involve the concept PLUS. The question, rather, is whether we have enough understanding of what it is for one content-involving property rather than another to be the causally explanatory one (given that the transition in question has both properties). Or, to put it another way, whether our understanding of the notion of a content-involving property's being causally explanatory is independent enough of our understanding of the notion of a state's having a particular content for the former to be used in a fully satisfactory explanation of the latter.

Peacocke does not develop the notion of a content-involving property's being causally explanatory. As I noted earlier, Peacocke's LOT example is his only attempt to say anything about what it could be for one content-involving form or type rather than another to be causally influential. And we are now considering the reading of Peacocke's argument according to which any subpersonal story, even if it is, for example, nomologically necessary for the truth of the personal-level account, is not a part of Peacocke's solution to the Kripkenstein problem. So we are in effect asked to take as primitive the idea of a form's being causally influential.

We should set aside two familiar ways of understanding the idea that a certain content-involving property is causally explanatory. The first takes for granted ordinary intentional explanation. We can explain my producing the answer 12 on the ground that I believe that 7 plus 5 is 12, or my continuing a series in a given way on the ground that I am consciously following a particular rule. Peacocke is obviously not relying on such an understanding of a form's being causally explanatory. It would simply beg the question to say that a thinker's willingness to make a certain inference is explained by his following the rule for addition or his intention to add. Peacocke is trying to give a deeper kind of explanation in terms of the transition's having the plus-form. It is not that the thinker intends to make inferences with the plus-form and believes that this inference has the plus-form. As Peacocke makes clear, the plus-form is supposed to explain a thinker's willingness to make inferences without the thinker's explicitly knowing the plus-form (pp. 145, 197).

Second, one could understand the idea that a certain content-involving property is causally explanatory as the idea that some lower-level property that realizes the appropriate content-involving property is the causally explanatory one. This way of understanding Peacocke's account will be considered in the next section. The present discussion concerns a way of understanding Peacocke's account that remains at the personal level.

We are left with a primitive notion of a content-involving form's being causally explanatory, where that notion does not correspond to either of the familiar notions in the area. We thus have little to go on in evaluating the personal-level version of Peacocke's account.

Consider a person who is primitively compelled to make certain inferences. The inferences all instantiate two different forms that diverge only for numbers that are larger than any the person is psychologically capable of considering. Peacocke needs it to be the case that there can be a fact of the matter as to which of the two inference forms is the causally influential one – the one that explains the thinker's being primitively compelled to make the inferences. He gives no argument for this proposition, but takes it for granted.

It is important that the worry is not some kind of general skepticism about causal explanation. Let us waive any such skepticism, for such worries are not special to the theory of content. For example, let us waive Goodmanesque problems, unrelated to mental or linguistic content, about whether "straight" or "bent" properties are the causally influential ones. In general, we can take for granted realism about non-content-involving causal explanations.

These waivers, however, seem to stop short of the proposition that an appeal to causal explanation will distinguish between content-involving properties, such as inference forms that diverge only for cases the thinker is incapable of considering (and will distinguish between them in a way that yields appropriate results). If in addition to waiving any general skepticism about causal explanation, we also assume that an appeal to causal explanation yields intuitively correct results with respect to content-involving explanations of thinkers' willingness to make mental transitions, it is arguable that we simply beg the Kripkenstein question. Taking for granted that a transition's having the plus-form, rather than its having the quus-form, is what is causally influential is very close to taking for granted that the thinker is following the rule for PLUS rather than the rule for QUUS. At any rate, before we could be satisfied that an appeal to causal explanation solves the problem of content, we would to need to know more about *what it is for* a particular content-involving property to be the causally influential one. As things stand, it is difficult to evaluate whether realism about causal explanation implies that one of the forms and not the other can be causally explanatory when, for example, the forms diverge only for cases the thinker has not considered or is not capable of considering.

It is also not clear that the causally influential content-involving property, if there is a unique one, will be the one that yields intuitively correct contents. For example, in order for Peacocke's account to yield the intuitively correct result, it must be that the thinker's finding certain inferences compelling is explained by their having the plus-form. One problem is that it is plausible that many thinkers have learned, for instance, some simple algorithm that enables them to produce correct results for many simple addition problems, but have not acquired the appropriate sensitivity to the plus-form. Pre-theoretically at least, we would attribute some such thinkers thoughts involving the concept of addition – they think, for example, that two plus two equals four.

I address such cases in chapter 6, under the heading of incomplete understanding. It is important to note, however, that although Peacocke's possession condition for the concept of addition implies that such thinkers lack mastery of the concept, it is not the case that all such thinkers are naturally classified in that way. Many ordinary people who seem to be perfectly competent users of the concept may not have the appropriate sensitivity to the plus-form, and the point is even more clear in the case of more complex concepts. In general, for a wide range of concepts with which we are ordinarily competent, the causal explanations of our finding certain canonical transitions compelling may regularly fall short of the ideal explanations specified in Peacocke-style possession conditions. We may be skeptical that the contents of our thoughts are hostage to the precise details of such causal explanations.

A different problem is to avoid the result that the causally influential property of inferences is not having the plus-form, but having the plus-form (or perhaps any form equivalent to it for numbers less than w) *and* involving numbers smaller than or equal to w (where the thinker is incapable of considering numbers larger than w). In other words, an account of causal explanation will have to face the question of whether background conditions are to be distinguished from causally influential factors and, if so, how. The need to address this question is not a problem that is peculiar to the theory of content, so it is not itself an objection to Peacocke's appeal to causal explanation. The objection, rather, is that it is not clear that a general account of causal explanation will draw a distinction between causally influential factors and background conditions in a way that has the consequences that a theory of content requires. For example, it may well be that in general there is no interest-independent distinction between background conditions and causally influential factors. The full explanation of why a rock slid down a hill may well include that the temperature was within a certain range, and indefinitely many other such factors (though which factors it is appropriate to cite in a particular context depends on the interests that are at stake).[13] In response to such suggestions, a theorist attempting to develop a Peacocke-style account of content has a variety of options. The present point is that we have another reason that more is needed before an appeal to causal explanation can yield an adequate account of content.

Peacocke poses the problem of whether his possession condition would yield "as the reference of the concept PLUS the classical addition function defined on all natural numbers [or] . . . merely a function defined only on accessible ('surveyable') natural numbers" (p. 137). He states that his possession condition would yield the intuitively correct reference "only if we can give philosophically adequate reasons for saying, of two inference forms that disagree only on (concepts of) the inaccessible numbers, that one rather than the other is causally influential in the accessible cases" (p. 137). But he does not attempt to provide such reasons.

Peacocke thus recognizes that he needs to do more to address the case of forms that differ for inaccessible numbers. Even in the case of two forms that differ for accessible natural numbers (that have not yet been encountered), however, we may be unconvinced that there are adequate reasons for saying that one form rather than the other is the causally influential one. Again, this is not because of any lack of realism about causal explanation. Without an account of what it is for a particular

[13] Biological explanations may be special because a notion of biological function may give rise to a distinction between background conditions and causally influential factors. Some theorists have explored the idea that the notion of biological function can solve the content problem, but that is not Peacocke's approach.

content-involving property to be the causally influential one, assuming that an appeal to that notion solves the Kripkenstein problem is too close to assuming the problem away.

Peacocke himself seems to indicate doubts about the adequacy of the personal-level solution.[14] In response to Kripkean skepticism, Peacocke points out: “Kripke’s skeptic does not consider the response that it is the *form* of the transitions and principles relied on in counting that explains their compellingness and to which the antiskeptic should appeal.” He continues: “I have been emphasizing that what explains why certain transitions and principles are found compelling is a form that extends beyond instances that have been considered” (p. 145). Immediately thereafter, however, Peacocke writes that he has “yet to deal with the idea that when someone possesses a concept, there are facts about him that determine whether it would be correct for him to apply it to an as yet unencountered object, facts given in advance of any judgement he actually makes about the new case” (pp. 145-146). This is tantamount to saying that he has not yet given his solution to the problem of content. An account of what makes it the case that a thinker is using a particular concept is an account of facts about him that determine which concept he is using. And whatever determines which concept a thinker is using *ipso facto* determines what it is correct for him to apply it to. (For a discussion of this critical issue, Peacocke again refers the reader to chapter 7, where he gives the subpersonal account of how the form of a transition can be causally explanatory.)[15]

In sum, it is unclear whether the personal-level account can provide a satisfactory answer to the content question. At very least, what it is for one content-involving property rather than another to be causally influential – as opposed to how such a property’s causal influence could be implemented – requires further development.

5.5 The subpersonal-level solution

Let us therefore turn to the second explication of the notion of an inferential form’s being causally explanatory, according to which the subpersonal story is part of the solution to Kripke’s problem. The theory postulated in the example presumes the correctness of positions such as the LOT

---

[14] As noted above, however, Peacocke says that he intends the personal-level solution.

[15] Also, as noted above, before giving his hypothetical subpersonal explanation, Peacocke maintains that (without that explanation) “it is far from clear how the form of [a transition between contents] can causally influence a thinker” (pp. 183-184). Finally, Peacocke argues (p. 189):

> The relational properties in virtue of which a state has its content are not normally causally efficacious in producing the judgements mentioned in the possession condition. So again the question arises of how the content-involving states can be explanatory of the judgements mentioned in the possession condition. . . . [W]e can give an answer in terms of subpersonal representational properties that are locally causally

hypothesis, which are controversial (p. 184). Since the example is the only attempt that Peacocke makes to explain how the form of an inference could explain a thinker's disposition to make it, I will assume the correctness of the underlying positions and argue that, even so, the appeal to the forms of token mental representations fails to address Kripke's problem.

We can grant that, since a token representation is a physical thing, there is in principle no problem of how T's mental representations' having a particular form could causally influence T. By hypothesis, all of T's past mental representations that have been instances of the plus-form have also been instances of the quus-form. Thus, whenever she has been causally influenced by a representation of an inference with the plus-form, she has equally been causally influenced by a representation with the quus-form. The LOT example shows how the form of a representation could causally influence a thinker's dispositions. But that does not yet show how appeal to causal explanation can distinguish T's attitudes' involving PLUS from their involving QUUS. We cannot assume the assignment to the relevant mental representations of contents involving PLUS rather than QUUS. That would utterly beg the question and make appeal to causal *explanation* unnecessary. Similarly, it would beg the question to assume attribution of the plus-form and not the quus-form, to transitions between token mental representations with contents within the range for which the plus-form and quus-form are indistinguishable. So the form of token representations can causally influence T's judgments, but that does not determine whether T's attitudes involve PLUS rather than QUUS.

This is where *explanation* is supposed to provide a solution. Let us grant Peacocke a robust notion of causal explanation such that there is a fact of the matter as to which physical or syntactic property of token representations explains T's finding the relevant inferences primitively compelling. Further, let us again waive any Goodmanesque problem about which such property – *e.g.*, a "straight" or "bent" one – is the causally explanatory one, for that is not a type of problem that is peculiar to content. So a particular property P of representations of inferences, characterized in physical or, at least, non-content-involving terms, causally explains T's relevant inferential dispositions.

Peacocke makes no attempt to say what determines that the syntactic property P corresponds

efficacious.

to the plus-form, rather than the quus-form. (He comments only that the example "presumes, of course, the correctness of several philosophical accounts underwriting its assignment of particular contents to mental representations" (p. 184).)

A token representation that has P would be a token of *both* the plus-form and the quus-form in every case where k is less than or equal to 10^{70}. Whether the syntactic property P corresponds to the plus-form or the quus-form depends on how those forms are represented in LOT. The plus- and quus- forms were specified in terms of concepts of succession, identity, and natural numbers,[16] so whether P is the LOT representation of the plus-form depends on the assignment of such concepts to LOT representations. But we cannot take for granted such assignments of content to syntactic representations, for Kripke's problem arises again for them: what determines that a given LOT symbol means succession, rather than, say, quccession? We have simply pushed the problem down a level, and nothing in Peacocke's account offers any hope of solving the problem at this level. (Not that Peacocke claims any such thing; to be clear, Peacocke does not suggest that the problem can be solved by assigning contents to subpersonal representations.)

Pushing Kripke's problem down to the level of the content of mental representations is no advance in solving the problem. In the end, the evident futility of the subpersonal way of developing Peacocke's theory supports the interpretation of the theory according to which the LOT example plays no constitutive part, and the theory of content remains entirely at the personal level.

5.6 Conclusion

In this chapter, we have considered Peacocke's theory of content on the assumption that thinkers who have thoughts involving a concept have mastery of the relevant concepts. We have seen that, even given that strong assumption, it is at best unclear that Peacocke's theory is successful. The theory depends crucially on the notion that the (content-involving) form of an inference can explain the thinker's finding the inference primitively compelling. But Peacocke does not develop that notion. As things stands, there is a serious worry that Peacocke's appeal to causal explanation simply begs the question that it is designed to answer. In the next chapter, it will be clear that, whatever the verdict on Peacocke's theory while the assumption of concept mastery remains in place, Peacocke's

[16] The concepts may be any that have as their semantic values the successor function, the identity function, and natural numbers, but for convenience I will write "the concept of succession," and so on.

theory cannot handle cases of incomplete understanding.

Part III: Incomplete Understanding

Chapter 6: Incomplete Understanding

6.1 Introduction

In the last three chapters, I have argued that standard theories of content are inadequate *on their own terms* – that is, even given their own conception of what such a theory needs to accomplish. In particular, standard theories attempt to explain what determines the content of thought for the case of thinkers who fully grasp the relevant concepts. My critique of those theories has largely not challenged the assumption that a theory of content can take concept mastery for granted.

Most theorists accept, as I have mentioned, that thinkers can have thoughts involving concepts they do not fully grasp. Theorists tend to assume that such thoughts can be accounted for by an appeal to deference. Thus, they give an account of content for the case of a thinker who fully grasps the relevant concepts, and then throw in a proviso that a thinker who incompletely grasps a concept can have thoughts involving it by deferring to others who fully grasp the concept.

In this chapter, I argue that such appeals to deference are unsatisfactory. A theory that relies on deference to account for cases of incomplete understanding needs to explain why, given the theory's account of what it is to have a thought involving a given concept, deference could enable a thinker to have a thought involving a concept she incompletely grasps. Standard theories of content fail to do so.

The problem for standard theories of content that is raised by deference-dependent contents goes beyond the problem of finding a way to account for such cases. Once it is accepted that there are cases in which thinkers, by deference to others, manage to have thoughts involving concepts they incompletely grasp, we need an account of what it is to have a thought with a particular content such that deference is one way to do so. If standard theories of content cannot make sense of the role that they assign to deference, that counts against their general view of what it is to have a thought involving a given concept. Thus, the argument in this chapter reinforces the argument that standard theories of content fail on their own terms.

In section 6.2, I begin by exploring why standard accounts of content assume the demanding view (DV) – the view that in order for a thinker to have a thought involving a particular concept, the thinker must fully grasp the concept or defer to others who do. In section 6.3, I explain that, perhaps contrary to first appearances, the well-known work of Putnam and Kripke on natural-kind concepts does not reject DV. In section 6.4, I show that standard theories of content are committed to DV because they take the achievement view for granted.

In the following five sections, I challenge the assumption that inferential-role and reference theories of concept individuation need be linked to the achievement view. First (section 6.5), I show that the achievement view runs into serious difficulties in trying to account for claims about the role of deference that are widely accepted (even by achievement theorists). Second (section 6.6), one of the most influential criticisms of inferential-role theories of concept individuation rests on a failure to separate an inferential-role theory of concept individuation from the achievement view. As a view about the individuation of concepts, the inferential-role view is much more plausible without the achievement view than with it, and more plausible than a reference-based view of concept individuation. Third (sections 6.7-6.8), once we carefully separate the question of what it is to *have[1] a concept from the question of what individuates a concept, the achievement view is not a very plausible or satisfying answer to the former question because it does not account for normative features that are (in part) constitutive of *having a concept. Also, the achievement view seems less obvious or natural once we have seen that the relation between, on the one hand, dispositions and abilities, and, on the other, *having a particular concept need not be type-type identity, for different dispositions and abilities might realize *having a particular concept in virtue of their sharing some higher-order property. Finally (section 6.9), I argue that ordinary practices of attitude attribution accept that there are Independent Incompletely Understood Contents (IIUCs)[2], which are incompatible with the achievement view.

6.2 The demanding view

Standard accounts of content assume DV, though this may not be obvious at first glance. It will be useful to begin by examining how different types of accounts of content come to hold the assumption and the form that the assumption takes in those accounts. We will have to look separately at, on the one hand, broadly Fregean accounts of content – accounts that individuate content by cognitive significance

[1] As explained in section 1.9 above, for a thinker to *have a concept is for the thinker to have (or to be able to have) thoughts involving the concept.
[2] As defined in section 1.5, IIUCs are incompletely understood contents that do not depend on deference.

– and, on the other, accounts that do not accept the sense/reference distinction and individuate content in terms of reference.

6.2.1 Inferential-role theories

Let us begin with inferential-role accounts and trace the source of their assumption that thinkers fully understand the concepts they use. Inferential-role accounts begin from the idea of content as cognitive significance. If a concept is the unit of cognitive significance, it is plausible to individuate each concept by its inferential role, by its contribution to the reasons that propositions provide for judging other propositions. The question then arises how this neo-Fregean picture relates to the fundamental folk psychological notion of a person's having an attitude with a particular content. What is it for an individual human being, a particular physical thing, to have a thought involving one concept as opposed to others? As we have seen, inferential-role theories answer this question by means of an isomorphism between the network of abstract conceptual relations of propositions and the network of causal relations of states of a person: to have a thought with content P is to *instantiate*, in flesh and blood, the conceptual relations that are individuative of P.

As we have seen, the most plausible inferential-role theories will appeal to the thinker's abilities, in the sense of the causal basis underlying her conditional behavior, rather than to the conditional behavior itself. The distinction between ability and conditional behavior allows an inferential-role theory to allow for a Chomskyan competence/performance distinction. A thinker's conditional behavior will be imperfect because of interfering factors of various kinds. According to an ability-based inferential-role theory, to *have a particular concept is to have the ability that corresponds to the concept's inferential role.

This position leaves no room for a thinker to *have a concept without having the ability to make the inferences that are individuative of the concept.[3] On an inferential-role theory of content, as we saw in chapter 2, such an ability constitutes full grasp of the concept. (Since inferential-role theories hold that concepts are individuated by their inferential roles, full grasp of a concept is the ability to get

[3] On an inferential-role account, idealization cannot bridge the gap between incomplete and full understanding because an inferential-role account lacks the resources to say how the abilities of a thinker with incomplete understanding should be idealized. The ability that putatively constitutes a thinker's incomplete grasp of concept *C* is a candidate for incomplete understanding of an indefinite number of other concepts and may constitute full grasp of some concept *C*'. Why, then, should the thinker's ability be idealized to full grasp of *C*? An account that holds simply that using a particular concept is being in a state whose causal role corresponds to the inferential role of that concept has no principled reason for favoring some idealizations over others.

the concept's inferential role just right.) Thus, inferential-role theories are committed to the position that a thinker must fully grasp a concept in order to have thoughts involving it.

Peacocke's view has the virtue of making the requirement of full grasp particularly clear. Rather than first setting up an abstract structure of propositions and then finding an isomorphism between that structure and the network of a person's actual states, he specifies a concept's identity conditions directly in terms of the ability of a thinker who fully grasps it. He then explains what it is for an attitude to involve the concept by appealing to that ability. (I will come to the issue of deference below).

6.2.2 Informational accounts

Informational theories also assume DV, though this is less evident than in the case of inferential-role theories, and it will take some work to justify the claim. To begin with, talk of "full grasp," "incomplete understanding," and the like is less apt in the case of informational theories than in the case of cognitive-significance-based accounts. (The term "mastery" is somewhat better.) A requirement of understanding or grasping suggests some sort of constraint on the internal states of a thinker, while, as we have seen, informational accounts require only that occurrences of the appropriate mental representation covary with the concept's reference; what mediates that covariation is irrelevant to what constitutes content. Although it is true that an informational account does not make a thinker's exercising a certain ability *constitutive* of his having a thought involving a particular concept, an informational account nevertheless relies, as we will see, on thinkers' having what, *for an informational account*, amounts to mastery of the relevant concepts. More precisely, a person will not be able to have a thought involving a given concept without having mastery of the concept because, as a matter of nomological necessity, a person will not have the appropriate dispositions without mastery of the concept (leaving deference aside).

It is worth pointing out in advance that my argument does not turn on the use of the term "mastery" or "full grasp" – on what we should count as constituting "mastery" or "full grasp" of a concept for an informational view. What is at stake is whether inferential-role and informational theories of content are too demanding – whether they require too much of a thinker in order for the thinker to have a thought involving a particular concept. In section 6.9, I will argue that, at least by our ordinary standards of attitude attribution, thinkers apparently have non-deference-dependent thoughts involving concepts they incompletely grasp. So, if the examples are genuine cases of thoughts involving those concepts, the question is whether informational views can account for the examples. That is, what matters in the end is not whether the term "full grasp" or "mastery" aptly characterizes the minimum ability that, according to an informational view, a thinker will need to have, but whether the thinkers in the examples lack that minimum ability. With these remarks on the table, I will nevertheless show that the minimum ability that a thinker will need to have in order to *have a particular concept is, in fact, appropriately characterized as "mastery" of the concept.

