Four Essays on Self-Knowledge

by

Casey Brian Doyle

B.A., Bates College, 2007

Submitted to the Graduate Faculty of the

Kenneth P. Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This dissertation was presented

by

Casey Doyle

It was defended on

April 6, 2015

and approved by

Karl Schafer, Associate Professor of Philosophy

Edouard Machery, Professor of History and Philosophy of Science

Dissertation Advisor: John McDowell, Distinguished University Professor, Philosophy

Dissertation Advisor: Kieran Setiya, Professor of Philosophy, MIT

Copyright © by Casey Doyle

2015

Four Essays on Self-Knowledge

Casey Doyle, PhD

University of Pittsburgh, 2015

I defend a Constitutivist account of self-knowledge of the intentional attitudes consisting of two claims. First, for a class of self-beliefs, *believing* that you are in a mental state guarantees being in that state. Second, for a class of mental states, *being in that state* can be a source of knowledge that you are in it. The first thesis must be accepted in order to explain unique features of self-knowledge including its value, the roles it plays in reasoning, and the way in which it enables self-expression. The second thesis must be accepted in order to maintain the view that in inference a subject takes it that her premises support her conclusion.

TABLE OF CONTENTS

INTRODUCTION 1

0.1 A PUZZLE ABOUT SELF-KNOWLEDGE 1

0.2 VARIETIES OF CONSTITUTIVISM 7

0.3 SELF-INTIMATION AND HUMAN FALLIBILITY 12

1 TESTIMONY AND THE RATIONAL SIGNIFICANCE OF SELF-KNOWLEDGE 16

1.1 A PUZZLE ABOUT TESTIMONY 16

1.2 CONSTITUTIVISM ABOUT SELF-KNOWLEDGE 21

1.3 THE EPISTEMIC ADVANTAGE OF SELF-KNOWLEDGE 26

1.4 THE SHORTCOMINGS OF TESTIMONY 28

1.5 RATIONAL SIGNIFICANCE 31

1.6 AUTHORITY 39

2 SELF-KNOWLEDGE AS AN ABILITY 43

2.1 KNOWLEDGE AND SELF-KNOWLEDGE 43

2.2 AUTHORITY 46

2.3 HYMAN ON KNOWLEDGE 49

2.4 REASONS AND BELIEFS 52

2.5 ACTING FOR THE REASON THAT I BELIEVE THAT P 55

2.6 OBJECTIONS 61

3 EXPRESSION AND THE CONSTITUTION OF SELF-KNOWLEDGE 64

3.1 FIRST-PERSON AUTHORITY 64

3.2 TWO ASPECTS OF FIRST-PERSON AUTHORITY 66

3.3 AUTHORITY AND CONSTITUTIVISM 73

3.4 EXPRESSION.. 75

3.5 FACTIVITY 86

3.6 SELF-KNOWLEDGE AND SELF-EXPRESSION 88

3.7 A NEO-EXPRESSIVIST OBJECTION TO CONSTITUTIVISM.............. 91

4 INFERENCE AS A SELF-CONSCIOUS ACTIVITY 93

4.1 INTRODUCTION 93

4.2 SELF-INTIMATION 97

4.3 THE TAKING CONDITION ON INFERENCE... 103

4.4 THE ACCESS OBJECTION.. 108

4.5 CARROLL'S AND FUMERTON'S REGRESSES..................................... 112

4.6 INFERENCE AS A SELF-CONSCIOUS ACTIVITY................................ 115

4.7 OBJECTIONS... 120

BIBLIOGRAPHY 126

PREFACE

I am grateful to many people for assistance in producing this dissertation. First of all, I am grateful to my advisors, Kieran Setiya and John McDowell. Kieran has helped me see this project through from beginning to end and has put far more time in than could reasonably be expected of anyone who doesn't owe me money. I thank Kieran for teaching me how to write philosophy and how to properly appreciate dialectical context. John's philosophical work has been an inspiration to me since before I came to Pittsburgh. In many ways the dissertation is my attempt to work through in my own way (i.e. clumsily) some of the ideas I have inherited from him. I am also extremely grateful to the other members of my committee. Karl Schafer offered perceptive comments on many drafts of the chapters. Edouard Machery has been extremely kind to me both as a reader and teacher. I could not have asked for a better committee.

I would also like to thank the friends who have helped me find my voice over the course of countless conversations, philosophical and otherwise, especially Preston Stovall, Tom Marre, Stephen Makin, Joseph Milburn, Ben Schulz, and Eric Palmer.

Most of all I am grateful for my best friend and wife, Katelyn. In documents like this it is conventional to say of the person one loves that "without whom not." I am not sure that I wouldn't have completed a dissertation without Katelyn in my life, but I am sure that there wouldn't be much point.

INTRODUCTION

0.1 A PUZZLE ABOUT SELF-KNOWLEDGE

Each of us possesses a way of knowing our beliefs, desires, intentions, emotions and other attitudes that is unavailable to anyone else. Call this "first-personal self-knowledge"; it is to be contrasted with ways of knowing one's mind that are equally available to others, such as observation, inference from behavior, and testimony.[1] The motivating question of this dissertation is: what is the source of first-personal self-knowledge of the attitudes? Given that it is not based on observation or inference from behavior, in virtue of what do we possess it?[2]

An odd feature of this question is that it focuses on cases that non-philosophers typically find quite uninteresting. When people outside of the discipline think of philosophical reflection on self-knowledge they are likely to imagine accessing one's "true self," moral character, deepest motivations, values, and the like. One can imagine and sympathize with their disappointment, then, when they hear that philosophers are typically interested in understanding such things as how you know that you believe that it's Monday. However, viewed from the perspective of the question I have raised, the mundane case is revealed as puzzling in a way that, arguably, the other cases are not. Whatever such "deep" facts about oneself consist in and

---

[1] Many philosophers refer to first-personal self-knowledge as "introspective," but I prefer to avoid that term, which suggests that we know our minds through some kind of inner observation.