Of course, since full grasp is not the same thing under an informational account as under an inferential-role account, examples of thinkers who have attitudes involving IIUCs provide an argument against informational accounts only if they include thinkers who do not have what an informational account counts as full grasp. The examples of IIUCs in section 6.9 satisfy this proviso.

An informational account begins from the idea of individuating concepts purely by their reference. From that standpoint, it is natural to answer the question of what it is for a person to have a thought involving a particular concept in terms of the person's dispositions' bearing certain causal relations to the concept's reference. Just as the inferential-role account relies on the thinker's correctly instantiating the appropriate inferential relations, the informational account relies on the thinker's getting his relations to the concept's reference right. Now it is true that having the appropriate relations to the concept's reference does not involve mastering an inferential role. But in criticizing informational theories, we must assume *their position* on what a concept's identity consists in, and thus of what fully grasping a concept involves. Since, for an informational account, inferential role is not constitutive of the concept, a thinker who has not mastered a concept's inferential role does not, at least not for that reason, lack a grasp of any aspect of the concept's nature.

If a concept is individuated by its reference, I suggest that mastering the concept entails having some understanding or other of the property that is the concept's reference. Although no particular way of understanding the reference property is necessary, not just any understanding will do; the understanding must be sufficient, together with the world, to determine that it is that property that is the reference. Remember that, as I am using the term, mastery is an ability that is sufficient (together with the world) to determine (metaphysically) which concept it is. If a concept is individuated by its reference, what is needed to have mastery of the concept is an ability sufficient to determine the concept's reference. One can have an ability sufficient to determine a concept's reference without, for example, knowing exactly what falls under the concept or having an infallible test for telling what falls under the concept.

Suppose I know that water is whatever shares the underlying nature of certain samples I have encountered, but do not know that water is H_2O and am not particularly good at ascertaining whether a given liquid is water. My knowledge is sufficient (together with the world) to determine the reference of the concept of water, so it constitutes mastery of the concept (still assuming concepts are individuated at the level of reference). More generally, then, all one needs to have mastery of a concept is some understanding of what property an object must have in order to fall under the concept ("an understanding of the concept's reference," for short). One needn't have any particular understanding of the property, so long as one's understanding is sufficient, together with the world, to determine which property it is. For example, one might know a description of the property that, given the context, is uniquely identifying. Or, one might have a recognitional ability and some way of restricting the relevant range of objects, such that the ability is sufficient reliably to discriminate the property within the range. Thus, the understanding of the concept's reference may be tacit in the sense that a thinker need not be able to explain how he recognizes something as falling under the concept.

As remarked above, nothing turns on what we count as "full grasp" or "mastery" of a concept for an informational theory. The rest of this section argues that it follows from an informational theory of content that a thinker will not be able to have thoughts involving a particular concept unless she has an understanding of the concept's reference (an understanding that is sufficient, together with the world, to determine *that* property uniquely). Given that result, all that I will need to show in order to complete my argument (that informational theories cannot adequately explain incomplete understanding) is that a thinker who lacks an understanding of a concept's reference can nevertheless have thoughts involving that concept. I try to show that in section 6.9. For the rest of the section, I assume that "mastery" of a concept for an informational theory of content is, as argued above, an understanding of what property an object must have to fall under the concept.

That a thinker who lacks full grasp of a concept will lack the right dispositions is most obvious in the case of those informational accounts that hold that a symbol's reference is whatever the symbol would be applied to in certain ideal conditions. For there are no non-trivial conditions in which a person who does not have an understanding of what property an object must have to fall under a given concept is disposed to apply the concept to all and only those objects that fall under it. It will not do to argue that in order for a thinker to be in ideal conditions, the thinker must be provided with an account of the concept's reference. A thinker who produces the appropriate covariation relations only when she is given an understanding of the concept's reference is not a thinker who *has the concept in question. In order for covariation to obtain in a nonvacuous way, the thinker must have some understanding (or other) of what falls under the concept that, together with the world, is uniquely identifying of what falls under the concept.

Pettit's dispositional theory of content is worth mentioning here, though it apparently is not purely informational since it includes inferential dispositions. Pettit holds that a concept's correct application is determined by the dispositions of its users in ideal conditions, where ideal circumstances are those "that survive intrapersonal or interpersonal negotiation about how to handle discrepant responses" and "where we have all the relevant information that could ever be available," also as fixed by the participants' practices (1993, pp. 93-94). Pettit's account avoids the need actually to specify ideal conditions and allows that participants can never be sure in a particular case that ideal conditions are satisfied. The account nevertheless implies that there are conditions under which any thinker who has a thought involving a concept necessarily is disposed to use the concept correctly. The account therefore cannot account for IIUCs. Even with all the relevant information and under the best circumstances, a person who lacks an understanding of a concept's reference will not have the right dispositions. (As discussed above, the "relevant information" cannot include an account of what the concept refers to.)

Finally, let us turn to Fodor's theory. Let's assume for purposes of the following argument that in cases of full understanding, contrary to what I argued in chapter 4, the laws connecting properties and symbols will have the patterns of asymmetric dependency that Fodor's theory assumes and depends on. If this assumption is false, then Fodor's theory of content is not a contender anyway. More precisely, suppose a person thinks that objects must have a particular property P in order to fall under a given symbol.[4] What we are granting Fodor is, first, that there will be a law connecting P to (the causing of occurrences of) the symbol. And, second, that every law connecting a property other than P to the symbol will asymmetrically depend on the law connecting P to the symbol.

As we saw in chapter 4, Fodor's sufficient conditions for a symbol's having a particular content entail the necessary condition that the symbol not satisfy Fodor's conditions for meaning any other content.[5] Thus, for example, if (in a particular instance) someone means QUUS by "plus," then (in this instance) she does not mean PLUS. So if Fodor has given sufficient conditions, then *not* satisfying Fodor's conditions for a particular symbol's meaning QUUS – or indeed for meaning any of an infinite number of QUUS-like variants of PLUS – is a necessary condition for the symbol's meaning PLUS. A thinker's ability to recognize instances of addition must be such as not to mediate a primary[6] nomic dependence on any of the quaddition-like functions. Since such functions are as similar as you like to the addition function – indeed may be exactly similar under all conditions that a thinker will actually encounter – the requirement that the thinker's tokening of "plus" not have a primary nomic dependence on any of these quaddition-like functions may turn out to be quite a strong necessary condition. In general, Fodor's conditions imply that *if* a thinker's tokening of a representation has a primary dependence on *any* property, a *necessary* condition for the representation's expressing a particular concept is that the primary dependence be on the property that is the semantic value of the concept.

Fodor's theory therefore cannot explain IIUCs. In the case of people who misunderstand what property it is that an object must have to fall under a concept, there will often be a primary, nomic dependence of tokenings of the relevant symbol on the "wrong" property, that is on some property other than the one that is the semantic value of the concept the people incompletely understand.[7] The primary dependence won't be on the right property or the people would have complete understanding (given what complete understanding is *for an informational theory of content*). But if having a representation

[4] This needs to be qualified because of the point noted above that the person's understanding could be tacit. That is, the person could have an ability to recognize things that have P as falling under the symbol, rather than anything that the person could articulate.
[5] I am simplifying by leaving aside the case of ambiguity, which does not affect the present argument.
[6] As defined in section 4.3.1, a law L governing the occurrence of tokens of a symbol S (*i.e.*, a nomic dependence of the tokening of S on a given property) is *primary* just in case all other laws governing the occurrence of tokens of S asymmetrically depend on L.
[7] Given the assumptions we are granting Fodor (see text above), if a thinker has a mistaken or partial understanding of the property, occurrences of the symbol will plausibly be nomically dependent on some (perhaps disjunctive or weird) property or other. It is a difficult question how likely it is that one nomic dependence will be primary if the thinker is confused. In cases in which the person with incomplete understanding has cottoned onto the wrong property, it seems that there would be a primary dependence on that

with a primary, nomic dependence on a property is sufficient for meaning that property, people with incomplete understanding will mean the wrong property by the relevant symbol, which implies that they will not mean the property that is the semantic value of the concept they (putatively) incompletely grasp. (In other words, Fodor's theory implies that they have not incomplete grasp of that concept, but full grasp of a different concept.)

We have not yet explored whether deference to other thinkers is a mechanism that might enable mental representations to have the right nomic dependencies in cases of thinkers with incomplete understanding. Aside from deference, however, thinkers who do not have full grasp of a concept will not have the right relations between their applications of the concept and the property that is its semantic value, for by hypothesis they lack even a tacit understanding of what falls under the concept. Thus, Fodor must accept *DV*.

In sum, informational accounts, given what counts as full mastery for them, assume DV. Since only full grasp or deference to others who have full grasp can ensure the required dispositions, I will continue to refer to the problem as a problem of incomplete understanding, although in the case of informational accounts the problem is really incorrect dispositions to token the concept, the explanation or source of which may be incomplete understanding. In this section, we have seen that both inferential-role accounts and informational accounts assume DV. The next section responds to an objection to that conclusion based on natural-kind concepts.

6.3 Natural-kind concepts and the Demanding View

It might be objected that DV cannot be widely accepted because the influential work of Putnam and Kripke on natural-kind concepts shows that people can have attitudes involving concepts they (and everyone in their community) incompletely understand. The idea would be that, for example, long before anyone understood the natural-kind concept GOLD, people had attitudes involving it.

A better understanding of the work in question, however, is that Putnam and Kripke provide a more sophisticated account than was previously available of what it *is* fully to grasp a natural-kind concept. In the case of GOLD, Putnam and Kripke show that fully grasping the concept is not a matter of knowing certain characteristics such as its superficial properties or its atomic structure. Rather, in

---

property. (I am still setting aside the possibility of deference at this point.)

order fully to grasp the concept, on their view, a thinker must understand (roughly) that gold is whatever shares the underlying nature of certain paradigm samples (or of most of them) of gold, where that underlying nature may not be known to any thinker. (Putnam also talked about the role of the division of linguistic labor in the case of natural-kind concepts like ELM, the idea being that full grasp of the concept, including the ability to identify the paradigm samples, may be had by only some members of the community. But deference to others with full understanding is not peculiar to natural-kind concepts, so there is no need to address it in this context.[8] The important point about natural-kind concepts for present purposes is what is required fully to grasp them, regardless of whether all who use them or only experts fully grasp them.)

Despite incautious remarks to the contrary, the lesson that subsequent theorists have taken from Putnam's and Kripke's work on natural kinds is not that thinkers can have attitudes involving a natural-kind concept that no one in the community fully grasps, but that fully to grasp a natural-kind concept is to think of the substance or objects to which it refers as having an underlying nature. Thus, theorists assume that to use a term to express a natural-kind concept, thinkers must intend to use the term to mean, roughly, whatever shares the underlying nature of certain paradigm samples (or must intend to defer to experts who in turn intend to use the term in that way).[9] On a more subtle account, there need be no explicit intention; rather, the thinker must have appropriate dispositions.

Peacocke, for example, suggests that part of what is required fully to grasp (one sort of) natural-kind concept *N* is to find it natural to infer from the premise that an object falls under *N* to the conclusion that the object is of the same underlying kind as other samples of N (1992, p. 26). And, as we have seen, he holds that a thinker cannot have an attitude involving a concept unless she fully grasps it or defers to others who fully grasp it. On Fodor's account, in order to use a natural-kind term, a thinker's tokenings of the term must have the appropriate nomic relations to the natural-kind property to which the term refers. Fodor explains how these nomic relations would be expected to come about by means of the thinker's intention to apply the kind-term "to all and only stuff of the same (natural) kind as paradigmatic local samples" (1990, p. 115). In chapter 4, I questioned whether such an intention would

---

[8] I discuss deference in section 6.5 below.

[9] See, *e.g.*, Schwartz 1977, pp. 40-41; Putnam 1981, 24-25, 46-47; Hurley 1989, p. 197; Fodor 1990, p. 115; Peacocke 1992, pp. 26, 144. I believe that the view described in the text is also the best interpretation of Kripke's and Putnam's original arguments, though I will not engage in Kripke/Putnam exegesis here. In his earliest presentations, Putnam especially makes a variety of confusing and confused suggestions about the significance of his arguments. In certain places, however, he suggests that he thinks an intention to use a natural-kind term in the right way is necessary for a term to mean a natural-kind concept. See Putnam, 1975,

in fact establish the appropriate nomic relations, but the present point is that a thinker who incompletely grasps a natural-kind concept would lack the appropriate intention (or tacit understanding) and therefore *a fortiori* would not have the appropriate nomic relations, unless by some semantically irrelevant mechanism.

6.4 The achievement view and the demanding view

Now that we have seen that both inferential-role and informational accounts assume DV, the next few sections challenge the motivation for that assumption. In chapter 2, we discussed a line of thought that begins from a view about the individuation of concepts and moves to a view about what it is for a thinker to have a thought involving a particular concept – the achievement view – from which it follows that the thinker must fully grasp the concept (or defer to others who fully grasp it). Let us reexamine that line of thought.

We begin from a view about what a concept is: a concept is individuated by its inferential role or by its reference. We want to give an account of the folk psychological notion of a thinker's having a thought involving a particular concept. We want the account to be naturalistic. The obvious resources available to a naturalist are the thinker's dispositions and abilities. As we saw, a plausible approach is to take each concept to be associated with a condition specifying that a thinker do something. Those who hold an inferential-role-based theory of the individuation of concepts will hold that a concept is associated with a condition specifying that a thinker make certain inferences. And those who hold a reference-based theory of the individuation of concepts will hold that a concept is associated with a condition that the thinker apply or activate a symbol or state in a way that covaries with the property that is the concept's reference. (See section 2.3.2.) It is then apparently very natural to take for granted that for a thinker to *have a particular concept is for the thinker to have the disposition or ability to satisfy the concept's associated condition. To have a thought involving the concept is to exercise that disposition or ability.

We thus arrive at the position that, in chapter 2, I dubbed "the achievement view." And, as I argued in section 2.3.1, the achievement view implies that a concept's associated condition must be such that an account of what individuates the concept can be derived from it. The achievement view holds,

pp. 225, 231, 238; also Kripke, 1980, pp. 118-122.

roughly, that a thinker must be successful in fulfilling this condition. But to have an ability the specification of which is sufficient to individuate the concept is to fully grasp the concept. The achievement view, then, is the reason that inferential-role and informational theories of content are committed to DV. As we saw in chapter 2, however, inferential-role-based and reference-based theories of concept individuation do not compel us to accept the achievement view. The responsibility view is an alternative view of what it is to *have a concept that is consistent with both inferential-role-based and reference-based theories of concept individuation. In the next section, I begin my attack on the achievement view by arguing that it cannot account for the role of deference.

6.5 Achievement-based theories and the role of deference

6.5.1 The basic options for an achievement theorist

In section 2.4.2., I pointed out that the basic form of the achievement view makes it difficult to account for incompletely understood contents. A thinker who lacks mastery of a concept will not, in general, have a disposition or ability to satisfy the concept's associated condition. At least officially, however, standard theories of content accept that there are cases of incomplete understanding. They tend to assume that those cases can be accounted for as cases in which the thinker who lacks mastery of a concept defers to others who have that mastery.

In this section, I will argue that theories that presuppose the achievement view have difficulty explaining the role of deference. Given an achievement-based position on what it is to *have a concept, how could deferring to someone else enable a thinker to *have a concept? Ultimately, I will suggest (in chapter 7) that deference is much more plausibly a way of becoming responsible to a standard than a way of acquiring the dispositions or abilities of the person to whom one defers. The phenomenon of deference-dependent thoughts, which is widely accepted by achievement theorists, thus supports the responsibility view over the achievement view.

There is no standard view about exactly what deference is, but we can say roughly what is involved. Deference is generally taken to be mediated linguistically. Certain speakers use a word to express a concept that they fully grasp. Another speaker, who does not fully grasp the concept, also uses the word. If she notices that her use of the word diverges from that of the other speakers, she tends to revise her own use to conform to theirs. This pattern of behavior constitutes the latter speaker's deferring to the others with respect to the use of the word in question. Somehow, such deference is supposed to enable her to have thoughts involving the concept that the others express by the word.

Sometimes philosophers talk of "deference to the world" rather than to a person who fully grasps the concept. We need to be careful about what this could mean. On a plausible account of certain kinds of concepts, in order to fully grasp such a concept, a thinker must understand that its reference is determined by facts or theories that go beyond human knowledge or abilities. Thus, as I discussed in section 6.3, in order fully to grasp a natural-kind concept, a thinker plausibly must understand that the reference of the concept is determined by the real underlying nature of certain objects that the thinker has encountered – even if *everyone* is wrong about (or ignorant of) that nature or about which objects share it. On a dispositional version of such a story, a thinker must, for instance, be disposed to withdraw a provisional judgment that a substance is water if it turns out to have a different underlying structure than other samples. The phrase "deference to the world" is sometimes used to refer to such a disposition. On this usage, deference is part of *mastery* of certain kinds of concepts, *not* a way of having a thought involving an incompletely grasped concept. Deference, so understood, is no help in defending the position that deference can account for cases of incomplete understanding. I will not use the term "deference" in this way; instead, I will refer to the kind of understanding involved (or the corresponding dispositions) as a *natural-kind understanding* (or *natural-kind dispositions*).[10]

We can now return to the question of the achievement view's potential to account for the role of deference in enabling thinkers to have thoughts involving incompletely grasped concepts. In what follows, I will use "deference" in my sense of a disposition to accept correction from other people.

On the achievement view, a thinker's attitude involves a particular content in virtue of the thinker's having and exercising a certain disposition or ability. A person who incompletely understands a concept will not in general have the appropriate disposition or ability. How exactly is deference supposed to enable such a person to have an attitude involving the incompletely grasped

[10] The phrase "deference to the world" could also be used to describe a thinker's tendency to defer to the world on the question of what kind of concept to acquire. For example, a thinker could have a disposition to acquire a natural-kind concept for a given substance just in case the world provides reason to think that the substance is in fact a natural kind. See section 7.7 below for further discussion. I will set aside such dispositions here since they do not provide a way of defending DV.

concept? Standard theories of content do not address this question. In order for such a theory to be able, without significant revision, to integrate the phenomenon of deference, deference must enable a person who lacks full grasp of a concept to come within the theory's standard explanation of how a person's attitude involves that concept. That is, deference has to be a means by which a person who incompletely grasps a concept nevertheless comes to have and exercise the appropriate disposition or ability. (What is the reason for this requirement? Suppose a theory allows that deference enables a person who lacks full grasp of a concept to *have a concept *without* coming to have and exercise the appropriate disposition or ability. The theory has *ipso facto* abandoned the achievement view's tenet that *having a concept is having the appropriate disposition or ability. Rather than integrating deference into its standard account of *having a concept, the theory allows that deference offers a different way of *having a concept – the option discussed in the next paragraph.)

If deference does not enable thinkers with incomplete understanding of a concept to fit within an achievement-based theory's account of what constitutes a person's attitude involving that concept, the theory can hold that deference offers an additional, different way of having an attitude involving a concept. A thinker's attitude involves a particular concept *either* in virtue of her exercising a certain disposition *or* in virtue of her deferring to another thinker's exercise of that disposition (leaving aside the complication of the linguistic mediation of deference). One trouble with this approach is that an account that merely specifies two different ways in which content may be realized loses its claim to say what it *is in virtue of which* an attitude has a particular content. Without an understanding of what the different realizers of content have in common (that makes them both realizers of a thought's having a particular content), we do not have an answer to the content question.

A related problem for the achievement theorist is that allowing that there are two different ways of having a thought involving a given concept undermines an argument that apparent attributions of Independent Incompletely Understood Contents are not genuine attributions of content. Suppose that an achievement theorist were able to account for all of the data except for my putative cases of IIUCs (see section 6.9 below). The theorist could then argue that the success of the theory provided a theoretical reason for discounting the apparent cases of IIUCs. Adding an apparently *ad hoc* clause for Dependent Incompletely Understood Contents undermines such an argument, however. Once an achievement theorist has allowed that there are two different ways of having a thought involving a particular concept – exercising the appropriate ability *or* deferring to others who have that ability – there is no longer much force to the claim that the theory's simplicity and explanatory power militates against the genuineness of cases the theory cannot account for.

We have outlined two options for an achievement theorist. First, she can argue that deference enables a thinker to come within the theorist's standard account of what it is to *have a concept. This option requires the theorist to show that deference enables the deferring thinker to acquire the appropriate dispositions or abilities of the other person – call her "the expert." I will turn to this option immediately below. Second, the achievement theorist can take the position that deference is simply a second, different way of having a thought involving a given concept. This option gives up the achievement view's attempt to offer a unified account of what it is in virtue of which a thinker *has a concept. (A theorist could, of course, offer a different, non-achievement-based account of what it is that unifies the different cases.) Moreover, this option undermines the achievement theorist's theoretically-based opposition to the possibility of IIUCs.

The only other possibility for the achievement theorist is to bite the bullet and deny that there can be deference-dependent thoughts, claiming that apparent attributions of such thoughts are just a convenient but inaccurate way of talking.

This is a highly unattractive option. (As we will see below, Peacocke flirts with this option, but understandably does not embrace it.) Burge (1979) has famously and persuasively argued that people can have thoughts involving concepts that they incompletely grasp. And this view has achieved widespread acceptance. The opposing view requires the theorist to attribute massive error to our ordinary practices of attitude attribution. The consequence would be that many of our ordinary thoughts would not in fact have the contents that we intuitively attribute to them, but instead would have various idiosyncratic contents. Such a sharp distinction between attitudes involving fully grasped concepts and attitudes involving incompletely grasped concepts seems to be driven by the assumptions of the achievement view, rather than found in the folk psychological notion of having an attitude involving a particular concept. In what follows, I will not take seriously the option of dismissing all attributions of thoughts involving incompletely grasped concepts.

Let us turn to the prospects for fitting deference within an achievement-based theory's standard account of what it is to *have a concept. Current theories of content typically make no attempt to show that their conditions for content could be satisfied by means of deference. How deference could enable a person to acquire the dispositions or abilities of the person to whom she defers is not straightforward. A person who defers to another in the use of a word does not thereby simply acquire the appropriate dispositions or abilities of the latter person.

It might be possible for an achievement-based theory to be formulated in such a way that deference-based contents would fit smoothly within the theory's account of content. Such a theory would have to be based on a dispositional condition that is common to thinkers who fully grasp a concept and thinkers who do not, but defer to others who do. Of course, the theory would have to be otherwise plausible as well. Whether such a theory can be found is difficult to answer in the abstract. I will therefore turn to particular theories of content and examine their prospects for explaining the role of deference.

6.5.2 Simple informational theories and deference

Consider a simple informational view according to which *having the concept of a butterfly is having a mental symbol that, under specified optimal conditions, covaries counterfactually with butterflies. (*Prima facie*, informational theories have better prospects than inferential-role theories for accounting for the role of deference because informational views abstract from the inner mechanisms that mediate the occurrence of the thinker's mental representations.) Consider Will who (according to a reference-based view of concept individuation) does not have mastery of the concept of a butterfly – he does not distinguish between moths, hummingbirds, and butterflies – but who would accept correction from another thinker, Vladimir, who does have mastery of the concept. Does Will's mental symbol covary with butterflies, in virtue of his deference to Vladimir?

Suppose Vladimir is not around much, or suppose that when he is around, he has no time to correct Will. It will not be true that Will's mental symbol will covary with butterflies, even under what we would normally think of as ideal epistemic conditions – good light, no distractions, and so on. The achievement theorist might point out, however, that *if* Vladimir is present, taking the job of correcting Will seriously, and so on, Will's mental symbol will covary with butterflies.

This counterfactual falls short in several respects. First, no non-intentional specification will plausibly capture the condition that Vladimir is present, taking the job of correcting Will seriously, and so on.

Second, even if we grant for the sake of argument that Will's mental symbol covaries with butterflies when Vladimir is present, taking the job seriously, and so on (and waive the problem that this condition is specified in intentional terms), the achievement theorist has not shown what she was supposed to show. She was supposed to show that Will, by deferring to Vladimir, acquires the crucial disposition (or set of dispositions) that Vladimir has. Vladimir's disposition is such that his mental symbol covaries with butterflies under the specified optimal conditions. By contrast, Will's disposition is such that his mental symbol covaries with butterflies under optimal conditions *when Vladimir is present and taking the job of correcting Will seriously*. We are still lacking an explanation of why having this disposition is a way of *having the concept, given the theory's position on what it is to *have a concept. It seems that the informational theorist is left with the position that there is a second, different way of *having a concept.

Third, it seems that Will's symbol will covary with Vladimir's use of the term "butterfly," not with butterflies. For if Vladimir misapplied the word (for example, because he loses his mastery of the concept, forgets the use of the English word, or has poor eyesight), Will would misapply his mental symbol. His application of the mental symbol tracks Vladimir's use of the word "butterfly." So the informational theorist may be forced to hold that the reference of Will's symbol is not the property of being a butterfly but the property of being that to which Vladimir applies "butterfly," or perhaps of being linked in some complicated way to Vladimir's utterances.

The informational theorist may be tempted to move to a position on which a thinker does not count as deferring to an expert unless the thinker satisfies some condition stronger than merely that he would accept correction from the expert. The idea is to make the condition strong enough that a thinker who defers will necessarily have the same dispositions as the expert. But if what counts as deference is revised in this way, many of the cases of apparent incompletely grasped contents that are supposed to be accounted for by deference will not count as cases of deference at all. It is plausible that ordinary speakers of English would accept correction from experts with respect to a wide range of terms, such as “elm,” “arthritis,” “megabyte,” “fortnight,” “baritone horn,” and so on. But it is not plausible that ordinary speakers thereby acquire the experts’ dispositions. Again, it is perhaps plausible (though far from clear) that young children are in general willing to accept correction from adults on at least a wide range of terms, but the children do not thereby acquire the adults’ dispositions.

In sum, the informational theorist (and the achievement theorist more generally) is caught between two constraints. On the one hand, if she wants deference to account for the wide range of cases that it is generally held to account for, the condition that a thinker must satisfy to count as deferring to an expert cannot be made too strong. On the other hand, if an achievement theorist is to be able to make sense of how deference enables a thinker to have a thought involving a concept she incompletely grasps, the condition that a thinker must satisfy to count as deferring to an expert must be strong enough that the thinker will not satisfy it unless she acquires the expert’s dispositions.

A further point is that folk psychology does not seem to test the extent or effects of deference before awarding attitudes. Mere membership in a community and lack of an intention to depart from its conventions is all that is assumed in typical presentations of Burge-style examples of deference-dependent attitudes. Yet those assumptions are sufficient to make attitude attributions involving incompletely grasped concepts seem intuitively correct. The claim that if such deference fails to effect the right dispositions, the thinker cannot have attitudes involving the concept has no foundation in folk psychological attitude-attribution practices.

There are, no doubt, various other moves that the informational theorist could make. The point, however, is to show that, even in the case of an informational theory, which compared to an inferential-role theory is a relatively promising case, accounting for the role of deference is not straightforward. I will not further pursue the hypothetical debate with the simple informational theorist because we have already (in chapter 3) removed purely dispositional informational and inferential-role theories from serious consideration.

6.5.3 Fodor's asymmetric-dependence theory and deference

Fodor's asymmetric-dependence theory is our only example of a non-dispositional informational theory, so let us consider its prospects for accounting for the role of deference. I have already argued that even in the case of full understanding, a thinker will not have the patterns of nomic relations required by Fodor's account. If a thinker with full understanding will not have the required patterns of nomic relations, a thinker with incomplete understanding certainly will not acquire the right nomic relations by being willing to accept correction from others.

For the sake of argument, we therefore need to assume that Fodor's theory successfully accounts for cases of full understanding, and ask whether deference would ensure that a thinker with incomplete understanding would have the same pattern of nomic relations between mental representations and worldly properties as a thinker with full understanding. Questions about the existence of such patterns of nomic relations are, as we will saw in chapter 4, difficult to evaluate. (And the problem is exacerbated by the difficulty of making sense of the assumption that, contrary to what I argued in chapter 4, a person with full understanding will have the pattern of nomic relations that Fodor's theory requires.) On the face of it, however, it is not at all clear that deference would have the necessary effect.

The constraints that we encountered above are relevant here. If deference is to cover the cases it is supposed to cover, the most that can be required for a thinker to count as deferring to an expert is that the thinker would accept correction from the expert. But it is not at all clear that the fact that a thinker would accept correction from an expert ensures the obtaining of a primary nomic relation between the property to which the expert applies the word and the thinker's mental representation. Even if deference establishes a nomic relation between the relevant mental representation of the deferential thinker and the use of the relevant word by the expert, all that is clear is that the former thinker's mental representation now has a new nomic relation unlike the nomic relations of the expert's mental representations. And if the reliance on the expert did establish a nomic relation between the deferring thinker's mental representation and the appropriate property, that nomic relation would not be primary. For that nomic relation could be eliminated – for example by altering the deferring thinker's ability to recognize the expert – without eliminating other nomic relations of the deferring thinker's mental representation. Fodor makes no attempt to argue that his theory can account for deference-dependent attitudes, and since we found that Fodor's theory conclusively failed on its own terms, I will not further pursue the issue of its potential to account for the role of deference.

6.5.4 Peacocke's theory and deference

Let us now turn to inferential-role theories of content. We have seen that purely dispositional inferential-role theories are not serious contenders. The most plausible inferential-role theories hold that *having a concept is having the ability that corresponds to the concept's inferential role. Peacocke's theory is our model of such a theory.

Before turning to the details of Peacocke's attempt to address the issue of deference, it is easy to see that this general type of theory is in a worse position than informational theories to make sense of the role of deference. On an informational theory, *having a particular concept is just having a mental representation that covaries appropriately with the concept's reference. It does not matter what abilities mediate that covariation. But the inferential-role theories in question hold that *having a particular concept is not having the right inferential conditional behavior but having a particular ability, a particular causal basis of the behavior. It is difficult enough to see how deference could be a way of acquiring an expert's conditional behavior. But if a thinker does succeed in acquiring the expert's conditional behavior, it will not be because he has acquired the ability that is the causal basis of the expert's behavior.

Consider the following example. In explaining to a non-native speaker of English an example involving a child who does not understand the concept of a stevedore, I realized that I incompletely understood the concept myself: I was not sure whether the concept includes only workers who load and unload ships or whether it can include workers who load other kinds of carriers, such as trains. It seems plain that by ordinary standards of attitude attribution, I have occasionally had thoughts involving the concept of a stevedore (or have meant the concept by the word "stevedore"). For example, I might have seen a large person near the docks and wondered whether he was a stevedore. But I couldn't have deployed the ability that corresponds to the inferential role of the concept since I did not have that ability: for example, I did not recognize that the fact that someone is a stevedore provides a reason for judging that he works unloading ships.

It might be argued that by deferring to other thinkers who incompletely understand a concept, a thinker manages vicariously to acquire the ability that corresponds to the concept's inferential role. But a person who would be willing to accept another's judgment over his own does not thereby come to be able to exercise the ability of the other person to recognize the reasons that judgments provide for making other judgments. Merely because I would have accepted correction had I been informed that a stevedore must be a dock worker, it is not the case that when I had the thought that, say, a particular man looked like a stevedore, I was exercising the ability that corresponds to the inferential role of the concept. Strictly speaking, accepting that there are *any* attitudes involving incompletely understood contents is incompatible with a view on which using a particular concept is exercising an ability that corresponds to its inferential role.

Let us turn to Peacocke's theory, our model of an inferential-role theory. Peacocke makes an explicit attempt to address deference, so we need not speculate as to how he would treat the issue. Peacocke is clear, as I have noted, that a possession condition is a condition for fully grasping or mastering a concept. (And Peacocke's argument that possession conditions individuate concepts – whatever its merits otherwise – would be a non-starter were possession conditions not conditions for fully grasping a concept (pp. 5-6).)

Peacocke officially recognizes, however, that a thinker need not satisfy a concept's possession condition in order to have attitudes involving it (pp. 27-30). He accepts that Burge has shown that thinkers with incomplete understanding of a concept can still be correctly attributed attitudes involving that concept. For this reason, Peacocke acknowledges that possession conditions, which state "what is required for full mastery of a particular concept," are different from "attribution conditions" – the conditions for correctly attributing a thinker a thought involving a concept (p. 29).[11] But Peacocke fails to integrate the phenomenon of thoughts involving incompletely grasped concepts into his theory. After pointing out that thinkers can have thoughts involving incompletely grasped contents, he ignores the issue and often relies implicitly on the assumption that thinkers fully grasp the concepts that figure in the contents of their thoughts. For example, Peacocke's explicit account

[11] See also p. 249, fn. 7.

of what determines the content of thought – his solution to Kripke's rule-following problem – does not address the issue of incomplete understanding (pp. 133-145, 190-197). Deference dependence, the only aspect of the issue that Peacocke recognizes, is barely mentioned after the section in the first chapter that introduces it, though he notes that it "merits extensive independent discussion" (p. 238, n. 19).[12]

As noted in chapter 5, Peacocke contrasts possession conditions, which state "what is required for full mastery of a particular concept" with "attribution conditions," which state the conditions under which a thinker has a belief involving a particular concept.[13] He offers three conditions that he asserts are "jointly sufficient" for a thinker to have a thought involving a particular concept, the concept RED, for example:

a. The subject is willing sincerely to assert some sentence of the form "___ red ___" containing the word "red" (or some translation of it).

b. He has some minimal knowledge of the kind of reference it has (*e.g.*, that it is a color word).

c. He defers in his use of the word to members of his linguistic community.

As Peacocke emphasizes, these conditions are much weaker than a possession condition; a thinker who does not satisfy a concept's possession condition can satisfy the three conditions.

Peacocke has little more to say about the nature of attribution conditions and their relation to possession conditions. For example, he does not address the issue whether there are other sufficient conditions for a thinker who does not have mastery of a concept to have a thought involving it. (In fact, he does not explicitly give attribution conditions for a thinker with full mastery of a concept, though they can be inferred from his view of what it is to have a thought involving a particular concept.) And he does not attempt to explain *why* the three attribution conditions are sufficient for having a thought involving a concept.

As explored in chapter 5, on Peacocke's main view, what makes it the case that a person's thought involves a particular concept is that the person finds certain inferences (involving that

[12] For other examples of Peacocke's reliance on an assumption of concept mastery, see Peacocke's treatments of radical interpretation (pp. 37-38), of Gareth Evans's generality constraint (pp. 42-45, 57), and of the relations between possession conditions and belief states (pp. 106-115).
[13] Rather than simply stating that attribution conditions are the conditions under which a thinker has a thought involving a concept, Peacocke says that attribution conditions (for the concept RED) are "the conditions under which something of the form 'x believes that ___ red ___' is true." I discuss below why Peacocke resorts to the formal mode.

thought) compelling and that his finding them compelling has a certain causal explanation. The underlying idea seems to be that *having a concept is having a certain ability – the causal basis of the thinker's finding certain inferences compelling – and having a thought involving the concept is exercising that ability. We would expect Peacocke to say why, given this view, satisfying his three attribution conditions for a particular concept is a way of having a thought involving that concept, or, at least, ensures that one is having a thought involving that concept.

Peacocke given us two different sets of sufficient conditions for a thought's involving a particular concept. A thinker can have and exercise the ability that corresponds to satisfying the concept's possession condition. Or a thinker can know a little about the kind of thing that the concept word refers to and defer in his use of the word. Peacocke does not suggest that the two conditions correspond to two ways of having some more fundamental property in virtue of which a thinker has a thought involving a particular concept. And it is clear that he does not think that the deferring thinker will acquire the appropriate ability of an expert to whom he defers. So Peacocke's acceptance of deference-dependence is the thin end of a wedge. If deference is a way of having a thought involving a concept one incompletely understands, then what it is to *have a concept cannot be to have the ability that corresponds to the concept's inferential role. Peacocke's account of what it is to have a thought involving a particular concept thus lacks a way of making sense of how deference enables a thinker to have a thought involving a concept he incompletely grasps.

Suppose it were suggested that Peacocke could say that it is the expert's finding certain judgments compelling and the causal explanation of her doing so that determines the content of the deferential thinker's thought. Such a technical device would not help to make sense of the role of deference. It would leave us with two different ways of having a thought involving a particular concept: one's inferential dispositions could have the right causal explanation or one could defer in one's use of a word to one whose inferential dispositions have the right explanation. We would still lack an account of what a thought's involving a particular concept *is* such that these are two ways of accomplishing it. And we would still be unable to preserve the simple view on which what it is for a thought to involve a particular concept is for the thinker to exercise the ability that corresponds to the concept's inferential role.

Peacocke displays a revealing ambivalence with respect to deference-dependent attitudes. On the one hand, as we have seen, he officially accepts that it is false that a thinker who has attitudes involving a concept must fulfill the concept's possession condition (Peacocke 1992, pp. 27-30). On the other hand, Peacocke's discussion of deference-dependent attitudes suggests that his view of their status is not unequivocal. He notes, for example, "that while possession conditions are given for concepts, attribution conditions and deference-dependence primarily concern words and sentences" (pp. 29-30). Peacocke does not elaborate, and the force of this remark is unclear. Perhaps Peacocke is simply calling attention to the fact that his three conditions are cast in terms of the concept word, rather than the concept. For example, the conditions require deference in the use of a word. [14]

It is clear from Peacocke's discussion, however, that, in a more important sense, deference-dependence concerns concepts and thoughts. Peacocke accepts that what is achieved by means of deference in the use of a concept-word is *a thought involving the concept*. He introduces deference dependence by showing that the claim that "[a] thinker who has attitudes to thoughts containing the concept *red* must fulfill its possession condition" must be false (pp. 27-29). The falsity of this claim implies that a thinker can have attitudes to contents involving a concept that she only partially understands, which is what motivates Peacocke's discussion of deference dependence. That Peacocke accepts that deference can enable thinkers to have attitudes toward incompletely grasped concepts is all the more clear because Peacocke is led to deny the quoted claim by the propositions: 1) that a thinker with incomplete understanding of the concept RED may believe, for example, that an object is not red; 2) that "it follows that he stands in a propositional-attitude relation to a thought containing the concept *red*" (p. 28).[15]

Peacocke also characterizes attribution conditions in a way that hints at reservations. As noted above, he says that attribution conditions for RED are "the conditions under which something of the form 'x believes that ___ red ___' is true," rather than simply saying that attribution conditions for RED are the conditions under which someone has a belief involving that concept (p. 29).

[14] It is possible to question Peacocke's assumption that deference must be deference in one's use of a word to someone else's use of that word, but I cannot explore the issue here.

[15] Peacocke also indicates that the attribution conditions are *not* conditions for merely a metalinguistic attribution of belief "such as the belief that whatever property is expressed by the word 'red,' it is thus and so" (p. 29). Further evidence is Peacocke's reference to a correct "attribution of an attitude with a content containing the concept *red*" in a case of partial understanding (p. 29). For other examples, see Peacocke's approval of Putnam's point that, without understanding what distinguishes elms from beeches, he can "still believe that no elm is a beech" (p. 30) and Peacocke's discussion of what can be informative for someone with incomplete understanding "who can still properly be said to express beliefs involving the concept" (p. 32).

Peacocke's use of the formal mode here might be taken to suggest that, despite his official view, he has an inclination to view the attribution of deference-dependent attitudes as merely a convenient way of talking. Such a tendency would also explain the insignificant position to which Peacocke relegates deference-dependence, which would otherwise be surprising given the apparent inconsistency of deference-dependence with an achievement-based inferential-role theory such as Peacocke's.

Further evidence of Peacocke's ambivalence is the following passage recognizing that deference-dependent attributions are useful:

> An individual for whom only deference-dependent attributions involving a certain expression are true may nevertheless be a reliable gatherer and transmitter of information about the referent of the expression. If we want to report the mental state of such an individual, we will often have to resort to a deference-dependent attribution, for we will often not know in what idiosyncratic and partial way he is thinking of the referent of the expression. (1992, p. 33)

This passage hints that a person who incompletely grasps a concept cannot have attitudes involving it in a full-blooded sense because he lacks the appropriate way of thinking of the concept's reference. The suggestion seems to be that deference-dependent attributions are not genuine attributions of the content mentioned in the that-clause, but merely the most informative way, given our limited knowledge and the limited expressive power of our language, of talking about how the person is thinking about the object in question.

This undercurrent in Peacocke's theory is in tension with Peacocke's explicit acceptance that, because of deference-dependence, possession conditions are not "attribution conditions" and that it is correct to attribute attitudes involving an incompletely grasped concept to a thinker who defers to others.[16] If deference to others can make it correct to attribute an attitude involving a particular concept, an account of what it is to have an attitude involving a concept cannot hold that it is to exercise an ability or way of thinking that corresponds to the concept's inferential role. Of course, it could be argued that folk psychology must, strictly, be in error in accepting deference-dependent attitudes because the

[16] 1992, pp. 27-30.

achievement view has to be right. Peacocke makes no attempt to argue this claim, however, and it is not obvious how the argument for an error theory would go.

The point is not that a view that individuates concepts in terms of their inferential role cannot accommodate deference-dependent attitudes. As I argue below, such a view of concept *individuation* does not force us to accept the achievement view's position on what it is to *have a concept. The present point is that Peacocke has, in accepting deference-dependent attitudes, departed from the achievement view's model according to which using a concept is exercising an ability that corresponds to the concept's inferential role. The apparent naturalness of the achievement view cannot be traded upon by an inferential-role theory that accepts deference-dependent attitudes.