[2] Unless otherwise noted, when I speak of "self-knowledge" I will have the first-personal kind in mind.

however exactly we access them, it seems that we do so by broadly third-personal means, drawing inferences from behavioral and other evidence, relying on the testimony of those close to us, and so on. Indeed, what makes these facts about ourselves troubling for us and worthy of our time is precisely the fact that we lack the kind of immediate access to them that we typically have to our propositional attitudes. Put another way: in one way or another, learning about our characters and the like is the result of a kind of *discovery*. By contrast, when it comes to our beliefs and other attitudes, it seems wrong to think of us as having to discover them at all. Typically, you do not have to figure out what you believe, desire, or intend to do: you know these things straight away. Our relation to our attitudes is intimate in a way that our relation to our characters and true selves is not. Indeed, the very same feature that makes knowledge of attitudes *uninteresting* to us in our everyday lives (what could be easier?) makes it philosophically puzzling.

The solution to this puzzle offered in this dissertation is surprisingly straightforward. To the question, how do I know that I believe that p? I answer: by believing that p. For a class of mental states, a subject knows that she is in one of those states simply by being in it. Put another way, being in a first-order state can be metaphysically sufficient for knowing that you are in that state. In the literature this view is called "Constitutivism" because it explains our self-knowledge in terms of the constitutive relations between first-order states and self-knowledge. Since these relations hold in both directions the view is best understood as the conjunction of two claims.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.

Authority holds that if you first-personally believe that you are, say, angry, it just follows that you are. Part of what it is to first-personally believe you are angry is to actually be angry. As I like to put it, first-personally believing that you are angry is a *way of being angry*. Self-Intimation holds that if you are angry it just follows that you know that you are; part of what it is to be angry is to know that you are. Such states are necessarily self-conscious.

Two quick qualifications about the scope of these claims will help to make a start at rendering them more attractive. First, it should not be thought that Authority holds of just any self-ascription of a mental state, any belief with the content "I am in M." Authority is restricted to our first-personal self-beliefs. Thus, if I believe that I am angry on the basis of your testimony it does not follow that I am, in fact, angry. That leaves the defender of Authority the task of identifying what unifies the class of first-personal self-beliefs other than their truth. Second, Self-Intimation does not hold of all mental states. There are many kinds of mental states of which, by their nature, the subject is ignorant: implicit biases and the inhabitants of the Freudian unconscious, for example. These are not counterexamples to Self-Intimation. That leaves its defenders the task of demarcating its scope: which mental states are self-intimating and why?

The chapters that make up this dissertation pursue these two tasks. The first three are focused on Authority, they seek to establish that it offers the best explanation of distinguishing features of first-personal self-knowledge. These chapters pursue roughly the same strategy. I begin by highlighting a mark of first-personal self-knowledge, something that distinguishes it from knowledge of one's mind based on third-personal sources. Importantly, these features are not narrowly epistemic: I do not focus on the high degree of justification we have for our self-beliefs. Instead, I focus on non-epistemic facts about first-personal self-knowledge: that it holds a special *value* for the subject who possesses it, that it plays distinctive roles in *reasoning*, and

that it enables the subject to *express* her mind by avowing it. To make sense of these differences between first-personal self-knowledge and other ways of knowing one's mind, I argue that we should think of the former as constituted by dispositions not found in the latter. Then I argue that these dispositions are sufficient for possession of the relevant first-order state. I argue further that these dispositions are features of the capacities that give rise to first-personal self-knowledge, and are realized in first-personal self-*beliefs* as well. Thus, possessing a first-personal self-belief that one is in a mental state involves possessing dispositions sufficient for being in that state, which means first-personal self-beliefs are guaranteed to be true, as Authority claims.

This strategy differs from what is standardly pursued in the contemporary literature. In discussions of self-knowledge it is common to start by pooling data about its epistemic features and then propose a mechanism or capacity that plausibly yields beliefs with these features. The primary question is: how do we arrive at beliefs with these epistemic features? In these chapters I postpone this question. Instead of focusing on processes of arriving at self-knowledge, I'm interested in understanding what self-knowledge is like as a standing state of mind, what roles it plays in our mental economy. Of course, if Authority is true then first-personal self-knowledge is *infallible*, which explains its epistemic privilege. But I think that we go astray in assuming that what is most in need of explanation about first-personal self-knowledge are its epistemic features.

Here is a brief summary of the chapters that pursue this strategy:

The first chapter focuses on the value of first-personal self-knowledge. Like other matters of fact, we can know our minds on the basis of the testimony of others. Yet, intuitively, this is inferior to first-personal self-knowledge. Other things being equal, it is better to possess first-personal self-knowledge than to have to depend on another's say-so. I argue that the value of

self-knowledge cannot be accounted for in terms of either epistemic asymmetries between the first and third-person perspectives or the demerits of testimony in general. Instead, first-personal self-beliefs are special for the role they play in reasoning. A first-personal self-belief can play the same roles in reasoning as the state represented in its content. Because a disposition to reason in this way is sufficient for possession of the relevant first-order states, explaining the value of self-knowledge commits one to Authority.