6.6 Detaching theories of concept individuation from the achievement view

One of the most influential objections to inferential-role theories of content is that there is no inferential role that actual users of a concept all instantiate. For example, Fodor points out in criticizing inferential-role theories: "The pattern of causal dispositions actually accruing to a given mental state must surely diverge very greatly from the pattern of inferences characteristic of its propositional object" (1985, p. 96).[17] In his review of Peacocke's *A Study of Concepts*, Fodor conjectures "that there aren't any non-trivial inferences" that a concept "requires its possessors to have come what may" nor "perceptual judgments" that a concept requires its possessors to make (1993, p. 15). Since Fodor does not distinguish possessing a concept from being able to have thoughts involving it, his point is that there is no inferential ability or disposition that a thinker must have in order to have thoughts involving a particular concept.

Like Peacocke, Fodor does not distinguish inferential-role theories of concept individuation from achievement-based inferential-role theories of content, never considering the question whether an inferential-role view of concept individuation can be combined with a theory of what it is to *have a concept according to which using a concept does not require instantiating its inferential role. He therefore thinks that the fact that there is no inferential ability that a user of a concept must have makes hopeless Peacocke's attempt to individuate concepts by their inferential roles. If an inferential-role theory of concept individuation can be detached from the achievement view, however, it would not be

[17] For another example, see Schiffer 1987, p.109.

susceptible to objections of this sort.

Unsurprisingly, an achievement-based theory that individuates concepts at the level of reference – in particular, an informational theory of content – may be subject to a parallel objection, for it is plausible that there is no disposition to token symbols that a thinker must have in order to have attitudes involving a given concept, at least according to our ordinary practices of attributing content (see, *e.g.*, Schiffer 1987, pp. 84-85, 102). At any rate, if section 6.9 is right that thinkers who do not understand what a concept refers to can have attitudes involving it, informational theories, including Fodor's asymmetric-dependence theory, are untenable. For, as we saw in section 6.2, informational theories are committed to DV. My argument in section 6.9 that there are IIUCs will provide further ammunition for attacks of the kind Fodor sketches, but my argument will cut against informational as well as inferential-role versions of the achievement view. Thus, the type of objection Fodor sketches cuts against the achievement view, rather than against an inferential-role-based theory of concept individuation.

There are powerful reasons, going back to Frege, for thinking that concepts are individuated more finely than reference. Informational theories struggle to find the resources to deal with the apparently fine grain of content (*e.g.*, 1994, chap. 2 Fodor, 1998, chaps. 1-4). With respect to the question of concept individuation alone, a inferential-role-theory is thus *prima facie* more attractive than a informational theory. As noted, if we do not separate a theory of the individuation of concepts from a theory of what it is to use a concept, an inferential-role theory seems to face the objection that there are no inferences that users of a particular concept necessarily make. From that point of view, an informational theory's failure to capture a concept's inferential-role may seem to be an advantage. Once we separate the question of concept individuation from the content question, however, it makes sense to take an inference-based theory of concept individuation as a working hypothesis and to consider whether it can be combined with a satisfactory theory of what it is to *have a concept. In the rest of this chapter and the next, I will therefore take as a primary focus theories of content that rely on inferential-role theories of concepts; I will continue, however, to consider the applicability of the argument to theories that rely on reference-based theories of concepts.

6.7 Normativity and theories of content

In this section I argue that, once we have distinguished the question of what individuates a concept from the question of what it is for a person to *have a given concept, an answer to the latter question won't be plausible if it gives us no reason to think that it accounts for the normative aspects of *having a concept. I begin with a reminder about why this long section comes at this point in the overall argument. In the next section, we will come to the putative data of IIUCs. Since standard theories of content cannot account for IIUCs, there will be theoretical pressure to reinterpret or otherwise reject the data. That theoretical pressure is weaker to the extent that standard, achievement-based theories are otherwise problematic. In previous chapters, we have raised doubts about the success of achievement-based theories on their own terms. And in earlier sections of this chapter, we have seen that achievement-based theories cannot comfortably handle even deference-based cases of incomplete understanding. In this section, we see that achievement-based theories arguably face an additional problem that cuts across cases of complete and incomplete understanding.

The gist of my argument in this section is this: suppose that all that is required of a correct constitutive account of *having a concept (with respect to the normativity issue) is that it specify a non-intentional property that is sufficient for the relevant normative properties in all nomologically or metaphysically possible worlds. So the supposition is that the account need not provide an intuitively satisfying explanation of *why* the specified property is nomologically or metaphysically sufficient for the relevant normative properties. Even if this is true as a metaphysical matter, as an epistemological matter we would not have reason to accept that the account is correct unless it did give some such explanation. The reason, in a nutshell, is that all the serious candidates for theories of content take a relatively *a priori* approach. But no attempt to show *a priori* that a non-intentionally specified condition is nomologically or metaphysically sufficient for certain normative features can get off the ground unless it demonstrates some *a priori* or intuitive connection between the proposed condition and the normative features.

Let us turn to the idea, discussed in chapter 1, that an answer to the question of what it is to *have a concept must account for what has been called the "normativity" of content. Kripke's (1982) argument that dispositions cannot provide an account of content did not rest merely on the claim that dispositions can be systematically mistaken, are finite, and so on, and therefore will not correspond to correct applications. Kripke also made a second, apparently different claim: that dispositions are altogether the wrong kind of thing with which to account for content because content is fundamentally normative in a way that dispositions are not. Specifically, the content of what you mean or think – whether, for example, your thought involves concept *C* or concept *C*' – determines which judgments are correct and which are incorrect. But having a particular disposition, Kripke argued, cannot make it correct to act in accordance with it. Therefore, your having a disposition to judge in a way that accords with concept *C* cannot be what makes it the case that your thought involves concept *C*. Kripke draws a general lesson: "The relation of meaning and intention to future action is normative, not descriptive (1982, p. 37)." Since what dispositions you have is a descriptive fact, he concludes that what content your thoughts have cannot be a matter of what dispositions you have.[18]

Paul Boghossian has pursued this idea. As he points out, Kripke's first argument is, in effect, that we cannot specify in non-intentional and non-semantic terms "a property M such that: possession of M is necessary and sufficient for being a disposition to apply an expression in accord with its correctness conditions" (1989, p. 532).[19] Kripke's second claim, according to Boghossian, is that:

> even if there were a suitably selected disposition that captured the extension of an expression accurately, that disposition could still not be identified with the fact of meaning, because it still remains true that the concept of a disposition is descriptive whereas the concept of meaning is not. In other words, according to Kripke, even if there were a dispositional predicate that logically covaried with a meaning predicate, the one fact could still not be identified with the other, for they are facts of distinct sorts. (p. 532)

Boghossian thus interprets Kripke as suggesting that even if it is *logically necessary* that a dispositional property and a content property (for example, one's thought involving the concept PLUS) are always coinstantiated, a theory of content that specifies the dispositional property will have failed to naturalize the content property.

Boghossian goes on to consider "what precisely has been left over" if an attempt to provide a reduction of the notion of extension succeeds in specifying "dispositions that mirror the extensions of expressions correctly" (p. 533). He considers the idea that a reduction would "not merely specify [an

[18] As Paul Boghossian points out, a number of writers have accepted Kripke's argument (on both points). Boghossian cites Simon Blackburn, John McDowell, and Crispin Wright (Boghossian 1989, p. 533 fn. 44).

[19] As Boghossian notes, though not in these terms, the proviso that M is necessary is required merely in order to state an upward entailment claim (p. 532 fn. 43). If M were merely sufficient, there would be no guarantee that the condition would encompass all the objects in the expression's extension, so we would not even have a sufficient condition for the expression's having that extension. It is unclear whether Boghossian intends to assume a downward as well as an upward entailment; I will assume that he does in what follows.

expression's] extension correctly, it would also reveal that what it is specifying is an *extension* – namely a *correctness* condition" (p. 533).[20]

Boghossian ultimately formulates the following requirement.[21] An adequate theory would need not only to specify a necessary and sufficient condition for being a correct application, it must also show how its proposed condition "could amount to something that deserves to be called a correctness condition, something we would be inherently motivated to satisfy" (pp. 533-534). Boghossian explains that "[t]his would ensure the intensional equivalence of the two properties in question, thus paving the way for an outright reduction of meaning to dispositions" (p. 534). He also gives one further characterization of what he calls the "intensional" requirement: "that possession of the favoured *M*-property appear intuitively to resemble possession of a correctness condition" (p. 535).

As argued in the discussion of the Kripkenstein problem (section 1.3.2), the notion of normativity on which Kripke's argument depends (and that is at issue in this section) is a thin one. There is no suggestion, for example, that one should use certain concepts as opposed to others, or that one should use words with the same meaning as in the past or with the meaning that they have in one's community. Rather, the relevant normative feature is that the concepts involved in one's thought determine standards of correctness for one's thought. For example, whether it is correct to make a certain judgment about an object depends on the concepts involved in the judgment. If concepts are individuated by inferential role, the relevant standards may include not only standards for applications of concepts to objects, but also standards for transitions between mental states (see also section 3.4.2). Perhaps the concept WALK determines that a thinker who judges that someone is walking should, if she considers the question, judge that the person is moving. It is worth noting that such standards may conflict with standards of truth. If the example just given is correct, a thinker who falsely judges that a stationary person is walking is required (if she considers the

[20] He puts it another way: "To be told that 'horse' means *horse* implies that a speaker ought to be motivated to apply the expression only to horses; whereas to be told, for instance, that there are select circumstances under which a speaker is disposed to apply the expression only to horses, seems to carry no such implication."

[21] Although Boghossian agrees that there is an additional "intensional" requirement, he explicitly rejects Kripke's conclusion that a dispositional account is therefore ruled out of court. Boghossian suggests that it might not be true that no dispositional condition could satisfy this additional requirement. He thinks we cannot rule out in advance the possibility that a particular dispositional proposal could "captur[e] the normative force of an ascription of meaning." I want to focus for now on the nature of the requirement, as opposed to the question of whether a dispositional account has any prospect of satisfying it.

question) to judge, falsely, that the person is moving. In this section, for simplicity, I will focus on standards for applications of concepts to objects.

Let us consider two preliminary objections to the idea that the normativity of content is something that theories of content need address. First, an achievement theorist might deny that content is truly normative. It might be claimed, for example, that the sense in which it is correct to apply a concept to objects in its extension is analogous to the sense in which the planets should move in an elliptical orbit. By saying that the planets should move in an elliptical orbit, one means only that, other things being equal, they would do so.[22] So the suggestion would be that if my concept *C* is the concept HORSE, the only sense in which it is correct to apply *C* to horses and not to cows is that I would do so, other things being equal. If this is the only sense in which content is normative, an achievement view is well suited to explain it. A thinker who has a disposition to satisfy a condition will do so if no interfering factors are present.

This suggestion denies the apparent normativity of content, and offers a reinterpretation of the data according to which content is no more normative than the motion of the planets. A full discussion would require evaluating the strength of the reasons that could be offered for the reinterpretation. In line with my goal of seeing how far we can go in accounting for our ordinary notion of content, I will proceed on the assumption that content is genuinely normative, that, for example, the extension of a term is the set of objects to which it is *correct* to apply the term, not just the set of objects to which thinkers would apply the term, if they were not distracted or tired, and so on.

The achievement theorist might try to argue, differently, that even if content *necessarily* has certain normative properties, those properties are not *constitutive* of content. Therefore, the objection continues, an account of the nature of content need not explain the normative aspects of content. To take an analogy, it is plausibly a necessary truth that the fact that an action causes pain counts morally against that action.[23] But we would not hold that a proposed neurophysiological reduction of pain would fail on the ground that it would fail to account for the moral badness of pain. Intuitively, the reason we would not impose this additional requirement is that moral badness is a necessary property of pain, but not a constitutive one. Moral truths, not just truths about the nature of pain, are

[22] The *ceteris paribus* clause has to be interpreted in a non-normative way. Pietroski and Rey (1995) develop such an interpretation.
[23] Martin Davies suggested this example.

needed to explain the moral badness of pain. Similarly, an account of the nature of the number seven need not account for all of the necessary properties of that number.[24] The point, then, is that a constitutive account of a phenomenon need not by itself entail all necessary truths about the phenomenon. They may be entailed by the constitutive account plus other necessary truths, *e.g.*, other moral or mathematical truths. Thus, the achievement theorist's claim would be that a theory of content need not account for the normativity of content. If the theory of content is true, the normative consequences will follow.[25]

The achievement theorist's position is *prima facie* implausible. As Boghossian says, "the notion of the extension of an expression just is the notion of what it is correct to apply the expression to" (p. 530). We understand what it is for something to be a pain in a way that is independent of its moral qualities: its being a pain consists in how it feels. Similarly, we understand what it is for something to be the number 7 in a way that is independent of (say) its property of being the smallest Mercenne prime. The point seems not to be merely epistemic: it's not just that we lack a way of discovering whether someone means PLUS without knowing what judgments are correct for the person to make; rather, we have no grip on what it *would be* for someone to mean PLUS but for the appropriate judgments not to be correct. (In contrast, we do have a grip on what it would be for someone to be in pain, but not to be in a state that is even *prima facie* morally bad.) In fact, as an epistemic matter, it is not that we somehow ascertain what judgments are correct and then use that information to determine what the person means. When we decide that someone means PLUS we *ipso facto* decide what judgments are correct, and *vice versa*. For the rest of this section, I will take it that content is constitutively normative – that certain judgments or transitions being correct and incorrect (under certain conditions) is part of what it is for a person to be using one concept rather than another. (I believe that a version of this section's argument is plausible even if the seemingly constitutive relation is merely epistemic – that it is, even if it is merely that we cannot identify

[24] In a recent talk in Oxford, Paul Horwich argued that though it was conceptually necessary that meaning was normative, its normativity was not "intrinsic" to it. According to Horwich's explication of this claim, however, to say that normativity was not to intrinsic to meaning was to say merely that a constitutive account of meaning would not need to use normative notions. So Horwich's point may be simply to deny Kripke's over-strong claim that we can know in advance that no descriptive account of meaning can succeed. Let's grant that there is no advance requirement that we use normative notions in a constitutive account. A constitutive account may still need to account for the normativity of content, as I discuss in the text below.

[25] This could be understood as a burden of proof claim: in order for me to show that my account of content is true, I don't have to show that my account makes it possible to explain the normative features of content. Rather, you have to show that my account makes it impossible for them to be explained. But this is not the point of the objection that I am considering. Rather, the objection is that the constitutive account need not itself contain the materials to account for the normativity of content.

content independently of its necessary normative consequences. I will not attempt to argue for this point here, however. Those who reject the view that content has constitutive normative properties may wish to skip the rest of the section.)

Since normativity is a constitutive part of content – that is, since, *e.g.*, part of what it is for a judgment to involve a particular concept is for the judgment to be correct in some circumstances and incorrect in others – a theory of content does need to account for normativity. As discussed in earlier chapters, naturalistic accounts of content have taken it to be their task to offer accounts of natural facts that could separate what falls under a concept from what does not. But they have generally ignored Kripke's claim about normativity or denied that it is a distinct problem. On the other hand, even Jerry Fodor acknowledges some ambivalence. After developing his asymmetric-dependence theory, Fodor admits that he is "sort of in sympathy" with the need to account for normativity. Further, he writes that it is "not plausible" that asymmetric dependence accounts for normativity: "There's no obvious reason why the fact that one way of using a symbol is asymmetrically dependent on another implies that we should prefer the second way of using it to the first."[26] At the same time, Fodor adds a footnote outlining his doubts:

> The trouble is that requiring that normativity be grounded suggests that there is more to demand of a naturalized semantics than that it provide a reduction of such notions as, say, *extension*. But what could this "more" amount to? To apply a term to a thing in its extension *is* to apply the term correctly; once you've said what is it that makes the tables the extension of "table"s, there is surely no *further* question about why it's *correct* to apply a "table" to a table. It thus seems that if you have a reductive theory of the semantic relations, there is no job of grounding normativity left to do. (1990, p. 135, fn. 35)

It is certainly true that "if you have a reductive theory" of a semantic notion such as the notion of extension, there is no further problem of grounding normativity. The claim is rather that in order to qualify as a reduction of a semantic notion, an account must successfully ground the normative aspect of those notions. For example, as Fodor says, to apply a term to a thing in its extension is to apply it correctly. On the view that normativity must be grounded, whatever it is in virtue of which my term has a certain extension must be something that makes it *correct* to apply the term to certain things and not to others. It is no answer, but a begging of the question, according to this argument, to say that what makes it correct to apply the term to these things is that they fall within its extension.

[26] In passing, he gestures at a possible way to "finesse" the normativity issue by appealing to biological function not to determine content, but to ground the normativity of an independent, nonteleological account of content.

Let us consider more generally what is required of a theory of content. That is, let us consider what a theory needs to do to account for the constitutive features of content, whatever those features may be. For simplicity, assume that a theory of content takes the form of specifying necessary and sufficient conditions for content.[27] (For example, an informational theory might propose that a mental representation means a particular worldly property if and only if the representation bears a specified relation to the property.) Consider a few possibilities, in order of increasing demandingness, as to what is required for a theory of content to be successful.

The first possibility is that all that is required is that the proposed necessary and sufficient conditions hold with nomological force. That is, it must be the case that, in all nomologically possible worlds, whenever the proposed conditions for, say, a thought's having content *K* are satisfied, the thought has content *K*.

According to the second possibility, nomologically necessary and sufficient conditions are not enough. The proposed conditions must be successful in all metaphysically (or logically)[28] possible worlds.

A third possibility is that the proposed conditions must satisfy some more stringent requirement than nomological or metaphysical necessity. One version, suggested by some of Boghossian's remarks, is that a theory of content must provide a conceptual analysis of the notion of content. Another, weaker version – also suggested by Boghossian's discussion – is that a theory of content must offer an intuitively satisfying explanation of the constitutive features of content. With respect to normativity, a theory must explain why a person's having a certain property – a dispositional property in the case of most leading theories of content – makes it correct for her to apply a term in one way rather than another. In other words, the demand is for some kind of intuitive explanatory, or perhaps even *a priori*, connection between the relevant property and correctness that will give us an understanding of how the property could realize or constitute content. Since I will argue only that there needs to be an explanatory connection between candidate conditions for content and the appropriate normative features of content, we can set aside the stronger requirement of a conceptual or analytic connection.

---

[27] As we saw in chapter 4, Fodor maintains that a theory of content need offer only sufficient conditions for content. This difference won't matter to the argument of the present section.
[28] For present purposes, the difference that some philosophers recognize between logical and metaphysical possibility won't matter.

We thus have a range of possible views about the requirement on a theory of content. 1) It must offer conditions for content that are necessary and sufficient for content in all nomologically possible worlds. 2) It must offer conditions for content that are necessary and sufficient for content in all metaphysically possible worlds. 3) In addition to 1) or 2), there must be an explanatory connection between the proposed condition and all constitutive properties of content. (In requirement 3), an explanatory connection must be more than a reliable correlation; otherwise 3) would not be an additional requirement. Rather, it must be intelligible how the proposed condition could realize the properties to be explained.) I am going to argue that, for present purposes, we need not resolve the question of which is the appropriate requirement on a theory of content.[29] Even if 1) or 2) is the appropriate requirement, a theorist of content will have to establish an explanatory connection between conditions for content and normativity in order to demonstrate that 1) or 2) is satisfied. Readers who accept that 3) is the appropriate requirement can skip the next eight paragraphs.

Suppose that an achievement theorist proposes a condition J (which might be, for example, that a thinker's mental representation covaries in a specified way with a worldly property) that she claims is necessary and sufficient for a thought's having content *K*. (For simplicity, I will write in terms of a condition for a particular content, rather than a schema for generating the necessary and sufficient conditions for different contents.)

To begin with, the achievement theorist cannot establish an explanatory connection between condition J and content *K* without establishing an explanatory connection between J and the correctness conditions appropriate to *K*. Since the relevant normative properties are constitutive of content, an explanation of why a thought that satisfies J has content *K* has to be an explanation of why a thought that satisfies J has the correctness conditions appropriate to *K*. For example, if it is constitutive of a judgment's involving the concept HORSE that it is correctly applied to horses, then nothing counts as an explanation of why a judgment that satisfies condition H has the content *that's a horse* unless it explains why a judgment that satisfies condition H is correctly applied to horses.

Suppose, next, that we accept for the sake of argument that requirement 1) – a condition for content must be necessary and sufficient in all nomologically possible worlds – is the appropriate

[29] For purposes of understanding the place of content in the natural world, it is arguable that we need not worry about possible worlds in which the laws of nature do not hold. Thus, the difference between nomological and metaphysical (or logical) necessity may not be significant. This point will not affect my argument in the text.

requirement on a theory of content. Is it possible for the achievement theorist to demonstrate that condition J is necessary and sufficient for content *K* in all nomologically possible worlds, without establishing an explanatory connection between condition J and the correctness standards associated with content *K* (and therefore, as just argued, without an explanatory connection between J and *K*).

Obviously the achievement theorist won't be able to rely on *a priori* arguments; if there were such arguments, there would be an explanatory connection between J and *K*. One thought might be that the theorist could use empirical methods to show that there is a lawful connection between J and *K*. Suppose that a theorist of content had discovered a regularity between a non-intentionally described condition J and a content *K*. And suppose the theorist had then checked thoughts that satisfied condition J and found that they all had content *K*. Or suppose, further, that the theorist had found a way of generating thoughts that satisfied condition J, and confirmed that such thoughts always had content *K*. If something like this were the situation, we might have to face the question of whether a content theorist could establish a lawful connection between J and *K* on purely empirical grounds. But nothing like this has ever been proposed or attempted. And given the nature of theories of content, we can't really expect empirical methods to be helpful. (Imagine trying to establish empirically that a thought's content involves PLUS just in case the explanation of the thinker's being primitively compelled to make certain transitions is that those transitions have the plus-form. Or that whenever all the other laws governing the tokening of a mental symbol asymmetrically depend on the horse law, the symbol means HORSE.)

As we saw in chapter 2, current theorists of content take a completely different approach. First, they propose a property that seems intuitively to be a candidate for meaning. Second, they make an *a priori* argument that the property would have certain features of content, most importantly that there would be room for a distinction between representation and misrepresentation. Thus, current theories of content stand or fall entirely on *a priori* arguments. Even Fodor, who is generally more sanguine than most philosophers about the idea that nomological necessity is all that philosophers should want,[30] is explicit that he does not claim that his asymmetric-dependence conditions are ever satisfied in the actual world (1990, p. 131). He thus makes no attempt to identify a correlation and show that it is lawful. The argument that anything that satisfies his conditions for

[30] *E.g.* Fodor (1994), pp. 25-29.

having a particular content will have that content[31] is entirely *a priori*. In sum, there is no serious prospect of demonstrating empirically that proposed conditions for content are nomologically sufficient for content.

One might wonder whether a theorist could use inference to the best explanation to argue for a lawful connection between J and *K*. A theorist could not infer to the best explanation of an empirical regularity since no such regularity has been identified. The theorist would need to argue that a lawful connection between J and *K* is the best explanation of the philosophical data. As argued above, however, a central part of the philosophical data is that certain normative properties are constitutive of content *K*. It is thus difficult to see how J's being nomologically sufficient for *K* could provide a serious candidate explanation of the philosophical data without there being some explanatory connection between J and the relevant normative features of *K*.

Let us turn to requirement 2). We have seen that a theorist will not be able to establish that condition J is necessary and sufficient in all nomologically possible worlds without establishing an explanatory connection between J and content *K*. The same arguments apply all the more strongly with respect to the question whether condition J is necessary and sufficient in all metaphysically possible worlds. If there is no serious prospect of relying on purely empirical methods when what is at issue is nomological possibility, things are that much worse in the case of metaphysical possibility.

Finally, requirement 3) explicitly requires an explanatory connection. Thus, regardless of which of the three requirements is the appropriate one, a theorist of content will have to establish an explanatory connection between condition J and content *K*.

As we saw above, an explanation of why a thought that satisfies J has content *K* has to be an explanation of why a thought that satisfies J has the correctness conditions constitutive of *K*. Thus, a theorist won't be able to support her theory of content without showing that there is an explanatory connection between her proposed condition and the appropriate normative features of content. In the next section, we will consider an achievement theorist's prospects for satisfying this requirement.

6.8 Normativity and the achievement view

An answer to the question of what it is to use a particular concept should offer facts that make it the

[31] See section 4.4 for some discussion of the thinness of Fodor's argument on this point.

case that the person *should* make certain judgments (or transitions), in the sense that it is *correct* for him to make those judgments. (If the proffered facts do not make it the case that the person should make the relevant judgments, the facts cannot constitute the person's using the concept in question.) The key difference between a person's meaning PLUS and a person's meaning QUUS, for example, is in the judgments involving the candidate concept that it is correct for him to make. This point is obscured by treating together the questions of what individuates a concept and of what it is to *have a concept, for the primary worry then tends to be what facts about a person could uniquely determine a particular concept.

Once we have answered the individuation question by specifying an inferential role, the remaining question is: what facts about a person would make it the case that he is using a concept with that inferential role, which is to say, would make it the case that it is *correct* for him to make the judgments that correspond to that inferential role? When the question is put in this way, an achievement-based theory seems a poor answer since it is not at all plausible in general that having the ability to make certain judgments makes it the case that a person should make those judgments. (Similarly, if we give a reference-based answer to the individuation question, it is not plausible that a person's having dispositions to apply a symbol to certain objects makes it the case the person should judge the symbol applicable to those objects. And the corresponding point holds for Fodor's version of informationalism.)

It might be responded that if the thinker's dispositions or abilities determine that the concept has a particular inferential role or a particular reference, it is *ipso facto* the case that the thinker should make the appropriate judgments. But this response (the *question-begging response*) manages to beg the question of what it is to use a particular concept by confusing it with the question of what individuates the concept. Granted that a particular inferential role or reference individuates a particular concept, what follows is of course that it is correct for a thinker who is using the concept to make judgments in accordance with that inferential role or reference. This truism presupposes *which* concept a thinker is using, however. To put it another way, the question is what makes it the case that the thinker's thought involves a certain concept, the one with a particular inferential role or reference. Since it follows immediately from something's being the inferential role or the reference of the concept a thinker is using that the thinker should make certain judgments, the difficulty in giving an account of what makes it the case that a thinker is using *this* concept is to specify facts about the thinker that make it correct for him to make the relevant judgments.

Given this situation, it is crucial that a theory of content explain why a thinker *should* make the judgments that correspond to the concept the account attributes to him. But current achievement-based theories do not do so. We've already seen that Fodor, though he expresses some uneasiness about not addressing the normativity issue, gives the question-begging response. Peacocke addresses the normativity question by appealing to his theory's way of satisfying the requirement that a concept (together with the world) must determine its reference. His theory specifies roughly that the semantic value of a concept is whatever semantic value would make the inferences and non-inferential judgments specified in the concept's possession condition truth-preserving and true, respectively (1992, pp. 18-19). Peacocke points out in effect that, given this specification, the belief-forming practices that are specified in the concept's possession condition will be correct; thus, he maintains, his theory accounts for their normative status (1992, p. 138-140).

This point, however, cannot be intended as a response to the normative demand I have explicated, for it would constitute a version of the question-begging response. Given how Peacocke's theory specifies that references are determined from a concept's possession-condition, it follows that *if* a thinker is using a particular concept, it is truth-preserving for him to make the inferences that correspond to the concept's inferential role. And it is true that the fact that an inference is truth-preserving respects the normative demand of the level of reference. But none of this goes to the question of whether the thinker *is* using the concept in question – whether he is subject to its normative demands. Peacocke's points assume that the thinker is using the concept in question, that he is subject to its requirements. Peacocke cannot, of course, answer that his theory holds that a thinker is using a particular concept just in case he satisfies such and such conditions specified by Peacocke's theory. For the question is precisely whether satisfying those conditions makes it the case that a thinker is using the concept. It is in order to defend the conditions his theory specifies that Peacocke must explain why a person's satisfying them makes it correct for him to make the relevant judgments.

In sum, both Fodor and Peacocke emphasize the point that if their theories get a concept's semantic value right (and in Peacocke's case, in a way that makes the canonical inferences truth preserving), normativity is taken care of. Neither makes any attempt to explain, *without assuming that his theory is right about which concept a given thinker is using*, why a thinker who has the dispositions or abilities specified by their respective theories *should* make the relevant judgments rather than others. To say that a thinker should make a given judgment because it applies a concept to an object in the concept's extension or because the judgment accords with the concept's inferential role is to appeal to the content of thought to explain normativity, rather than to show that the theory has the resources to explain content's normative features.

This discussion suggests that the facts about a thinker that make it the case that he is using a particular concept must do so in virtue of their coming under a description other than that of the thinker's having a particular type of disposition or ability to token representations, draw inferences, or, generally, make judgments. For it is not plausible that it is correct for someone to make certain judgments *in virtue of* his having a particular disposition or ability to make judgments. That is not to say that the appropriate resources for an account of what it is to use a particular concept do not include a thinker's symbol-tokening dispositions and inferential abilities, but that we should not expect a type-type identity between using a particular concept and exercising a particular symbol-tokening disposition or inferential ability.

The discussion of deference-dependent attitudes should already have made this point clear. Deference-dependent attitudes are realized by very different sorts of symbol-tokening dispositions and inferential abilities from attitudes involving fully grasped concepts. A good account of what it is to use a concept will say in virtue of what higher-order property the dispositions or abilities of the thinker with full understanding and the dispositions or abilities of the thinker with incomplete understanding (who defers to others) both realize the state of having an attitude involving the same concept. The point that type-type identity is not necessary, nor even to be expected, removes any reason for thinking that, whatever its problems, an achievement view is the only conceivable option. For we are not seeking a way to specify, for each concept, a symbol-tokening disposition or inferential ability the exercise of which is *having the concept.

Let us take stock of the position. When we consider the question of what it is to *have a particular concept, an achievement view looks implausible, for it fails to satisfy the desideratum that it explain why the thinker *should* make the appropriate judgments. Moreover, since we do not need a type-type identity of using a concept and exercising a disposition or inferential ability, an achievement view is not forced upon us. In the next chapter, I explore the responsibility view's potential to answer the content question.

Before doing so, however, we need to turn to my argument for IIUCs. I present the argument at this stage for several reasons. First, the arguments against the achievement view and the argument for IIUCs are mutually supporting. Since the achievement view is a major reason for thinking that there can be no IIUCs, the reasons I have given for doubting the necessity and plausibility of the achievement view should make the idea of IIUCs less problematic. Conversely, if the arguments for IIUCs are right, DV and, *a fortiori*, the achievement view are untenable. Second, the examples of IIUCs are an important category of data that, in addition to deference-dependent attitudes, an account of what it is to *have a concept must account for. Third, relatedly, the examples of IIUCs will be suggestive of how to go about giving an account of what determines that a thought involves a particular concept.

6.9 Independent Incompletely Understood Contents (IIUCs)

We have seen that the achievement view, which is the form that DV takes in standard accounts of content, has difficulty making sense of the role of deference. We turn in this section to Independent Incompletely Understood Contents (IIUCs), which directly challenge DV. The section can thus be seen as the second stage of a two-stage attack. First, leading theories of content cannot give a satisfactory account even of the cases of incomplete understanding that they do acknowledge. Second, worse, there are other cases (IIUCs) that they cannot begin to address.

Let us begin by asking why, aside from the achievement view, it should be thought that deference is necessary for a thinker with incomplete grasp of a concept to have attitudes involving it. Here is Peacocke's argument for his claim that deference is "essential" in his "attribution conditions": "[i]t is what distinguishes the case we are interested in, partial understanding (and partial misunderstanding) of a word in a communal language, from the quite different case of an individual's taking over a word from his community and using it in his own individual, different sense" (p. 29). It is certainly true that deference is what serves this function in the case of deference-dependent attributions, but that does not settle the question of whether anything other than deference could enable someone to have a thought involving a concept she incompletely grasps.

Let us ask at a more general level what function deference serves. Something is needed to distinguish the case of a thinker's using a particular concept that she incompletely grasps from her using a different concept she fully grasps (and from her using other concepts she incompletely grasps). A natural idea is that deference to others who fully grasp the concept in question can determine that the thinker is using that concept. The initial plausibility of DV is explained by the difficulty, if no one in the community fully grasps the concept, of solving the problem of what could make it the case that a person's attitude involves *that* concept. If DV is false, it must be possible for something to determine the content of such an attitude.

The responsibility view has the potential to show the way to a solution. If what it is to use a concept (in a thought) is to be responsible (with respect to that thought) to the standards that concept imposes, thinkers who do not fulfill those standards may be able to use the concept, for people can be responsible to standards they do not fulfill. Deference to a thinker who satisfies a concept's standards is one good way of making oneself responsible to those standards. But there is no reason for thinking it is the only way for a thinker to become responsible to conceptual standards he does not fulfill. Before we consider the problem further, let us look at reasons for thinking that thinkers can have attitudes involving a concept that no one fully grasps.

I will use examples to suggest that DV is not true as a claim about our ordinary practices of content attribution. Each example will be of a thinker who does not have full grasp of some concept, yet who, we would want to say, has thoughts involving the concept that are not dependent on deference. Of necessity, the examples appeal to our intuitions about content attribution since the point is to test DV against those intuitions.

My plan is to sketch a broad range of examples in order to suggest that putative cases of incomplete understanding are a pervasive and familiar phenomenon. I will first concisely describe several classes of examples, without addressing objections. I will then consider briefly a variety of objections and replies to objections, showing that the merits of different examples are controversial enough that, at least, we cannot take DV for granted. Given our working hypothesis that an inferential-role theory is the best answer to the question of what individuates a concept (section 6.6 above), the discussion takes as its primary target an inferential-role theory that accepts the achievement view; the examples are selected to show that such a theory cannot account for IIUCs. Peacocke's theory often serves as a representative of such a view because it is based on a well-articulated inferential-role theory of concept individuation. The details of arguments and counterarguments with respect to each example, though perhaps important in their own right, are somewhat beside the main point. No doubt philosophical ingenuity can provide some way of reinterpreting any particular example, but if such objections are to vindicate DV, they must be driven by a coherent account of our intuitions. To the extent that the objections appear to be *ad hoc*, there is reason to suspect that what lies behind them is merely the assumption that DV must be retained come what may. The process of considering the objections and replies will, I believe, support the view that the objections are piecemeal reactions to the examples. Finally, I will suggest that we have at least a very abstract idea of the nature of content attribution according to which we would not expect it to respect DV. Chapter 7 discusses the prospects for developing this idea.

Consider first *non-dependent* examples, cases in which, though there may be experts who fully grasp a concept, other people's use of the concept is not dependent on the experts' mastery. It is plausible in many cases that thinkers who incompletely grasp a concept do not in fact defer to the experts or, at least, could use the concept even if they did not so defer. For example, philosophy students often apparently use a concept they do not fully grasp before they know that others have come up with the concept (and given it linguistic expression), and thus before the students defer in its use to anyone. The students seemingly have insights, make arguments, or draw distinctions that require the concept, but their attempted explications of the concept do not differentiate it from other concepts and manifest confusion. Examples are the concepts of supervenience, of necessity, and of a type/token distinction.

That the students would recognize and adopt a proffered correct account of the concept does not show that they defer to others. For a thinker to defer to another is for the thinker to treat her as an *authority*, relying on her to determine the content of his attitudes, not merely for him to be willing to accept a correct explanation of the concept from her. (It is only to be expected that a thinker who is using a concept that he has not mastered would often be willing to accept a correct explanation of it, even if the person who offers it is not one on whom he relies to determine the content of his attitudes.)

Before most judges had encountered economic theory, judges arguably used economic concepts like EXPECTED UTILITY, OPPORTUNITY COST, and EXTERNALITY in their reasoning, though they plausibly lacked full grasp of those concepts and would not have known that there were experts to whom they could defer. Similarly, many informally trained but skilled musicians gropingly rely on concepts like KEY and IN TUNE and may not defer to others because they do not realize their own grasp is incomplete or because there are no experts in their community.

A more radical possibility is that there may be many concepts – perhaps EXPLANATION, JUSTICE, and CONSCIOUSNESS are examples – that contemporary thinkers use but no one fully grasps. In many of these non-dependent cases, the thinkers may lack a tacit understanding of what falls under the concept and thus their abilities may be insufficient to differentiate the concept even from concepts with different references. They therefore have incomplete grasp of the relevant concepts even by the standards of reference-based views of concept individuation.

Since it is easier to identify cases of incomplete understanding in retrospect, after we know what full understanding involves, let us turn to the second class of cases – *historical* examples. These involve thinkers who used sophisticated theoretical concepts before anyone developed the theory necessary to understand them fully. A full defense would require detailed discussion of historical sources, but I will merely try to show how plausible it is that such cases can occur. Since Ancient Greece, thinkers arguably have used the concept of continuity – as in the continuous ordering of points on a line or of the real numbers – though no one could provide a rigorous account of the concept until Dedekind did so in the late nineteenth century.[32] Indeed, long before Dedekind's work, Newton and Leibniz used the concept, or related, poorly grasped concepts like that of an infinitesimal differential, in developing the calculus. The concept of entropy has a similar history. Like the philosophy students or judges, historical thinkers make arguments, draw conclusions, and engage in activities that suggest they are using the concepts in question, while their confused attempts at precise definition or explanation convince us that they incompletely grasp those concepts. (Again, it should be obvious that even by the standards of reference-based views of concept individuation, the thinkers may not be able to differentiate the concept from other concepts.) A concept is a good candidate for this sort of example if it is natural

[32] For a concise summary of the relevant history, see Bynum (1983), pp. 78-79, also pp. 48-49. I am grateful to Christopher Peacocke for calling this example to my attention.

and useful to someone engaged in a particular intellectual enterprise, yet can be fully explained only in the context of a theory. Potential examples can be found in diverse areas of theory: MARGINAL PRODUCTIVITY in economics; INTENSION in logic; COMPLEX NUMBER or LIMIT in mathematics; RES JUDICATA, BONA FIDE PURCHASER or CONSIDERATION (in contract doctrine) in law. (For that matter, pre-scientific societies lacking even systems of arithmetical notation probably use the concept PLUS long before their understanding of it involves a notion of counting that extends to infinitely many natural numbers.)

Such concepts are plausible candidates for incomplete understanding because, regardless of how sophisticated and reflective the innovative thinkers are, without the surrounding theory it may not be possible for them fully to grasp the concepts in question. Their reflectiveness about their enterprise and a concept's aptness for that enterprise can make it intuitively attractive to attribute to them thoughts involving the concept, despite their lack of full understanding. A concept may be so fitted to a certain stage of an argument, so needed for a certain progression of thought – highly structured fields like mathematics, music, and language may be the clearest source of such cases – that the facts under investigation themselves support imposition of the standards imposed by the concept. If a reflective thinker leaves no doubt that, as part of her pursuit of the truth about that subject matter, she is trying to work her way through the relevant course of reasoning, it can be difficult not to attribute to her thoughts involving the concept. We might want to say that a thinker who commits herself to seeking the truth in such a case thereby makes herself responsible to the standards imposed by the concept.

A third class of cases – *natural-kind* examples – involves the attribution to people in pre-scientific societies of beliefs that, at the level of reference, are about natural kinds like water and gold. For example, we would say that a Mycenaean could have had the thought that gold sinks in water. Some may be unsure whether, in this context, "gold" and "water" express natural-kind concepts, but it is plausible that the Mycenaeans had *some* way of thinking of the properties that are the *references* of our concepts of gold and water. But, as I discuss below, of the potentially available concepts, only natural-kind concepts have the right reference. So, *if* our practices of content attribution respect DV, these attributions to pre-scientific people presuppose that it is an empirical fact that the people had quite specific sorts of intentions or dispositions – those that (depending on whether we take an inferential-role or a reference-based view of concept individuation) are necessary for mastery of a natural-kind concept. Our view that the Mycenaeans could have had beliefs about gold or water is more secure, however, than any views we have about their intentions or dispositions – their level of understanding. Our attribution of such beliefs is better explained by the view that ordinary attributions of content do not respect DV than by the view that they rely on detailed and unsupported empirical suppositions.

Burge (1986b) sketches a way of constructing hypothetical examples that, I will argue, can be used to support the claim that in general, for any plausible possession condition for a given concept, someone could have attitudes involving the concept without satisfying it. Burge's examples, unlike those we have so far considered, do not involve thinkers who lack sophistication or theory. To the contrary, Burge's idea is that one who understands a term and knows its ordinary use could develop a "nonstandard" theory about the objects to which it applies, thereby coming to doubt even the characterizations most central to giving the meaning of the term (1986b, pp. 707-710). For example, someone who fully understands the term "sofa" and is fluent in its use could propose as a testable, empirical hypothesis, and could even come to believe, that sofas are not items of furniture made for sitting on, but religious artifacts or works of art. Burge points out that "the argument can be adapted to any substantive notion that applies to physical objects, events, stuffs, properties" (1986b, p. 709).

Burge's idea is relevant in the present context because believing a Burgean nonstandard theory can prevent a thinker from satisfying a possession condition she once satisfied or, more generally, fully grasping a concept she once grasped. One who is primitively compelled to infer that Q must at least find it natural to believe that Q, if not actually believe that Q.[33] So one who satisfies a Peacockean possession condition must find it natural to believe certain things. But Burge's examples suggest that a thinker who uses a given concept need not believe, indeed can believe to be false, any particular proposition, and there is no reason why the examples need allow that the thinker who believes a proposition is false nevertheless must find it natural to believe it is true. In general, believing a nonstandard theory can interfere with a thinker's instantiating a concept's inferential role and can interfere with a thinker's having symbol-tokening dispositions that correspond to a concept's reference.