Chapter Two addresses the relation between self-knowledge of belief and reasons for action. John Hyman has argued that we should think of knowledge as the ability to be guided by the facts. One knows that p if and only if the fact that p can serve as one's reason for action. This would suggest that when one knows that one believes that p one can act for the reason that one believes that p. But there is an ambiguity in this expression. When one possesses observational or inferential knowledge of one's belief, acting because one believes that p is like acting for any other reason, in this case a fact about oneself. But when one possesses first-personal self-knowledge acting because of the fact that one believes that p is just another way of acting on the basis of the fact believed, or aspiring to. In these latter cases having self-knowledge is simply a way of knowing, or aspiring to know, some extra-mental fact. To explain this, I propose that self-knowledge is a distinct kind of ability, the ability to act for reasons self-consciously, in the awareness of one's reasons as reasons. Rather than thinking of self-knowledge as knowledge of a fact about oneself, we should think of it as a way of being in a first-order state. The role of self-knowledge of belief in practical reasoning supports Authority.

Another way in which self-knowledge is special is its connection to *self-expression*, the topic of Chapter Three. A subject who first-personally knows that she is angry is thereby in a position to express that anger by avowing it. This suggests that there is an intimate connection

between our capacities for self-knowledge and self-expression. I argue that this connection should be captured as follows: first-personal self-beliefs are partially constituted by a disposition to express the states that are represented in their contents. Now expression is a factive phenomenon in that expressing that one is in a mental state entails being in that state. Therefore, first-personal self-beliefs are partially constituted by a disposition which guarantees the presence of the states that are in their content. The expressive character of self-knowledge supports Authority.

In the fourth chapter I turn to Self-Intimation and the demarcation task, focusing on the case of belief for a reason or inferential belief. A necessary condition on the inferences of rational creatures like us is that the subject takes it that her premises support her conclusion. However, the idea that this attitude of "taking" functions as a premise standing in need of justification leads to familiar Lewis Carroll-style regresses. If one's taking is justified inferentially then one must have made a prior inference in order to draw an inference, rendering inference impossible. If it is a premise then there must be a further attitude of taking it as a premise, which then must also function as a premise, ad infinitum. The regress worries can be avoided if there is noninferential knowledge of justificatory relations between propositions the source of which is not a premise in inference. I propose that our capacity for self-knowledge can provide this, but only if we accept Self-Intimation. If Self-Intimation is true then a subject who believes that p on the ground that q knows this noninferentially. Furthermore, in possessing this self-knowledge she represents herself as believing that p for that reason. We can retain the attractive view of inference by endorsing Self-Intimation.

0.2 VARIETIES OF CONSTITUTIVISM

The central Constitutivist idea is that there is a dependence between first-order states and self-knowledge. This dependence has been understood in at least two ways. According to "grammatical" versions of Constitutivism, the dependence is merely a feature of the grammar of psychological concepts. Thus, according to Crispin Wright (1989) and Richard Rorty (1979) it is a brute feature of those concepts that their first-person present tense use is incorrigible. There is no interesting fact about the nature of mental states that explains this. On the other view, defended here, there are such facts. Constitutivism is true in virtue of the dispositions constitutive of a class of first and second-order states. Call this the "ontological" version of Constitutivism.[3]

The most prominent version of the ontological view is that defended by Sydney Shoemaker (1996, 2012).[4] It will be helpful to contrast the view and argumentative strategy pursued here with Shoemaker's. Like Shoemaker, I think that we should look to philosophy of mind to explain first-personal self-knowledge because self-knowledge is a feature of mental states rather than a mode of access to them. And that means that the view must be placed in the context of some general account of the nature of the propositional attitudes. While Shoemaker accepts Functionalism, and takes the arguments for Constitutivism to require Functionalism, I do not assume this. I make the weaker assumption that the attitudes should be thought of as clusters of dispositions to reason and behave in a variety of ways. Thus the Constitutivist claims should be understood on the following schema: if a mental state M1 is partially constituted by dispositions x, y, and z, and those dispositions are sufficient for possessing M2, then M1 is

---

[3] I borrow these terms from Bar-On 2009. I discuss the views further in chapter two.
[4] See also Heal 1996, O'Shaughnessy 2000, Rödl 2007, and Boyle 2011.

partially constituted by M2. This can be understood in Functionalist terms, though it needn't. I take no stand on the truth of Functionalism.

Shoemaker's most well-known argument for Constitutivism involves the idea of self-blindness. A subject is said to be self-blind about her beliefs just in case she possesses unimpaired rationality and the concept of belief but lacks first-personal access to her beliefs. Such a subject is in a position with respect to her beliefs that someone who becomes blind is in with respect to red objects: she understands what they are, they exist, but she cannot know of them, at least in the standard way. Shoemaker argues that self-blindness is impossible, a subject with the relevant concepts and unimpaired rationality is guaranteed to possess self-knowledge of at least some of her attitudes. Reasons for this impossibility include the following. First, deliberation requires reflection upon one's mental states, a subject with unimpaired rationality is capable of deliberation, and so such a subject must be capable of possessing self-knowledge. Second, a self-blind subject would be disposed to make Moore-paradoxical assertions about her beliefs such as "it's raining, but I don't believe that it is." However, anyone who possesses the concepts of the attitudes will avoid making such assertions, and so will behave the same as someone who possessed self-knowledge of her beliefs. Given a broadly Behaviorist assumption about the possession conditions on the attitudes, she thereby counts as possessing self-knowledge, or, at least, a second-order belief.[5]

Shoemaker's arguments have been objected to on a number of fronts (Finkelstein 1999, Kind 2003, Siewert 2003). One might reject the behaviorist claim invoked in the argument from Moore's paradox; or else one might insist that pragmatic facts about the appropriateness of *asserting* Moore-paradoxical propositions tell us nothing of interest about *believing* them. While