[33] If it were not the case that finding something primitively compelling were being disposed to judge it true, but something along the lines of believing that it is conventionally judged true, then that notion would not be a plausible candidate for the work for which Peacocke intends it. The notion is supposed to capture the idea of how we naturally go on, beyond the point where reasons can be given (pp. 13, 190-191). Also, if Peacocke's account is not to be circular, a thinker's finding something primitively compelling must be a state of which a subrational account can be given (pp. 179-190).

Moreover, the thinker who believes a nonstandard theory obviously does not defer with respect to the relevant concept to other members of the community who believe the standard theory, for she knows how they apply the concept and believes they are wrong. Thus, deference-dependence cannot account for the examples. Indeed, although Burge does not address the issue explicitly, the argument does not depend on the assumption that other members of the thinker's community continue to believe the proposition in question.

So if Burge is right that one who is a competent user of a concept term can doubt the characterizations that are supposed to state its meaning, something must give: either those characterizations cannot be incorporated in the possession condition for the concept or it cannot be maintained that a competent user of a concept must satisfy its possession condition. Thus, if a competent user of the concept SOFA can doubt that sofas are pieces of furniture made for sitting on, then, given DV, the possession condition for the concept SOFA cannot require that a thinker be primitively disposed to infer from the premise that X is a sofa that X is a piece of furniture made for sitting on.

Consider, to take a more interesting example, Peacocke's suggested inferential requirement for natural-kind concepts that, for a given concept *F*, one must be primitively compelled to infer from the premise that X is F that X is of the same underlying kind as other samples of F (1992, p. 26). A fully competent user of the natural-kind concept WATER, one who, for example, has consciously intended to use that natural-kind concept in thinking about water, could coherently form the hypothesis that water is not a single natural kind, but two or more different kinds of substances. If she did so, then she would not be primitively compelled to infer from the premise that X is water that X is of the same underlying kind as other samples of water. It might be said that, in this case, the thinker would no longer be using the natural-kind concept WATER. Yet, intuitively, we would not want to say that her forming (or coming to believe) her hypothesis instantly worked a change in meaning.

We must now turn to a range of objections to the examples. Short of simply denying the intuitions, there are two general strategies of objection. First, an objector can claim that the possession condition of the concept in question is weaker than an example assumes, so that the thinker satisfies it. Second, an objector can reinterpret an attribution of content, arguing that what is attributed is not in fact an attitude involving the concept in question. Both of these strategies can take a variety of forms in different cases.

The most natural objections to the Burgean examples are those of the first type, so let us consider them first. Peacocke could argue that Burge's examples show only that particular inferences are not constitutive of the concept in question. Burge's argument, however, points in the direction of a general reason for thinking that no inference will necessarily be compelling for any competent user of a concept. Burge's examples suggest that one who is a competent user of a concept could be wrong about any proposition involving the concept, regardless of how central it is to the concept's identity. There seems no reason why the method of developing a nonstandard theory according to which a supposedly meaning-giving proposition is false cannot be generalized indefinitely.

Moreover, there is an important constraint on any objection of the first strategy: a concept's constitutive inferential role must include inferences that are sufficient uniquely to individuate the concept in question. As a result, there are two opposing pressures on an inferential-role account like Peacocke's that combines an inferential-role theory of concept individuation with the achievement view. On the one hand, Burge's examples suggest that if an inference is to be found that any competent user must find compelling, it will have to be a very limited inference, one even more uncontroversial than the inference from X's being a sofa to X's being made for sitting on. On the other hand, the inference must be strong enough uniquely to individuate the concept in question. Mastery of an ordinary empirical concept, for example, cannot be simply an ability to pick out (some of) the objects or substances to which the concept applies. Such an ability is not sufficient uniquely to individuate a concept since there can be different concepts with the same reference.

Of course, Peacocke can accept that we may be seriously wrong about what is constitutive of our concepts. There is an important limit, however, on the extent to which an inferential-role theory that accepts the achievement view can allow that we may be wrong: ordinarily competent users of a concept (those who are competent enough that they can, without reliance on others who are more expert, have thoughts involving the concept) must instantiate the concept's inferential role. But Burge's examples provide reason to believe not just that a competent user of a concept could fail to be able explicitly to formulate what constitutes the concept but that she could explicitly hold incorrect views about propositions central to the concept. If a competent user of the concept *X* could be utterly wrong about what Xs are in any given respect, then it is difficult to see how it can be maintained that a competent user necessarily instantiates an inferential role that is strong enough uniquely to individuate the concept.

A possible response for an inferential-role theorist who holds the achievement view would be to propose that concepts can have disjunctive inferential roles; in other words, that in order to have the inferential ability that corresponds to a concept's inferential role, a thinker must make only one of several inferences. This response would account for the way in which any particular inference proposed as part of a concept's inferential role seems subject to challenge, but there is no reason to think that nonstandard theory cannot be developed to challenge an indefinite number of inferences. Moreover, it is not intuitive that such disjunctive inferential roles correctly individuate ordinary empirical concepts. Most importantly, the inferential-role theorist would then lack an account of what distinguishes a thinker's using such a concept, for which she instantiates only, say, one of the disjunctive inferences, from her using a different concept whose inferential role includes only that inference (or that inference and some other, non-overlapping inferences). To put it another way, what would make it the case that thinkers who instantiated different disjuncts of the putative disjunctive inferential role were using the same concept?

Let us now turn to the second type of objection, which involves reinterpreting the ostensible attribution of content. An objector may claim that ordinary language has vague and shifting standards for using sentences that apparently attribute attitude contents, with different standards in different contexts and for different purposes. The more the subject is distant in time, place, or culture, the greater the tolerance in the use of attitude-attribution sentences. Therefore, the objection runs, we cannot put much weight on the fact that people would unreflectively say that, for example, the Mycenaeans believed that gold sinks in water or that a judge believed that the expected utility of taking certain precautions was less than the cost. A word like "gold" in such attributions should not be taken as expressing a particular concept in all contexts; in some contexts, the word may be merely a vague and convenient way of indicating that the subject has some way of thinking of the things to which the concept in question applies.

But the objection cannot be that the content attribution uses the concept word to indicate that the subject uses a different, fully grasped concept with the *same reference* as the relevant incompletely grasped concept. In both the natural-kind and historical examples, the reference of whatever concept the subject fully grasps is different from the reference of the incompletely grasped concept. For instance, our inclination to use the term "water" in attributing thoughts to the Mycenaeans cannot be taken as merely a convenient way of indicating that they had some way of thinking of H_2O, for if they did not fully grasp the natural-kind concept WATER, the relevant concept that they fully grasped almost certainly did not have the extension H_2O. At least sometimes, it is the attribution of the same concept that explains attribution of the same reference (rather than sameness of reference's accounting for the use of the same word).

It might be objected that, in the historical examples, an attribution ostensibly involving a given theoretical concept should be interpreted as an attribution involving a way of thinking of something as, roughly, *that which falls under the concept that would play the appropriate role in the correct theory*. This *placeholder* concept would necessarily have the same semantic value as the theoretical concept for which it stands in, and the thinker could fully grasp it.

It is important that the placeholder concept must take this second-order form, rather than being a concept such as *whatever is the nature of this phenomenon* (where the phenomenon is identified by examples). As we have seen, the thinker may have no (first-order) way of identifying the reference of the concept. A few examples, perhaps including some mistakes, cannot suffice to determine the concept's reference. The point of the placeholder-concept proposal is that, by going second order, the thinker might be able to refer to the semantic value of the theoretical concept indirectly, via reference to the theoretical concept itself.

A placeholder concept does not capture the intuitive force of the examples, however. The sorts of things the historical thinkers would say suggest that they are not just using empty placeholders for theoretical concepts. A thinker may use the concept in question – *C* – to give reasons in substantive arguments; judging that X is C commits her to conclusions beyond the proposition that X falls under the concept that has such and such theoretical role.

Moreover, in order for a thinker fully to grasp the placeholder concept, she would need fully to grasp the notion of (roughly) the concept that would play the appropriate role in the correct theory.

And, given that she fully grasps the placeholder concept, in order for it to be plausible that she uses it she would need to have second-order attitudes toward the theoretical concept (*e.g.*, the belief that the concept plays the appropriate role in the correct theory) and awareness of her lack of understanding of it (otherwise, she would not substitute the placeholder concept for it). There is no reason to assume that thinkers fulfill these conditions.

It is worth emphasizing the point that a thinker may not be aware that she does not understand the theoretical concept in question. In that case, not only will she not use a placeholder concept, she will not use a first-order concept of the *whatever-the-nature-of-this-phenomenon* sort.

Once we reject the suggestion that the thinker is using a different concept with the same semantic value as the incompletely grasped concept, there are a variety of less attractive claims open to the objector. The objector may claim that the ostensible attribution of content employs the relevant concept word to indicate that the subject: 1) uses a particular concept that applies to only some of the same objects as the incompletely grasped concept; 2) has a *de re* thought about an object to which the incompletely grasped concept applies, or about a property that is possessed by some of the objects to which the concept applies (*e.g.*, in the case of WATER, the property of being a clear, colorless, odorless, potable liquid or, in the case of SOFA, the property of appearing a certain way and being able to be sat upon);[34] or 3) has some way or other of thinking of some objects to which the concept applies.

Though these suggestions should make us reluctant to place too much weight on any given example of ordinary practice, they do not provide a satisfactory overall account of the examples. First, attributing to a thinker thoughts such that, even at the level of reference, she is not talking about the same thing as we are, is implausible in many cases. It implies that the thinker would have to effect a change in meaning merely in order to share a common subject matter with, and therefore in order to agree or disagree with, our beliefs involving the concept in question. For example, the suggestion that attributions of gold-thoughts to pre-scientific people use the word "gold" to express a functional-kind or recognitional concept would imply that those people had no thoughts about gold as opposed to about particular samples of it; for a functional-kind or recognitional concept would not refer to gold (the natural-kind substance).[35]

[34] Chris Peacocke brought to my attention the objection that relies on a thought that is *de re* of a property.
[35] Further, if we refuse to attribute thoughts involving incompletely grasped concepts, particular suggestions encounter their own problems: for instance, even if the Mycenaeans incompletely understood the notion of an underlying kind, they need not have believed that anything that had gold's function or appearance had to be gold. So we cannot explain our confidence in our content attribution by saying that, in this context, "gold" is

Second, ordinary practice – folk psychology – is the home of content attribution. I have claimed that ordinary practice frequently countenances (non-deference-dependent) attributions that apparently mention concepts the subject incompletely understands. In order to maintain DV, we would have to assume that, in these attributions, the concept word is not being used to express the concept or even, in many cases, to refer to its semantic value. That assumption should not be made without strong grounds.

To the contrary, however, there is reason to expect that folk psychology would license such attributions of contents involving incompletely grasped concepts. An important purpose of content attribution may be to indicate the standards to which thinkers are responsible. If the responsibility view is correct, that purpose derives from the constitutive truth that what it is to have an attitude involving a given concept is to be subject to certain standards, not to exercise a particular inferential ability. People with different levels of understanding can be responsible to common standards that some of them fail to meet, and fail to meet to differing degrees. (As discussed in chapter 7, however, it is consistent with the responsibility view that thinkers may typically need to have some degree of ability to satisfy a concept's standards in order to have thoughts involving it.) Thinkers might be responsible to such standards in virtue, say, of a common enterprise of seeking the truth about some field of inquiry. Thus, the purpose of indicating the standards to which thinkers are subject may be sufficient to explain why it is appropriate in ordinary practice to group together, as having thoughts involving a common concept, people with full understanding and people with diverse incomplete understandings of that concept: such attributions indicate that the subjects are considered responsible to certain common intellectual standards, perhaps because they are engaged in the same enterprise or are addressing the same overarching question. So we have at least an abstract and general justification according to which we need not expect the examples to satisfy DV. In the next chapter, we will explore the potential of the responsibility view to provide a detailed explanatory account of content.

I have not attempted to provide a definitive response to the objections to the examples. As I said above, the question is not whether we can devise ways of rejecting alleged counterexamples to DV one by one, but whether their rejection is motivated by the most plausible overall account of the phenomenon of incomplete understanding. My goal has been to suggest that the objections do not

---

used to express a functional-kind or recognitional concept that we can be confident the Mycenaeans fully grasped.

reflect a compelling account of our content-attribution practices that supports DV. At very least, the examples show that we cannot simply take DV for granted.

We have considered some affirmative reasons for doubting DV. As discussed above, DV is difficult to give up because if no thinkers, even collectively, fully grasp a concept, it is not easy to see what could determine that an attitude involves *that* concept, rather than an indefinite number of others. As we saw in chapters 3 through 5, however, even *given* DV, we lack a satisfactory answer to the content question. Thus, rejecting DV would not spoil a satisfactory account of content. On the other hand, rejecting DV may make the problem look harder, for if DV were false, we would need an account of what determines that a thinker uses a concept that neither her abilities nor those of anyone else in her community are sufficient to differentiate from an indefinite number of other concepts. In other arenas, however, people are held to standards they do not have the ability to fulfill, and not because they defer to others who do have that ability. Merely taking part in certain activities or enterprises is enough to make drivers, athletes, physicians, and generals responsible to standards that they may lack the skill to fulfill.

According to a simple but recognizable picture, competent speakers of a language fully grasp most everyday concepts. Learners defer to competent speakers, who include most native, adult speakers. The competent speakers defer to experts with respect to the use of scientific, technical, or otherwise specialized concepts, and experts fully grasp the concepts in their areas of expertise. Although competent speakers may in fact defer to each other in the use of everyday words, what determines the content of their attitudes is their own mastery of concepts, supplemented by deference to experts with respect to the use of technical terms. Any concept that figures in the attitudes of members of a linguistic community must be fully grasped by some members of the community individually.

This simple picture is rarely argued for, but frequently taken for granted. Perhaps our attachment to this picture is what makes it so difficult to imagine how, without DV, to go about giving an account of content. Since the picture has not yielded a satisfactory account of content, we should ask whether there are other, unrecognized determinants of content or, perhaps, unrecognized ways in which accepted determinants contribute to content. The simple picture and DV may have blinded us to the need to look for something other than a type-type identity between, on the one hand, *having a concept and, on the other, dispositions or abilities to token representations, make judgments, or draw inferences.

In the next chapter, I consider the possibility opened up by the responsibility view that facts may play a role in determining content, not only by determining reference, but by bearing on the question of which ways of thinking of objects thinkers should be held responsible to. So, for example, the historical examples suggest that if a thinker is pursuing mathematical truth, mathematical facts may play some part in determining to which concept's standards the thinker should be held responsible. This idea might provide the starting point for an account of what determines that a thinker is using PLUS, rather than QUUS. Perhaps physical, ethical, and other kinds of facts could play roles similar to those of mathematical facts.

Let us review where we are. We found that once we consider the question of what it is to use a concept independently of the question of concept individuation, the achievement view is not an attractive answer to the former question. Given the *prima facie* attractiveness of an inferential-role based theory of concept individuation (when we consider the question of concept individuation without regard to the question of what it is to use a concept), we decided on the strategy of taking an inferential-role theory of concept individuation as a working hypothesis and considering whether it can be complemented with a non-achievement-based theory of what it is to use a concept. In particular, the latter theory must provide a way of specifying, for each concept, facts about a thinker that make it correct for the thinker to make the judgments that correspond to the inferential role of that concept. Before considering the responsibility view of what it is to use a concept, we paused to consider section 6.9's argument for IIUCs. IIUCs provide further reason for rejecting an achievement view. Let us now turn to the responsibility view.

Chapter 7: The Responsibility View

7.1 Introduction

This chapter begins, in sections 7.2 and 7.3, with clarification of the responsibility view. In section 7.4, I turn to a distinction – between being subject to a standard and being disposed to satisfy it – that is crucial to the responsibility view's advantages over the achievement view. Next, in section 7.5, I review those advantages and respond to a number of objections. Finally (sections 7.6 and 7.7), I turn to the central challenge for a responsibility-based theory of content: developing an account of what makes thinkers subject to the relevant standards. I do not attempt to meet that challenge here, but merely sketch some programmatic ideas.

7.2 The basic idea

Let us review the responsibility view as a general account of what it is to use a concept.

> **The responsibility view**: a thought involves one concept rather than others in virtue of the thinker's being responsible or subject to standards distinctively imposed by that concept.

I want to begin by commenting relatively briefly on several aspects of this formulation. Many of the points will receive further development later in the chapter.

First, the responsibility view, like the achievement view, is offered as a constitutive account of content. The claim is not merely that a person who has a thought involving a particular concept is necessarily subject to certain standards. Rather, the person's being subject to those standards is what it is for him to be using that concept as opposed to others. (Or, at least, it is part of what it is for him to be using that concept. A moderate version of the responsibility view could hold that there are other aspects of what it is to use a concept.) The point I want to emphasize concerns the direction of explanation: the responsibility view appeals to a thinker's being subject to standards to explain a thought's involving a particular concept.

Second, what are the standards in question? As we have seen, achievement-based theories hold, implicitly or explicitly, that concepts are canonically associated with conditions on thinkers. A responsibility-based theory can treat a concept's associated condition as the relevant standard. On the achievement view, what makes it the case that one uses a certain concept is that one exercises a disposition to satisfy the concept's associated condition. On the responsibility view, what makes it the case that one uses the concept is that one *should* satisfy the concept's associated condition, not that one exercises a disposition to do so. As we will see, this does not mean that one's dispositions are irrelevant. Rather, one's dispositions are relevant *because* (and to the extent that) they play a role in determining what standards one is subject to.

The immediate point is that a responsibility-based theory's postulation of conditions canonically associated with concepts need not commit the theory to anything beyond what standard theories of content are committed to. (The issue on which a responsibility-based theory and an achievement-based theory must disagree is whether the conditions are standards to which thinkers are responsible.) A responsibility-based theory has the same range of options as an achievement-based theory with respect to views about the nature of concepts. For example, on Peacocke's achievement-based theory of content, each concept's possession condition specifies mental transitions that are canonical for that concept. Correspondingly, a responsibility-based theory could hold that for a thinker to use a particular concept is for the thinker to be subject to a standard requiring the thinker, if the relevant questions arise, to make the transitions specified in the concept's possession condition.

As we saw in section 6.7, many theorists would accept that a person who has a thought involving a concept is necessarily responsible to certain corresponding standards. Such theorists assume, however, that this responsibility to standards is somehow a consequence of what it is for a thought to involve a concept, a consequence that, I suggested, standard theories are not well-positioned to explain. By contrast, the responsibility view holds that it is the person's being responsible to the concept's standards that accounts for the person's having a thought involving the concept, rather than the other way around. Again, however, the responsibility view is not thereby committed to holding that concepts are associated with conditions (or, as the responsibility view understands them, standards) that go beyond what is postulated by achievement-based theories.

Third, what is it to be responsible to, or subject to, a standard? To be responsible to a standard is merely for it to be the case that the standard applies to, or governs, one's conduct, so that one's conduct is incorrect or wrong if it does not meet the standard.[1] The notion of being responsible

[1] As noted in section 2.9, it is important to distinguish this use of the term "responsible" from the use in which it means accountable, *e.g.*, blameworthy. An agent can act incorrectly – in violation of a standard – yet not be blameworthy for having done so.

to a standard is an extremely familiar one. (Arguably, not just agents are subject to standards, but, also, for example, actions, thoughts, artifacts, and biological organs. For the most part, I will be concerned with agents.) Legal examples are perhaps the most straightforward ones. In general, being an adult person physically present in a jurisdiction is sufficient to make one subject to its criminal laws. Dogs and infants are not subject to criminal laws, though their owners and parents are. Foreign diplomats notoriously are not subject to many of the local traffic and parking regulations. To take a very different kind of example, I am subject to rational and moral requirements; storms and viruses are not.

As the foregoing examples suggest, there is a generic notion of being subject to a standard or requirement, a notion that cuts across diverse domains. One can be subject to rational, epistemic, intellectual, aesthetic, moral, customary, and legal standards, and to requirements of particular organizations, communities, and activities. Thus, when the responsibility view holds that for a thinker to have a thought involving a particular concept is for the thinker to be responsible to certain standards, it does not rely on a notion of being responsible to standards that is special to the theory of concepts. The notion of being responsible to a standard is a fundamental normative notion, on a par with the very notion of a norm or standard. In order to understand what a norm is, one must understand what it is to be subject to a norm. Indeed, it seems that part of what it is to be a norm is to be the sort of condition to which agents (or other entities) can be subject. Since the notion of being responsible to a standard is familiar, very basic, and not special to the theory of content, I will not here try to say more about that notion itself.

Fourth, as noted above (section 2.1.2), a qualification is needed to the effect that the thinker is subject to the relevant standards *with respect to the thought in question*. Roughly, her mental activity involving that thought is subject to the relevant standards. (Also, what the standards require will depend on the position of the concept in question in the structure of the thought's content. To spell this idea out, each propositional content is associated with standards that derive from the standards of its component concepts, appropriately combined. For a thought to have a particular content is for the thinker to be subject to the standards of that content.) I will generally omit this qualification.

Fifth, a different question is what makes people responsible to standards. Just as we have a notion of responsibility to standards that is not domain specific, our understanding of what makes people responsible to standards cuts across domains. We have a great deal of familiarity with the kinds of factors that make people responsible to standards, though it is often difficult to articulate a precise account. Explicit commitments, such as promises and oaths, are a way of becoming responsible to standards. Participating in an enterprise or being a member of a community can be enough to make one responsible to its standards. On some views, merely having certain capacities is enough to make one responsible to moral and rational standards. One's position or circumstances can be at least an important factor in making one subject to norms. For example, on some views one can have moral obligations merely by being someone's son or being in position to rescue someone in urgent need. It is not peculiarly the job of a theory of content to say what makes people responsible to standards. A responsibility-based theory of content can appeal to our understanding of what makes people responsible to standards in general and apply that understanding to the case of concepts.

This is a good place to clarify the relation between the responsibility view and certain theories of content that hold that a thought's content is determined by biological function.[2] (Depending on the view, the relevant biological function can be that of the thought, the mechanism that produces the thought, or the mechanism that uses it.) The theories in question maintain that having a biological function is a normative notion: to have a certain biological function is to be subject to certain standards.[3] For example, on this understanding, to say that an immune system's function is to prevent disease is to say that an immune system should prevent disease or that a good immune system prevents disease. Such *teleological* theories of content typically hold that an organ or organ system has its function in virtue of the creature's evolutionary history. (It is also possible to give an account of biological function that does not appeal to evolutionary history, but, for example, to what the organism needs to flourish.)

The present point is that a teleological theory of content is a version of the responsibility view that takes a specific position on what makes thinkers responsible to the relevant standards. Teleological theories hold that evolutionary history is what makes thinkers responsible to conceptual

---

[2] See Ruth Millikan (1984; 1990) and Karen Neander (1991; 1995). I mention a different appeal to biology in chapter 3, note 18.

[3] Another possibility is to hold that a thought has its content in virtue of its biological function, but to deny that biological function is a normative notion. a theorist could accept that a thought has a biological function in virtue of the thinker's having a certain kind of evolutionary history, but would deny that having such a history makes the thinker or the thought subject to standards or requirements. In contrast to the kind of theory described in the text, such a position would not be a version of the responsibility view.

standards; competing responsibility-based theories of content appeal to other factors, holding that evolutionary history is only one factor or denying that it has any role. Teleological theories of content share with other responsibility-based theories the basic tenet that what it is for a thought to have a certain content is a matter of the thinker's being responsible to standards.

Sixth, finally, in the case of content, what makes thinkers responsible to the relevant standards? Since the responsibility view explains which concepts figure in a person's thought in terms of responsibility to standards, it cannot appeal to concepts to explain what makes thinkers responsible to those standards. Leading theories of content have identified a range of factors as determinants of content. Such factors include our own dispositions, other people, the external world, and our evolutionary history. As discussed below, the responsibility view has a unifying explanation of why these apparently diverse factors are, or are plausibly thought to be, determinants of content: the factors are all candidates for what makes us responsible to standards. In the next section, I sketch the way in which a responsibility-based theory might treat the cases of deference and concept mastery.

7.3 The responsibility view on deference and concept mastery

In Burge's well-known arthritis example, the thinker has thoughts involving the concept of arthritis, though he lacks mastery of that concept. As we saw in section 6.5, standard theories of content lack an adequate account of why the thinker's deference enables him to have a thought involving the concept. Their accounts are unsatisfactory even assuming that the thinker's deference amounts to dispositions to accept correction from others. As we noted, in many cases, it is plausible that the deference that apparently makes it possible to have thoughts involving a concept involves no more than membership in a community and the absence of an explicit intention to use a word in an idiosyncratic way.

The responsibility view offers a natural way of making sense of the role of deference. Dispositions to accept correction from others may constitute a commitment to the standards accepted by those others. Similarly, mere membership in a community may be sufficient to subject a thinker to community standards, as long as the thinker does not have explicit intentions to the contrary. Thus, on the responsibility view, deference enables thinkers to have thoughts involving a concept because it makes them responsible to the standards that are canonical for the concept.

What about a person who fully grasps a concept? As we will discuss, fully grasping a concept is not plausibly sufficient by itself to make a thinker responsible to its standards. That is not to say that full grasp is irrelevant. The thinker's mastery of the concept may determine which standards are in question. Something else is needed, though, to make the thinker responsible to those standards. One possibility is that a thinker's mastery of a concept allows her to identify the standards in question, thus enabling her explicitly to commit herself to those standards. Another possibility is that the factors that make thinkers with full grasp of a concept responsible to the relevant standards are the same as in the case of thinkers with incomplete understanding. For example, as sketched in sections 7.6-7.7, the relevant factors might include membership in a community or engagement in an enterprise. The relevant factors may vary from case to case, and a thinker's being responsible to standards may be jointly determined by different factors.

Even in the case of thinkers with full grasp of a concept, standard theories of content often appeal to some kind of idealization of a thinker's dispositions. But they have a difficult time explaining in which direction idealization should go, and, more fundamentally, why idealization is relevant at all. The responsibility view has a ready account of the role of idealization. When a thinker commits herself to the standards that correspond to her own dispositions or those of her community, she may be best understood as committing herself not to the actual dispositions but to an idealization of them. In other words, idealization is relevant because a thinker's commitment is plausibly to a standard that is captured by an idealization of the community's dispositions, not by their actual contours.

A responsibility-based theory must disagree with achievement-based theories with respect to what makes it the case that people's thoughts have their contents. However, a responsibility-based theory need not recognize cases of incomplete understanding other than those recognized by achievement-based theories. In principle, a responsibility-based theory could agree with achievement-based theories about what contents people's thoughts have. Once a theory claims, however, that the unifying feature – the feature in virtue of which people with mastery of a concept and people who defer to others have thoughts involving the same concept – is responsibility to standards, the assumption that concept mastery and deference are the only ways of *having a concept is undermined. For there may well be other ways of being responsible to a concept's standards.

7.4 Being responsible to versus being disposed to satisfy standards

The responsibility view's potential to account for deference cases and for incomplete understanding more generally depends on a crucial distinction between being responsible to a standard and being disposed to satisfy it. Call this the *Very Important Distinction*, or *VID*. As I have suggested, in general, people can be responsible to standards that they regularly fail to fulfill, and even that they lack the ability to fulfill.[4] Drivers of cars, for example, merely by operating their cars on public streets, become subject to standards that many of them obviously lack the ability to fulfill. To take a more relevant kind of example, the work of Amos Tversky and Daniel Kahneman has demonstrated that human beings have a deeply ingrained tendency to think irrationally in certain respects. But we are nonetheless *subject to* standards of rationality – we are incorrect or make a mistake when we, for example, make irrational inferences. Similarly, teleological theories hold that a biological entity such as an organ can have a function that it lacks the ability to perform (*e.g.*, Millikan 1984, p. 29).

Since ability is not necessary for responsibility, the responsibility view has better prospects than the achievement view for explaining how a thinker can have thoughts involving a concept, even if she is systematically disposed to misapply it or has an incomplete grasp of it. The responsibility view makes sense of our attributing beliefs involving a particular concept to people who lack full mastery of that concept: such attributions indicate that, despite diverse incomplete understandings, the people have something in common – their responsibility to the concept's requirements. VID makes it possible to see how an inferential-role view can be correct as a view about the individuation of concepts, without implying that thinkers who use a particular concept must have the ability or way of thinking that corresponds to the concept's inferential role.

Not only is ability to fulfill a requirement not necessary for being subject to it, it is not sufficient. A person can have an ability, yet be subject to no requirement that he exercise it. The ability to make irrelevant distinctions is not sufficient to subject its possessor to a requirement that he exercise it (evidence to the contrary notwithstanding). (It may be the case that we would not typically *describe* an ability as an ability to satisfy a requirement to which a thinker is not subject, but there is a simple pragmatic explanation for this: there is typically no reason to call attention to the fact that a thinker's ability is an ability to satisfy a particular requirement unless the thinker is subject to the requirement.) There seems to be nothing special in this respect about the case of conceptual

[4] In section 2.9, I addressed an objection based on the doctrine that "ought implies can."

standards. Having a disposition or ability to satisfy a concept's standards is not plausibly sufficient to make a thinker subject to those standards.

7.5 The responsibility view and the Three Problems

In section 2.5, I sketched the responsibility view's advantages in responding to the Three Problems, and deferred some objections. In this section, I return to the responsibility view's potential to solve or avoid the Three Problems, and respond to the long-deferred objections. I begin with the first and third problems – the problems of allowing for error and of accounting for incomplete understanding. I treat them together because, as we will see, the responsibility view tends to undermine the theoretical importance of the distinction between the two problems.

7.5.1 Error and Incomplete Understanding

On the achievement view, *having a concept requires having mastery of it. Thus, the achievement view must explain away apparent cases of incomplete understanding. In practice, achievement-based theories often try to address some cases of incomplete understanding by adding a proviso about deference, but, as we have seen, they lack a theoretically satisfactory treatment even of deference-based cases. In contrast, achievement-based theories' best option for allowing for error is to draw a competence/performance distinction and to hold that errors are mere performance. That is, they hold that thinkers who make systematic errors but nonetheless have (non-deference-dependent) thoughts involving the relevant concepts in fact have full grasp of the concept. Thus, achievement-based theories must distinguish sharply between cases of thinkers who have full grasp of the concept – who have the ability to satisfy its associated condition – and thinkers who do not.

The responsibility view requires no such distinction. Regardless of whether the thinker has full grasp of a concept, nearly fully grasp of a concept, or highly incomplete or mistaken grasp of a concept, the thinker's thoughts involve the concept, if they do, in virtue of the thinker's responsibility to the standards associated with the concept. (As noted above, a thinker's abilities and dispositions (including both mastery of a concept and deference to others) may help to determine what standards the thinker is responsible to, so abilities and dispositions are by no means irrelevant.) As we saw above, thinkers who lack full grasp of a concept can in principle be subject to its standards. So from a theoretical point of view, there is no fundamental difference between, on the one hand, cases of full grasp but systematic error and, on the other, cases of incomplete understanding. In both cases, the basic question is to what standards the thinker is responsible.

In fact, the responsibility view can dispense with the notion of full grasp of a concept. Its account of what makes it the case that a thought involves a particular content need make no appeal to the notion. Arguably the notion of full grasp of a concept is a technical notion, not found in our ordinary practices. Moreover, in earlier chapters, we saw that it is difficult to spell out the notion of full grasp in a fully general and precise way. It is thus arguably an advantage of the responsibility view that, in contrast with the achievement view, it places no theoretical importance on the notion.

Allowing for systematic error and accounting for incomplete understanding are problems for achievement-based theories of content because of their basic tenet that a thought's involving a particular concept is the thinker's exercising the right dispositions. The responsibility view holds instead that a thought's involving a concept is the thinker's being subject to the concept's standards. Since failing to satisfy a standard to which one is subject and, further, lacking the ability to satisfy such a standard are consistent with being subject to the standard, the problems of allowing for error and of accounting for incomplete understanding are not problems for the responsibility view at all. The task for a responsibility-based theory is, rather, to develop an account of what makes thinkers responsible to the relevant standards. I sketch some ideas for developing such an account in sections 7.6 and 7.7 below.

I have argued that the responsibility view has an advantage over the achievement view because, crudely put, it is less demanding of thinkers. As a result, it is better able to account for systematic error and incomplete understanding. It might seem that the responsibility view is vulnerable to an objection from the opposite direction. Is the responsibility view too liberal in granting contents to thinkers? Since it does not require that a thinker be disposed to satisfy a concept's associated condition in order to have thoughts involving it, won't the responsibility view award contents to thinkers who do not plausibly *have the relevant concepts?

The responsibility view has a strong response. In general, a person's dispositions and understanding can play an important role in determining the standards to which the person is subject. For example, in many situations, a person must have an ability to go some way toward satisfying a standard in order to be subject to the standard. Although a person who lacks the ability to obey the rules of chess perfectly may nevertheless be subject to the rules of chess, it may be that a person who is disposed to move the pieces randomly is not subject to them. Why should this be? It seems plausible that the person's having some ability "in the right direction" plays a role in determining which standard the person is subject to. On this picture, there are determinants of content other than the person's dispositions, but, at least in some circumstances, if the person lacks any disposition in the right direction, the other determinants may leave it underdetermined what the relevant standard is. For present purposes, the precise account does not matter. What is important is the plausibility of the idea that a person's having a disposition partially to satisfy a standard can help to make the person subject to the standard.

In sum, the responsibility view may have the potential to steer between two opposing pressures on theories of content. On the one hand, it avoids being too demanding on thinkers – it allows that thinkers can make systematic errors and, worse, can lack mastery of their concepts. On the other hand, it need not be too liberal in licensing attributions of content and can explain why a thinker tends to need an ability to go some way toward satisfying the relevant conceptual standards.

7.5.2 Normativity

The responsibility view cannot be accused of not taking the normativity of content seriously. As noted in section 2.5.2, however, one might worry that the view, by appealing to normativity directly, fails to be explanatory. I distinguished three different objections. According to the first, the responsibility view merely re-poses the question that a theory of content is supposed to answer. Second, because it appeals to normative facts, the responsibility view gives up the potential to explain the content of thought. Third, the view is circular because the crucial notion of responsibility to a standard itself depends on intentional notions. I now respond to the first two objections. (As I explain, I defer discussion of the third until the end of section 7.7.)

A natural reaction to the responsibility view is that it merely poses the question that a theory of content seeks to answer. This reaction can be elaborated as follows. Theories of content take for granted that having a concept involves being responsible to the concept's standards. The hard problem, the objection continues, is to say what makes a thinker subject to the relevant standards – and that is the problem to which leading theories of content are addressed. On this way of looking at things, such theories of content are all attempts to answer the question of what makes a person subject to the relevant conceptual standards.

Before responding to this objection, it is important to emphasize that, as noted above, the responsibility view does not hold merely that one who has a concept is necessarily subject to its standards. Rather, it holds that being subject to a concept's standards is (at least in part) what *makes* a thinker's thought involve that concept. Thus, a view according to which a person's being subject to a concept's standards is a *consequence* of the person's having a thought involving the concept is not a version of the responsibility view. The responsibility view takes the direction of explanation to be the other way around. An account of why a person is subject to certain standards explains why the person's thought involves the concept for which the standards are canonical.

In light of this clarification, I want to make several closely related points. First, it is not correct that theorists of content generally advance their theories as attempts to answer the question of what makes a person subject to a concept's standards. For the most part, neither inferential-role nor informational theorists would accept this characterization of their projects.[5] They certainly would not accept that a thinker's being subject to standards is what makes it the case that a thought involves a particular concept. They would not accept even that a theory of content must explain a thinker's being subject to a concept's standards.[6]

Second, if the objection were correct, what would be needed to defend a theory of content would be an argument that the theory explains what it is in virtue of which thinkers are responsible to conceptual standards. But this is not the way in which theories of content are typically defended. (For example, solving the problem of error would not be the focus of attention if the goal were to show that the theory explains what makes thinkers subject to conceptual standards.)

Third, and most importantly, achievement-based theories of content would be much less appealing if the question they were designed to answer were what makes a person subject to a concept's standards. Cast as an answer to that question, achievement-based theories of content would hold that what it is in virtue of which a thinker is subject to a concept's standard is the thinker's having a disposition to satisfy the standard. As we saw in sections 6.8 and 7.4, it is difficult to see what the argument for this position could be. In general, that one is disposed to φ does not make it the case that one is required to φ, not even other things being equal. (Think of all the illegal, immoral, irrational, and silly things that people are disposed to do.)

[5] See, *e.g.*, Fodor, 1990, pp.128-130, 135 fn. 35.
[6] As noted in section 6.7, a constitutive account of a phenomenon need not by itself entail all necessary truths about the phenomenon.

Fourth, even if being disposed to satisfy a concept's standard made a person responsible to that standard, this possibility would not support DV without an argument that deference were the *only* other way of being responsible to a concept's standards. But no such argument has been offered.

One aspect of the objection is correct. The responsibility view is not an answer to the content question on a par with informational or inferential-role theories of content. Rather, it is a very general view about the nature of mental content that competes with the achievement view. As such a general view, it re-poses the challenge of offering a theory of content as the challenge of explaining what makes us responsible to conceptual standards. But this feature of the responsibility view does not mean that it does no explanatory work. Establishing that a theory of content must answer the question of what makes thinkers responsible to conceptual standards would entail a major shift in the theory of content.

As we have seen, even without a developed theory of what makes people subject to conceptual standards, the responsibility view offers the prospect of explaining at an abstract level how thinkers can have thoughts involving incompletely grasped concepts. More generally, the responsibility view raises the possibility that we can draw on our understanding of what makes people subject to standards in other domains to explain what makes it the case that a thought involves a particular concept. (We have already implicitly relied on that understanding in arguing that an agent's having a disposition to φ does not make her subject to a standard requiring that she φ.)

I now turn to the second objection – that the responsibility view cannot be explanatory because it appeals to normative facts. A familiar idea is that the normativity of a domain or notion precludes further explanation, or at least further explanation in non-normative terms. Accordingly, appeals to normativity are often made in the service of claims that further explanation cannot be had. As we noted (section 6.7), for example, Kripke (1982) maintains that the normativity of content has the consequence that dispositions are the wrong sort of thing with which to account for content.

It is therefore important to emphasize that the responsibility view's appeal to normativity has the potential to expand explanation, rather than to cut it off. To begin with, as noted above, responsibility to standards is a familiar and basic notion that cuts across many domains. Explaining content in terms of such a notion offers explanatory pay-off even if the notion is itself not fully understood.

It is true that the responsibility view is not likely to yield an account of content that does without all normative notions. For the prospects of giving a fully non-normative account of what it is to be responsible to a norm seem poor. At the same time, however, the explanation of why a thought involves a particular concept need not end with the basic fact that the thinker is responsible to the relevant standards. As suggested above, we can explain what makes people (or mental states or actions) responsible to conceptual standards. In order to develop such explanations, we can draw on our understanding, from other domains, of what makes people responsible to standards.

Moreover, we may well be able to give such explanations in non-normative terms. The kinds of facts mentioned above that make us responsible to standards are plausibly non-normative (or can be accounted for in terms of non-normative facts). For example, the fact that a person accepts correction from others, participates in an enterprise, or has certain capacities may make the thinker subject to certain standards. Thus, there is a sense in which the responsibility view need not leave us with any unexplained normativity. There is no inconsistency in claiming that we can explain in non-normative terms what makes people responsible to standards though we cannot give a non-normative account of what it is to be responsible to a standard. Although having certain capacities may make me responsible to standards of rationality, what it is to be responsible to a standard is not to have those capacities; rather, having those capacities is simply an example of something that can make a person responsible to a standard.

Ultimately, a responsibility-based theory may take the following general form. 1) A thought's involving a particular concept is constituted by the thinker's being responsible to standards that are canonical for that concept. 2) The thinker's being responsible to those standards is explained by factors such as the thinker's deference to others, membership in a community, engagement in certain intellectual enterprises, and situation in the world. Such a theory would not have reduced the notion of a standard or of responsibility to a standard to something non-normative. Nor would it have given an account of what it is for a thought to involve a particular concept in non-normative terms. (Even if what it is for a thought to involve a particular concept is for the thinker to be responsible to its standards, and even if a particular thinker is responsible to a concept's standards by virtue of, say, deference to others, it does not follow that what it is to have the concept is to defer to others.) In sum, the responsibility view offers the prospect of giving a non-normative account of what realizes each thought's involving particular concepts. And the view can also explain, in normative terms, what each of the non-normative realizers have in common – they are all ways of being responsible to the relevant standards.

There is a final worry about the responsibility view's appeal to normative notions. It might be thought that the appeal is somehow circular because normative notions depend on intentional ones. We have to be careful about the kind of dependence that is at stake. For example, it may be that, as a matter of metaphysical necessity, only a creature with intentional states (or something created by such a creature) can be subject to standards (though perhaps evolutionary history can make it the case that very basic creatures or their organs are subject to standards). This kind of dependence, without more, would not necessarily be problematic for the explanatory potential of the responsibility view. Even if intentional content is *necessary* for responsibility to standards, intentional content might fruitfully be explained in terms of responsibility to standards.

A more troubling kind of dependence is suggested by the legal examples I have mentioned. At least in many kinds of cases, intentions play an important role in making people subject to legal standards. A theory that tried to explain content in terms of responsibility to standards would have the potential for problematic circularity if intentions or other intentional states were the only factors that could make people responsible to standards. It will be best to return to this objection at the end of section 7.7 after we discuss in more detail what makes people responsible to standards.