[5] Both of these arguments are developed in the essays collected in his 1996, especially chapters 2, 10, and 11, as well as his 2012.

some argue that Shoemaker overemphasizes the place of deliberation in our lives (Cassam 2015), others deny that self-knowledge is required for deliberation (Moran 2001, Setiya 2011), and others still hold out for the possibility that third-personal sources can provide the self-knowledge required for deliberation (Kind 2003). For my own part, while I agree with Shoemaker that self-blindness is impossible, I am less sure that Constitutivism follows. Call the claim that self-blindness is impossible No Self-Blindness:

> *No Self-Blindness (NSB)*: Necessarily, a subject with unimpaired rationality and the concept of belief is able to possess knowledge of her beliefs without relying on third-personal sources.[6]

Shoemaker holds that a version of Constitutivism follows from NSB:

> *Constitutivism*: necessarily, a subject with unimpaired rationality and the concept of belief who believes that p believes that she believes that p.

This is because any view that accepts the following must reject NSB:

> *Independence*: it is possible that a subject with the concept of belief and unimpaired rationality believes that p without believing that she believes that p.

Independence holds that believing that p is one thing and believing one so believes is another. In Hume's terms, first and second-order states are "distinct existences." It is Shoemaker's view that anyone who accepts Independence is committed to the possibility of self-blindness. The thought is that if Independence is true then self-knowledge must result from the operation of a capacity distinct from that involved in forming the first-order state. But then, if the connection between the belief and self-knowledge can be severed in one instance, in principle it can be severed for all cases, which would make a subject self-blind. And so if we accept NSB we must reject Independence and accept Constitutivism.

[6] I have couched NSB in terms of belief, though Shoemaker holds that the thesis also holds of other propositional attitudes and the sensations. I'll focus exclusively on belief here for the sake of simplicity and because it is arguably the most plausible case for the thesis.

The problem with this argument is that it conflates a dependence between *states* with a dependence between *capacities*. As we can put it, from Independence it follows that the exercise of a belief-forming capacity that yields the belief that p might not also yield the belief that one believes that p. But it does not follow that the self-belief cannot be produced through a different exercise of *the very same capacity* or anyways another capacity constitutive of either unimpaired rationality or possession of the concept of belief. That is, NSB and Independence can be reconciled so long as we accept that possession of either unimpaired rationality or the concept of belief is sufficient to put one in a position to know one's beliefs. Such a view is hinted at by a famous remark of Gareth Evans:

> A subject can gain knowledge of his internal informational states [i.e., his perceptions] in a very simple way: by re-using precisely those skills of conceptualization that he uses to make judgments about the world.
>
> Evans 1982, 227

I say that Evans hints at such a view because he did not develop it in detail. Drawing on Evans, and appealing to the "transparency" of belief, Alex Byrne and others have argued that self-knowledge results from the same capacities that give rise to first-order states, but it does not result from the very same exercises that give rise to those states (Byrne 2005, 2011, Setiya 2011, Fernandez 2014, Valaris 2014). Thus, thanks to my capacity for inference I can come to believe that p. That very capacity can then be put to use to discover that I believe that p. Byrne refers to such a view as "economical" (2005).

There are two reasons why this objection to Shoemaker is worth mentioning in the context of the chapters to follow. First, while I am inclined to believe that self-blindness is impossible, at least for the propositional attitudes, I do not assume that here, nor does it play a role in the arguments offered. So the problems facing Shoemaker's arguments do not threaten

Constitutivism as I develop it. Second it is worth mentioning the economical view here, since it will be a target throughout this dissertation, though not always directly. Constitutivism gained attention in philosophy largely as the result of Shoemaker's arguments from self-blindness. The primary target of those arguments was the Inner Sense model of self-knowledge, that is, the view on which self-knowledge is the result of a quasiperceptual faculty, a kind of inner eye, or simply a reliable belief-forming disposition, conceived as a capacity distinct from those involved in forming first-order states. While the impossibility of self-blindness may not entail Constitutivism, it does plausibly rule out this family of views. At the time when Shoemaker wrote those essays it may have looked like Constitutivism and the Inner Sense exhausted the possibilities. But the development of versions of the economical theories in recent years has shown that this is not so. A supposed advantage of those theories is that they can explain everything explained by Constitutivism (e.g. the impossibility of self-blindness) while avoiding what are thought to be its excesses. However, as I explain further in chapters two and four, this is not so. There are features of first-personal self-knowledge that only Constitutivism can explain.

In summary, the version of Constitutivism defended here holds that the epistemology of mind must be explained by facts about the constitution of mental states. Sydney Shoemaker has defended a version of this view. However, Constitutivism cannot be established on the basis of the arguments he has offered. I pursue a different, and more dialectically satisfying, strategy. Rather than arguing for Constitutivism from theories about psychological concepts or the impossibility of self-blindness, I start with plain observations about features of first-personal self-knowledge and first-order states. Thus, I take it, the starting points of the arguments to follow are less controversial than those of other familiar arguments for Constitutivism.

0.3 SELF-INTIMATION AND HUMAN FALLIBILITY

Many philosophers find Constitutivism implausible. While some will object to Authority, even more will object to Self-Intimation. That is because it makes a limited *omniscience* claim: for a class of mental states, if we are in one of those states we cannot fail to know that we are. This is considerably bolder than the *infallibility* claim made by Authority. Many epistemological views endorse the possibility of infallible justification. Infallibilists, of course, demand it for all knowledge. Disjunctivists hold that seeing that p yields infallible justification for perceptual beliefs. By contrast, many philosophers will deny that we are omniscient about any nontrivial conditions about ourselves. Thus, Timothy Williamson repeatedly calls versions of this idea "a quaint relic of Cartesian epistemology" (2000, 193). Others think that Self-Intimation cannot be squared with the myriad ways in which we are self-ignorant and self-deceived, as catalogued by social psychology. Thus we learn that people routinely mistake late night thirstiness for hunger; they confabulate in explaining their preference for one of two identical shirts; if primed to expect to be touched by a hot object they momentarily believe that it is hot when it is in fact cold, and so on. As Nisbett and Wilson put it: "people may have little ability to report on their cognitive processes" (1977, 247). Do these findings sap Constitutivism of all plausibility? Does Constitutivism wrongly portray our minds as always open to our unerring inspection?

Here are two responses to the worry that Constitutivism is implausible on broadly empirical grounds.[7] First, from the fact that a (no doubt surprisingly large) class of mental processes are opaque to the subject, it does not follow that they all are. As I mentioned in

[7] A point of orientation: the project pursued in the pages to follow is not primarily defensive. I do not aim to defend Constitutivism against a battery of counterexamples and objections. While I do consider many objections, the bulk of the effort is put into providing arguments for accepting the view. It will help, though, to locate the view within the broader terrain in which these worries arise.

qualifying Self-Intimation, the Constitutivist does not hold that all mental states are self-intimating, nor even that anything one might call a "belief" is self-intimating. The Constitutivist only holds that there is a class of mental states that are self-intimating, perhaps a small one. To identify this class, she must complete what I called the demarcation task. Here I only make a start at it, arguing that inferential belief is self-intimating. However, the argument there gives some indication of what this broader class may look like. Inference necessarily involves the subject taking her premises to support her conclusion. One of the strongest motivations for endorsing this idea is that it articulates a sense in which the subject is active in believing for reasons. I argue that we can hold onto that picture of inference as active if we endorse Self-Intimation. And one might think that this is merely an instance of a more general connection between mental agency and self-knowledge. So one strategy for pursuing the demarcation task further would be to consider the sense in which the various intentional attitudes are active. Given that only a relatively small amount of mental processes are active, this might leave room to reconcile Constitutivism with the empirical results.

Second, the Constitutivist *can* accept the possibility of impairments in our ability to know specific states of mind, but she must explain these as resulting from problems with *the first-order states themselves*. For example, a self-deceived subject is temporarily ignorant of one of her beliefs; the motivational forces at work in self-deception impair her ability to know this belief. But, plausibly, they also impair the rationality of the belief itself: a self-deceived subject will not be disposed to affirm the content of her belief or employ it in her reasoning. Another example: the occasional difficulty in knowing one's emotions may be explained by the possibility that sometimes one is not determinately in any particular state of mind. The idea here is that some attitudes come in degrees (Lear 2005). In the early stages of anger it may be difficult to tell

whether one is resentful or jealous because one's state of mind is not determinately one way or the other. That self-knowledge is difficult in these cases does not reflect negatively on Self-Intimation, which we may take to be restricted to mental states of suitable rationality, determinacy, and maturity. In general, the Constitutivist strategy will be to explain difficulties in knowing mental states by appeal to deficiencies in the states themselves, thus maintaining the connection between self-knowledge and first-order states, though negatively. Indeed, one might think that this defensive strategy, if successful, would lend considerable support to Constitutivism. As Brian O'Shaughnessy puts it, the strategy here is "to demonstrate that 'a well-formed state of self-conscious wakefulness is such that the present contents of that mind must be insightfully given to its owner', through demonstrating that 'a wakeful subject significantly self-ignorant must be improperly conscious or awake'" (O'Shaughnessy 2000, 113.)[8]

On the other hand, to the first objection I think the Constitutivist must simply plead guilty. Constitutivism is a Cartesian view in that it holds that some of our mental states are conditions we cannot be in without knowing it. The problem with Descartes' view, or at least, the view that has been dubbed "Cartesian" by philosophers, is that it accepts both Constitutivism and the thesis I earlier called Independence, here in the form of a broadly perceptual epistemology of mind. That is, on this view, when I believe that p I am guaranteed to know it, and I come to know it through an act of inner perception. But that cannot be. Any capacity for perception, inner or outer, is by its nature *fallible*. Hence self-knowledge cannot be both perceptual and constitutive (Shoemaker 1996, Bilgrami 2006, 12-16.) The Constitutivist resolves this by rejecting Independence. By contrast, the perhaps standard response to this difficulty is to retain Independence while denying that self-knowledge is constitutive and infallible.

[8] This strategy is pursued by O'Shaughnessy in his perceptive descriptions of the states of drunkenness and madness (ibid 125-153).

My point here is to make the following suggestion: however one develops this latter response it will be well placed to explain human fallibility. But it comes at the cost of losing touch with the *uniqueness* or *specialness* of self-knowledge. For all that is left for the opponent to think this uniqueness consists in is a very high degree of justification or reliability and a unique method for achieving it. Thus, Alex Byrne writes that explaining first-person authority consists in explaining why my beliefs about my mind are more likely to be knowledge than my beliefs about your mind (Byrne 2005, 81). In a number of places in what follows I will refer to this as the picture of the first-person perspective as endowing the subject with a kind of "self-expertise." While it can seem that the Constitutivist is hopelessly optimistic about our self-knowledge, I will argue that the picture of self-expertise mischaracterizes our relation to our own minds. Earlier I tried to bring this out by saying that, often, we do not have to *discover* or *learn* what we believe, desire, and intend. We aren't in the position of having to guess what is going on with us. In the next three chapters I survey a number of ways in which first-personal self-knowledge is fundamentally unlike other ways of knowing about oneself. The model of self-expertise is poorly positioned to explain these and should be rejected.