7.6 Responsibility to standards

The central challenge for a responsibility-based theory is to develop an account of where the relevant standards come from and of what makes thinkers responsible to them. I cannot begin to meet that challenge in this essay. In this section, I sketch an impressionistic picture according to which facts about the world – including, for example, mathematical and moral facts – are an important source of the standards to which thinkers are responsible and play a more radical role in determining the content of thought than is envisioned by current externalist theories.

My approach can be understood by opposing it to what I suggest is an overly intellectualized picture of the role of the thinker in determining the content of thought. Saul Kripke's and Hilary Putnam's seminal work on names and natural-kind terms focused attention on the role of the world in determining linguistic meaning and the content of thought. But, as discussed in section 6.3, their arguments do not undermine the Demanding View. The picture that has emerged is that *the* *individual thinker* determines the role that the world and social environment play. The thinker's dispositions, abilities, and practices – including deference to others – determine, for each concept, the respective roles of the world, other people, and the individual thinker. For example, the thinker's dispositions and abilities determine whether a given thought involves a natural-kind concept, such as WATER or COPPER, whose precise reference is determined by how things are in the world, or an appearance concept, such as WATER-LIKE SUBSTANCE or FUZZY, whose reference is determined by how things appear (or some other type of concept, a functional-kind concept, say).

According to this picture, before the development of scientific knowledge, it is already determined that "water" or "star" expresses (or does not express) a natural-kind concept, and that "weed" or "dust" does not. A consequence is that when the world fails to cooperate, a conceptual crisis arises, and we have to renegotiate our terms and concepts. The familiar, though perhaps historically inaccurate, example has it that we initially take "jade" to express a natural-kind concept. Since jade does not in fact have a single underlying nature, the putative concept, along with the beliefs that (apparently) involve it, is incoherent. When we discover the empirical facts, we have to decide on a new use for the term.

I suggest that this way of separating our role from the role of the world, of drawing the line between what is determined *a priori* and what is determined empirically, is artificially and implausibly neat. The picture is highly intellectualized in the sense that it presupposes a great deal of sophistication on the part of thinkers – for example, it presupposes that members of pre-scientific societies already (tacitly) understand that COPPER is a natural-kind concept, *e.g.*, their dispositions must determine that COPPER applies to whatever shares the underlying nature of certain samples and that DUST applies to whatever has a certain superficial appearance (assuming that is a correct account of the concept DUST). (Alternatively, the picture has to deny, implausibly, that people in pre-scientific societies have thoughts involving concepts such as COPPER and DUST). And the picture leaves no room for thoughts involving incompletely grasped concepts (that are not based on deference to other people). A thinker (or thinkers collectively) cannot have an incomplete understanding of the *concept*, as opposed to the objects or property to which the concept refers, for it is the thinker's understanding that determines *a priori* whether his thoughts involve a natural-kind or an appearance concept, for example. The real-life complexities of the relation between conceptual and empirical belie the idea that the demands of our concepts are well defined *a priori* and that the conceptual and empirical realms are cleanly separated (with the consequence that empirical discoveries can straightforwardly reveal our concepts to be incoherent).

On my alternative picture, the default position – operative unless our intentions or beliefs override – is that whether a thought involves a natural-kind concept, an appearance concept, or some other kind of concept *depends in the first place on the world*. This approach replaces reliance on a person's own mastery of a concept, or the person's intentional deference to other people, with the default position that the world controls. Thus, even before we discover that copper has an underlying physical nature and that weeds do not, those facts may play a role in determining that "copper" expresses a natural-kind concept and "weed" does not. On this picture, facts about the world thus play a larger role in determining the content of thought than they have on current, achievement-based externalist theories. Another consequence is that the division between the empirical and the conceptual is much less clean: empirical truths, and also moral, mathematical, and other truths, play a role in determining content that is not pre-defined by our *a priori* conceptual understanding. This possibility helps to explain how conceptual inquiry, including philosophical theorizing, can be informative, even profoundly so.

In this section, I have suggested that responsibility to standards provided by the world is a default position. So far, however, I have only gestured at an approach. In order to develop an account of content along the suggested lines, many questions would need to be addressed, but perhaps the most pressing is what could make us responsible to standards provided by the world. Why is responsibility to the world the default?

7.7 The world as a determinant of content

A suggestion that I hope to explore in future work is that we make ourselves responsible to standards provided by the world. It is because we have certain very general attitudes or dispositions that, in the absence of specific overriding intentions or dispositions in particular cases, we are responsible to the world. Thus, according to the suggestion, a thinker's background *commitment to the world* is what makes it the case that the default is that the thinker is responsible to standards provided by the world. Here's an intuitive story: a thinker's participation or engagement in a very general intellectual enterprise plays an important role in determining the conceptual standards to which she is subject. If a thinker is engaged, for example, in trying to understand some domain, such as mathematics, morality, or the empirical world, facts in that domain may play a role in determining the standards to which the thinker is responsible.

This story relies on the idea of a thinker's engagement in a certain enterprise, which may not be much of an advance over the idea of a thinker's being committed to something. We want to know what sorts of attitudes or dispositions might make it the case that one is committed to the world or engaged in exploring a certain realm? In section 6.5.1, I distinguished two different things that could be called "deference to the world." On the one hand, at least on one plausible account of natural-kind concepts, fully grasping such a concept involves having certain *natural-kind dispositions* – for example, dispositions to take back an initial judgment that something falls under the concept, in the light of certain kinds of evidence about the world. (One contrast is with *superficial-kind dispositions*, for example, dispositions to reject the possibility that something that is superficially indistinguishable from instances of the concept might not itself be an instance of the concept.) I do not use the term "deference to the world" for such natural-kind dispositions.

On the other hand, a thinker may defer to the world in the sense that she allows the world to determine what sort of dispositions she should have with respect to each phenomenon – for example, whether she should have natural-kind or superficial-kind dispositions. (If the responsibility view is correct, the thinker would, in committing herself to whatever standards the world provides, be allowing the world to determine what type of concept the thinker is using.) To take a very crude example, suppose the thinker has a general disposition to acquire natural-kind dispositions with respect to any particular phenomenon just in case the thinker obtains evidence that the phenomenon is in fact a natural kind. And suppose the thinker is disposed to acquire different dispositions – perhaps superficial-kind dispositions – if the evidence comes out differently. Let us use the term "deference to the world" for such a pattern of dispositions or attitudes. The suggestion, then, is that certain, perhaps complex, patterns of dispositions or attitudes – ones that are, in the sense explained, deferential to the world – could effect or constitute a commitment to the world – a commitment to let the world determine, among other things, what specific dispositions one should have with respect to particular phenomena, concepts, or terms.

Such a general commitment could be overridden in the case of a given concept (or phenomenon or term) by the thinker's doing something that makes her responsible to different standards with respect to that concept. This is a sense in which the general commitment establishes only a default. It need not be the case, however, that the default is effective only in the absence of any affirmative factors. A thinker's general commitment might for example be sufficient to override some degree of contrary dispositions.

My suggestion is helpfully compared with David Lewis's (1984; 1983, pp. 370-377) proposal that "objective sameness and difference in nature" plays a role in determining the content of thought. Lewis claims, in effect, that "naturalness" of a property is part of what it is for a property to figure in the content of a thought.[7] Thus, Lewis's proposal and mine are alike in the radical role that they allow the world in determining the content of thought. For example, suppose that a pre-scientific community has a term for the substance that makes up rivers and rain and that the dispositions of the members of the community with respect to that term are indeterminate between natural-kind and superficial-kind dispositions. On both Lewis's proposal and mine, the world might play a role in determining that the term expresses the natural-kind concept WATER, rather than the concept WATER-LIKE SUBSTANCE.

On Lewis's account, however, the role of the world seems insufficiently motivated. Lewis argues that there must be some additional determinant of content if we are to avoid paradox. He thinks that the world must be that additional determinant (1984, pp. 227-228; 1983, pp. 371-372). His account leaves the role of the world motivated only by the need for *some* additional determinant of content to avoid paradoxical conclusions; the account does not explain *why* the objective structure of the world should be that determinant. We want to know what it is for a thought to involve a particular concept such that the structure of the world plays the conjectured role. What is content such that a property's naturalness is part of what makes it figure in the content of a thought?

In contrast, my proposal offers a way of beginning to make sense of the role of the world. If having a thought with a particular content is a matter of being responsible to certain standards, and if our dispositions and attitudes make us responsible to standards provided by the world, then we have the beginning of an explanation of the relevance of the structure of the world.

My suggestion has another possible advantage over Lewis's. On Lewis's account the naturalness of a property is roughly how close it is to – how few "disjunctions" (Lewis 1983, p. 376) it is from – the properties that figure in the laws of the (true) basic physics. A worry about the account is therefore that naturalness in this sense will lead to (intuitively) incorrect assignments of

[7] Lewis credits G.H. Merrill (1980) with proposing the idea.

content. The way of dividing up the world that is appropriate for the explanatory purposes of basic physics may not be appropriate for the purposes of attributions of content. But until we have an explanation of what it is to *have a concept that explains the role of the world, we have no principled way of saying why one way of dividing up the world is more relevant than another. The responsibility view, combined with the idea that it is our dispositions that make us responsible to the world, has the potential for an account according to which the standards to which we are responsible derive from facts about the world that are relevant to humans. To put it another way, the responsibility view might support an approach according to which the right understanding of naturalness, for the purposes of the theory of content, depends on human limitations, needs, interests, and abilities.

I want to mention two important kinds of objections to the idea that *we make ourselves* responsible to – commit ourselves to – standards that derive from the world. The first objection is that, if our dispositions are what make us responsible to the relevant standards, the proposal is just a version of the achievement view.

This objection is off the mark. Dispositions play a crucially different role in the achievement view and the present proposal. According to the achievement view, having a certain disposition – a disposition to satisfy a concept's associated condition – is simply what it is to *have a particular concept. In contrast, according to the present suggestion, having dispositions of a general type – what we have called deferential to the world – is one way, among others, of being responsible to standards. A genuine achievement view denies that a thinker's *having a concept is a matter of a thinker's being responsible to certain standards. So, *a fortiori*, it does not claim that a thinker's being disposed to satisfy a concept's associated condition makes that condition a standard to which the thinker is responsible. This basic contrast between the achievement view and the responsibility view is reflected in some specific differences.

First, the dispositions involved differ between the two kinds of approaches. The achievement view appeals to a disposition to satisfy a concept's associated condition. In contrast, the proposal under discussion appeals to the thinker's commitment to the world, which could be constituted by any of a wide range of dispositions. Crudely speaking, the dispositions would be dispositions to allow the world to determine what kind of dispositions she should have. Another way of putting the point is that the achievement view is *not* merely any "dispositional view" – any view according to which a thinker's dispositions are what make it the case that she is using a particular concept. Rather, it is a narrow category of dispositional view.[8]

Second, the achievement view makes a concept-by-concept appeal to dispositions. It holds that, in order to *have a particular concept, a thinker must have a disposition specific to that concept. The commitment-to-the-world idea appeals to generalized dispositions toward the world. The proposal can still offer a concept-by-concept answer to the content question. It can say that to *have a particular concept is to be responsible to standards that are specific to that concept. The dispositions need not be concept-by-concept in order for them to play a role in making a thinker responsible to standards that are specific to each concept.

These points lead us to an advantage of the proposal over the achievement view. As we have seen, one of the problems for the achievement view is that it is not plausible that thinkers are disposed to satisfy the associated conditions for all the concepts they *have. The proposal is much less demanding of thinkers. It would require only that thinkers are disposed to defer to the world.

To sum up the contrast, the achievement view holds that having a *concept is having a disposition to satisfy that particular concept's associated condition. The present proposal is that *having a concept can be realized by any of indefinitely many dispositions that constitute a commitment to the world – and can also be realized in other ways, by, for example, deferring to others or, perhaps, understanding the concept's standards and deciding to hold oneself to them.

The second objection is that it is question begging to appeal to a commitment to the world in an attempt to answer the content question. Part of what makes it the case that one commits oneself to the world, or is engaged in an enterprise of trying to make sense of the world, is one's intentions. And, the objection continues, it is circular to appeal to a thinker's intentions in answering the content question.

I have four points in response. First, even if it is circular to appeal to deference to the world, the circle may be large enough not to be vicious. It is viciously circular to be told that what makes it the case that a thinker is using the concept PLUS is that the thinker intends to hold herself to the standards of counting, rather than, say, quonting. But it could be explanatory to be told that what

[8] As discussed in section 3.2.2, theorists tend to overlook the possibility of a dispositional theory that is not a version of the achievement view. For example, in their discussion of dispositional theories, Boghossian (1989) and Kripke (1981) both consider only achievement-based theories.

makes it the case that a thinker is using the concept PLUS is that the thinker intends to hold herself to the real structure of mathematics.

Second, we may be able to give an account of what is involved in committing oneself to the world, deferring to the world, or being engaged in an enterprise of trying to understand the world that does not depend on the contents of intentions or other propositional attitudes. An intention to defer is not necessary for ordinary deference to actual thinkers; it is in general enough that a thinker would in fact acquiesce, change her position, accept correction, and so on in appropriate counterfactual circumstances. It may be possible, without appealing to the content of thoughts, to provide an account of how dispositions or underlying abilities can constitute a commitment to standards provided by the world.

Third, a responsibility-based theory that is not a full-blown teleological theory can make a limited appeal to biology. The idea is that the contents of certain very basic attitudes are determined by biological function. (A separate question is whether the best account of biological function is in terms of evolutionary history or something else, for example, the needs and interests of the organism.) Here is a possible picture. Suppose that a thinker has a very general attitude, which we can describe as trying to understand her natural and social environment (or as a commitment to acquiring such understanding). It is not difficult to see how biology could be what determines that the thinker's attitude has this content. Roughly speaking, it could be that in virtue of evolutionary history (or whatever determines biological function), part of the thinker's cognitive apparatus has the function of coming to an understanding of her natural and social environment. Once biology has done its work, the attitude of trying to understand the natural and social environment fixes the default: the thinker is subject to standards provided by the natural and social world. More generally, the idea is that biology determines the content of a few very basic attitudes, and they, in turn, play an important role in making the thinker responsible to standards provided by the world.

Fourth, and finally, facts about evolutionary history may not be the only facts in virtue of which we can be responsible to standards without *making* ourselves responsible to them. In the case of moral, logical, epistemic, and rational standards, for example, it is arguable that one can be subject to standards merely by virtue of one's nature and position in the world. Along with many philosophers, I believe that people have moral obligations that do not derive from any commitments that they have undertaken. Possible examples include obligations to people in urgent need, to parents, and to oneself. Perhaps our nature and position in the world can subject us to conceptual standards, at least to the extent that we do not do something affirmatively to withdraw from or escape that responsibility. These are, of course, merely sketchy suggestions that require much development.

References

Antony, L. & Levine, J. "The Nomic and the Robust," in Loewer and Rey 1991, pp. 1-16.

Block, N. 1986: "Advertisement for a Semantics for Psychology," *Midwest Studies in Philosophy*, 10, pp. 615-678.

Boghossian, P. 1989: "The Rule-Following Considerations." *Mind*, 98, pp. 507-549.

----1991: "Naturalizing Content," in Loewer and Rey 1991, pp. 65-86.

Burge, T. 1979: "Individualism and the Mental." *Studies in Metaphysics*, *Midwest Studies in Philosophy*, 4, pp. 73-121.

----1982: "Other Bodies," in A. Woodfield, ed., *Thought and Object: Essays on Intentionality*. Oxford: Oxford University Press, pp. 97-120.

----1986a: "Individualism and Psychology." *Philosophical Review*, 95, pp. 3-45.

----1986b: "Intellectual Norms and Foundations of Mind." *Journal of Philosophy*, 83, pp. 697-720.

----1990: "Frege on Sense and Linguistic Meaning," in J. Bell and N. Cooper, eds., *The Analytic Tradition*. Oxford: Blackwell, pp. 30-60.

Davidson, D. 1986a: "A Coherence Theory of Truth and Knowledge," in LePore 1986, pp. 307-319.

---- 1986b: "Empirical Content," in LePore 1986, pp. 320-332.

Davies, M. 1991: "Concepts, Connectionism, and the Language of Thought," in W. Ramsey, S. Stich, and D. Rumelhart, eds., *Philosophy and Connectionist Theory*. Hillsdale, New Jersey: L. Erlbaum, pp. 229-257.

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

---- 1988: *Explaining Behavior*. Cambridge, MA: MIT Press.

Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

Fodor, J. 1975: *The Language of Thought*. Cambridge, MA: Harvard University Press.

----1986: "Why Paramecia Don't Have Mental Representations," in *Midwest Studies in Philosophy*, 10, pp. 3-24.

---- 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

----1988: Reply to Boghossian (unpublished), APA Eastern Division Meeting, December 1988, quoted in Boghossian, 1991, pp. 81-82

----1990: *A Theory of Content and Other Essays*. Cambridge, MA.: MIT Press.

----1991a: "Replies," in Loewer and Rey 1991, pp. 255-319.

----1991b: "You Can Fool Some of the People All of the Time, Everything Else Being Equal; Hedged Laws and Psychological Explanation." *Mind*, 100, pp.19-34.

----1993: "Unpacking a Dog" (review of *A Study of Concepts*). *London Review of Books*, 15, 19 (October 7, 1993), pp. 14-15.

---- 1994: *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge, MA: MIT Press.

----1998: *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Horwich, P. 1990: “Wittgenstein and Kripke on the Nature of Meaning.” *Mind & Language*, 5, pp. 105-121.

---- 1998: *Meaning*. Oxford: Oxford University Press.

Hurley, S. 1989: *Natural Reasons*. New York: Oxford University Press.

Kripke, S. 1980: *Naming and Necessity*. Oxford: Blackwell.

---- 1982: *Wittgenstein on Rules and Private Language*. Oxford: Blackwell.

LePore, E. 1986: *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. Oxford: Blackwell.

Lewis, D. 1983: “New Work for a Theory of Universals,” *Australasian Journal of Philosophy*, 61, pp. 343-377. Reprinted in Lewis, D., (1999). *Papers in Metaphysics and Epistemology*. Cambridge, UK: Cambridge University Press, pp. 8-55.

---- 1984: “Putnam’s Paradox,” *Australasian Journal of Philosophy*, 62, pp. 221-236.

Loewer, B. and Rey, G., eds. 1991: *Meaning in Mind: Fodor and His Critics*. Cambridge, MA: Blackwell.

Loar, B. 1981: *Mind and Meaning*. New York: Cambridge University Press.

Merrill, G. 1980: “The Model-Theoretic Argument against Realism.” *Philosophy of Science*, 47, pp. 69-81.

McDowell, J. 1984: “Wittgenstein on Following a Rule.” *Synthese*, 58, pp. 325-363.

McGinn, C. 1984: *Wittgenstein on Meaning*. Oxford: Blackwell.

Millikan, R. 1984: *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

1990: “Truth Rules, Hoverflies, and the Kripke-Wittgenstein Paradox.” *Philosophical Review*, 99, pp. 323-353.

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

Pettit, P. 1990: “The Reality of Rule-Following.” *Mind*, 99, pp. 1-22.

---- 1993: *The Common Mind*. New York: Oxford University Press.

Pietroski, P. and Rey, G. 1995: “When Other Things Aren’t Equal: *Saving Ceteris Paribus* Laws from Vacuity,” *British Journal for the Philosophy of Science*, 46, pp. 81-110.

Putnam, H. 1975: “The Meaning of ‘Meaning’,” in his *Philosophical Papers, Volume 2: Mind, Language and Reality*. Cambridge: Cambridge University Press, pp. 215-271.

----1981: *Reason, Truth and History*. Cambridge: Cambridge University Press.

---- 1992: *Renewing Philosophy*. Cambridge, MA: Harvard University Press.

Rey, G. 1997: *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Oxford: Blackwell.

---- 1999: “Resisting Primitive Compulsions,” in E. Margolis and S. Laurence, *Concepts: Core Readings*. Cambridge MA: MIT Press, pp. 339-343.

---- forthcoming a: “Physicalism and Psychology: A Plea for Substantive Philosophy of Mind,” in C. Gillet and B. Loewer, eds., *Physicalism and Its Discontents*.

---- forthcoming b: “Digging Deeper for the *A Priori*,” in *Philosophy and Phenomenological Research*.

Schiffer, S. 1987: *Remnants of Meaning*. Cambridge, Mass.: MIT Press.

Schwartz, S. ed. 1977: *Naming, Necessity, and Natural Kinds*. Ithaca, N.Y.: Cornell University Press.

Stalnaker, R. 1987: *Inquiry* , Cambridge, Mass.: MIT Press.

Stampe, D. 1977: “Towards a Causal Theory of Linguistic Representation,” *Midwest Studies in Philosophy, II: Studies in the Philosophy of Language*, 42-63. Morris, MN: University of Minnesota Press.