I take this to mean that there is, at least on first blush, a tension between our fallibility and what is distinctive about self-knowledge. And in attempting to understand self-knowledge we should aspire to reconcile these two ideas rather than simply deny one of them. Yet the project of reconciliation is considerable and can only be attempted with a proper appreciation of both sides. The goal of this dissertation is the relatively modest one of getting clear about the ways in which self-knowledge is special and arguing for a view that can explain them.

1 TESTIMONY AND THE RATIONAL SIGNIFICANCE OF SELF-KNOWLEDGE

1.1 A PUZZLE ABOUT TESTIMONY

We possess a way of knowing our minds that does not rely on observation or inference from behavior: a way that is only available from the first-person perspective. Call this first-personal self-knowledge.[9] Of course, other ways of knowing our minds are available to us. Like other matters of fact, we can come to know of our beliefs, desires, and other intentional attitudes on the basis of the testimony of others.[10] Yet, intuitively, knowing our minds in this way is inferior to knowing them first-personally. That is, first-personal self-knowledge seems to be valuable in a special way. But what is the source of this value and why does testimonial knowledge of one's mind lack it? Call this the Puzzle about Testimony. These questions are puzzling given that instances of both first-personal and testimonial self-knowledge can have the same content (e.g. "I am angry") and both can count as knowledge. (As we'll see, the puzzle extends beyond the case of testimony, but it is best to focus on this case to begin.)

One might think that the puzzle can be solved easily. It is problematic to rely on testimony about one's own mind because in the standard case we possess or are in a position to

---

[9] Unless otherwise noted, when speaking of "self-knowledge" and "self-belief" I will have this first-personal form in mind.

[10] My focus throughout will be on self-knowledge of the nonfactive intentional attitudes, including the emotions. See Boyle 2009 for an argument that the epistemology of phenomenal and intentional states should be given separate treatments.

possess first-personal self-knowledge. Thus it might reasonably be assumed that something has gone wrong when one has to rely on testimony. But this just means that something has gone wrong with our capacity for first-personal self-knowledge. Compare: suppose that I typically find out my roommate is home by seeing her pull into the driveway. If I'm temporarily blinded I may have to rely on my hearing. Given the normal run of things, that I have to rely on another source of knowledge reveals that there is a problem with my vision. But this doesn't tell us anything interesting about knowledge by vision or knowledge about one's roommate's whereabouts.

But first-personal self-knowledge is importantly different. It is not merely typical, but valuable in itself. When it comes to our attitudes, testimony is not an equally valuable, though atypical, source of knowledge. There is something problematic about having to rely on testimony in order to know one's mind regardless of what ordinary conditions are like. The subject who relies on testimony about his mind is, in some way, in an inferior position to the subject who possesses first-personal self-knowledge.

Consider the following example.

A and B are friends discussing their plans for the winter break. After A details the itinerary of her vacation to the Caribbean with evident joy and excitement, B describes his own plans to spend time with in-laws in some dreary and cold part of the country. As he mentions the uncertainties and frustrations associated with the trip, it becomes clear to A that B is angry. Perhaps he is angry at his in-laws, at his wife, or at all of them. He is clearly angry at someone, and A tells him this. But B resists, citing run of the mill holiday stress. A insists, saying, "Trust me. I know when you are angry, and you are angry now." Because he trusts her, B accepts A's testimony and comes to learn a fact about himself.

I assume that everyone is familiar with examples like this. We can imagine any number of similar cases involving different relationships, different mental states, and issues of varying significance. It should be clear, I think, both that B can gain knowledge by means of A's testimony and that this knowledge is in some way inferior to self-knowledge. Intuitively, it would be better for B not to have to rely on the say-so of his friend.

The Puzzle about Testimony can also be felt in therapeutic practice. Freud has it in mind in the following intriguing passage from the *Introductory Lectures on Psycho-Analysis:*

> From what I have so far said a neurosis would seem to be the result of a kind of ignorance – a not knowing about mental events that one ought to know of... Now it would as a rule be very easy for a doctor experienced in analysis to guess what mental impulses had remained unconscious in a particular patient. So it ought not to be very difficult, either, for him to restore the patient by communicating his knowledge to him and so remedying his ignorance…
>
> If only that was how things happened! We came upon discoveries in this connection for which we were at first unprepared. Knowledge is not always the same as knowledge: there are different sorts of knowledge, which are far from equivalent psychologically… The doctor's knowledge is not the same as the patient's and cannot produce the same effects. If the doctor transfers his knowledge to the patient as a piece of information, it has no result… The patient knows after this what he did not know before – the sense of his symptom; yet he knows it just as little as he did. Thus we learn that there is more than one kind of ignorance. We shall need to have a somewhat deeper understanding of psychology to show us in what these differences consist.
>
> Freud 1966, 347-348

Presumably one of the goals of therapy is self-knowledge. A patient is unwilling to recognize the operation of unconscious phantasy in his life or is otherwise unaware of some state of his mind. He is suffering in part because he is ignorant of his mind: curing him of the neurosis requires curing him of his ignorance. But, Freud tells us, merely possessing knowledge of oneself as a piece of information is insufficient to affect the cure. As we might put it, vaguely but helpfully,

testimonial self-knowledge (the kind the analyst is in a position to transmit to the patient) yields one an outsider's perspective on oneself, while therapy aims at knowledge of oneself from the inside. We can mark this contrast by referring to the self-knowledge that therapy aims at "self-knowledge" and other forms, including knowledge by testimony, "knowledge of oneself."

It seems fair to assume that the self-knowledge at which psychoanalysis and other forms of talk therapy at least sometimes aim is first-personal self-knowledge. Neurosis and other forms of psychic distress plausibly involve a disruption of one's ability to gain such knowledge, either because something like repression interferes with a subject's ability to know her mind or because the state in question is, by its nature, resistant to being known in this way.

Now it is not my concern here to determine whether a particular form of self-knowledge is a necessary component of psychic health, or why that might be, interesting as those questions are. Rather, I want to understand why, in general, knowledge of oneself by testimony is inferior to self-knowledge and what this tells us about the latter. As the case of A and B makes clear, Freud's point applies to fairly mundane examples. We needn't posit unconscious phantasy or repression to make sense of cases where a subject is blocked from accessing her mind. In a state of confusion or intoxication I may lack awareness of my mind; someone who knows me well can tell me what I want, believe, or intend, providing me with knowledge by testimony. Yet this knowledge is in some way inferior to self-knowledge. What is intriguing about Freud's observation is the suggestion that the inferiority is not *epistemic*. Rather, testimonial knowledge of oneself fails to play a non-epistemic role in the psychic life of the subject that self-knowledge can. While the patient can come to know of her state of mind, treatment requires that she know of it in a particular way.

The goals of this chapter are as follows. First, in line with Freud's remarks, I will argue that the value of self-knowledge is not epistemic, that is, it cannot be explained in terms of epistemic asymmetries between the first and third-person perspectives. Second, I will identify the non-epistemic role played by self-knowledge that gives it its special value and explain why, for failing to play this role, testimonial knowledge of oneself is intuitively problematic. More specifically, I will argue that accepting the following thesis can explain the special value of self-knowledge:

> *Rational Significance:* A first-personal self-belief that "I am in M" can play the roles in reasoning typically played by M.[11]

When I know that I am angry, I am thereby in a position to reason with and act on my anger. As I will put it, first-personal self-knowledge enables a subject to reason self-consciously with a mental state. I call this phenomenon the rational significance of self-knowledge. Testimonial knowledge of oneself is not rationally significant. Thus, a subject who had to rely on such knowledge would be prevented from reasoning self-consciously with her mental state.

The third goal of the chapter is to consider the consequences of accepting Rational Significance for our understanding of self-knowledge. Rational Significance holds not only of first-personal self-knowledge but also of first-personal self-beliefs. As we can put it, a mark of a first-personal self-belief is its rational significance. If Rational Significance is true then a subject who believes that she is in a mental state will be disposed to reason in ways subjectively rational given that she is in that state. On a plausible account of the nature of the intentional attitudes, to possess such a cluster of dispositions is *sufficient* for being in that state. Therefore, it follows from Rational Significance that self-beliefs are partially constituted by the first-order states

---

[11] I discuss the idea of a typical role in section 5 below.

represented in their content and so are guaranteed to be true. That is, I will argue that if we accept Rational Significance then we should accept the following:

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

Authority claims that self-beliefs are metaphysically sufficient for first-order states. It is a core commitment of Constitutivist accounts of self-knowledge on which it is explained by constitutive relations between first and second-order states. Thus, solving the Puzzle about Testimony lends support to Constitutivism.

I proceed as follows. In the next section I elaborate upon the Constitutivist view I favor as well as the assumptions I make in arguing for it. In parts 3 and 4 I argue against rival solutions to the Puzzle about Testimony. In part 5 I explain how Rational Significance solves the puzzle and offer independent support for it. In part 6 I argue that if we accept Rational Significance then we should also accept Authority. I conclude by considering objections.

1.2 CONSTITUTIVISM ABOUT SELF-KNOWLEDGE

Self-knowledge is special for at least two reasons. First, I can know my mind in a way unavailable to anyone else. In Alex Byrne's terms, my self-knowledge is possessed thanks to a source *peculiar* to the first-person perspective (Byrne 2005). Second, this self-knowledge is *privileged* in that it is not subject to standard forms of epistemic criticism (Hampshire 1979, Byrne 2005, Bar-On 2006). Typically, when I speak about my mind from the first-person perspective others owe me a kind of deference that is not owed me when I speak about the minds

of others. A philosophical account of self-knowledge ought to explain these features of self-knowledge.

The Constitutivist account is surprisingly simple: for a class of mental states, a subject knows that she is in one of those states simply by being in it.[12] Since only the subject herself can know that she is in a mental state by being in it, the Constitutivist explains the peculiarity of self-knowledge with ease. Likewise, since self-ascriptions are constituted by the very states they ascribe, the Constitutivist is committed to a limited infallibility claim (more on which below), which explains the epistemic privilege of self-knowledge. More specifically, the Constitutivist view is made up of two claims:

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, one thereby is in M.
>
> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.[13]

The restriction to first-personal self-beliefs leaves room for self-ascriptions that do not fall under Authority, such as those based on self-observation, inference, or testimony. Self-Intimation makes a similar restriction and so avoids the obviously false claim that all mental states are self-intimating. While it is controversial which states are self-intimating, I won't take that question up here since I'll be focusing on Authority.[14]

Authority claims that, in a class of cases, if a subject believes that she is in a mental state it just follows that she is. And that might make one worry that the view is committed to the Cartesian claim that we are infallible about our minds. But the point is subtle. We should

---

[12] Versions of Constitutivism are defended by Shoemaker 1996 and 2012, Heal 1996, Wright 1998, O'Shaughnessy 2000, Bilgrami 2006, Zimmerman 2006, Rödl 2007, Boyle 2011, and Coliva 2012.

[13] Bilgrami 2007 and Coliva 2012 refer to Self-Intimation by the name "Transparency." However, given that that term has been used so widely in recent philosophy and for such different purposes, it seems best to use a different name, one which more clearly conveys the content of the thesis.

[14] See Bilgrami 2007, Shoemaker 2012, and Coliva 2012 for discussion, as well as chapter four below.

distinguish two forms of infallibility: infallibility relative to contents and infallibility relative to sources of justification. A subject is infallible relative to a content X when her beliefs about X cannot be mistaken. By contrast, if she is infallible relative to a source of justification J, then her beliefs about X cannot be mistaken *so long as they are based on J*. Authority is only committed to the latter, but I take it that the problematic view is the former. Even if Authority is true, some of a subject's self-ascriptions of belief will be false, such as some of those based on observation or inference, so Authority doesn't entail that all of our self-ascriptions are infallible. Furthermore, Authority leaves room for the possibility of a subject who is mistaken about whether her self-ascription is a genuine first-personal self-belief. Perhaps this happens in some cases of self-deception.[15]

It is important to note that Authority is a claim about the metaphysics of a kind of mental state – first-personal self-belief – and does not trivially follow from the factivity of knowledge. Self-beliefs are guaranteed to be true because they are partially constituted by the states they represent. But how should we understand these claims of constitution?

I assume that beliefs and other intentional attitudes are clusters of dispositions to behave and reason in a variety of patterned ways.[16] Thus, possessing some conjunction of the following dispositions is sufficient for believing that p: to act in ways intelligible given that p, to affirm that p in speech and thought, to use p as a premise in practical and theoretical reasoning, to believe some of the logical consequences of p, to consciously judge that p, to entertain p in inner speech, to revise beliefs inconsistent with p, and the like. Notice that one can accept this general picture without settling the question whether these dispositions can be specified in a noncircular way. We can also accept the sufficiency claim without settling the question whether any of these

[15] I discuss self-deception further in chapter four.
[16] See Baker 1995 and Schwitzgebel 2013 for defenses of this idea.

dispositions are essential or necessary for belief. It also seems safe to assume that a person can possess a belief without possessing all of these dispositions. And finally, the term "belief" as it is ordinarily used may refer to states of varying degrees of cognitive sophistication. Even granting this, we should recognize that there are certain dispositions that are sufficient for the intentional attitudes typical of rational subjects. Such states possess conceptual contents, are typically formed on the basis of reasons, and play distinctive roles in reasoning. Thus, some combinations of dispositions to use p as a premise in reasoning, to affirm it in speech and thought, and to act on the basis that p are plausibly *sufficient* for possession of a rational, conceptual belief that p. Similar points plausibly hold of desires, intentions, and the emotions.

Now this view, even in broad outline, allows us the following thought: if, in virtue of being in a mental state M1, a subject possesses dispositions x, y, and z, and x, y, and z are sufficient for being in mental state M2, then M1 is partially constituted by M2. For example, suppose that to desire to φ is simply to be disposed to φ. Given that an intention to φ also involves such a disposition, it would follow that intentions are partially constituted by desires, so that if one intends to φ one thereby also desires to φ. Thus, intending is a way of desiring. (I am not endorsing this view of desire, which is obviously oversimplified, just setting it out as an example.)

This suggests a strategy for establishing Authority: show that a class of self-beliefs are constituted by dispositions that are sufficient for possession of the first-order states that are represented in their contents. This is the strategy pursued in the following. If Rational Significance is true then self-beliefs are partially constituted by dispositions to reason in a way

subjectively rational given that one is in the relevant first-order state.[17] Such a disposition is sufficient for being in the first-order state and so if one believes that one is in a mental state one is *thereby* in that state. As I will put it, believing that one is in a mental state is a way of being in that state: being in it self-consciously.

It is worth emphasizing that this strategy differs from what is standardly pursued in the contemporary literature. In discussions of self-knowledge it is common to start by pooling data about the epistemic features of self-knowledge and then propose a mechanism or capacity that plausibly yields knowledge with these features. That is, the primary question is: how do we arrive at beliefs with these epistemic features? While, as we have seen, Constitutivism can explain for the distinctive epistemic features of self-knowledge, I want to postpone this question. Instead, I start with a distinction in kinds of self-beliefs and ask in what their difference could consist. I'll argue that differences in the value we place upon these beliefs are explained by a metaphysical difference between them: a difference in the rational dispositions that constitute them.

However, before defending this view, I'll consider alternative solutions to the Puzzle about Testimony.

[17] By "subjectively rational" I mean an action or thought that is rational given the state of mind of the subject, independently of the objective correctness of that state of mind. For example, it is subjectively rational to retaliate against those who you believe have wronged you, even if that is not what one objectively ought to do. I remain neutral on the source of what one objectively ought to do or think.

1.3 THE EPISTEMIC ADVANTAGE OF SELF-KNOWLEDGE

Most philosophers accept that, other things being equal, when a subject self-ascribes a mental state she has some kind of epistemic advantage over others. There are many different versions of this view. For our purposes it will suffice to say that it is generally easier for my beliefs about my mind to count as knowledge than yours (Byrne 2005). This might be thought to offer an explanation of what goes wrong with B: his knowledge of his anger is inferior because it lacks the credentials that self-knowledge typically possesses. While he can learn of his mind by testimony, testimony can't come close to lending the kind of support that a capacity for self-knowledge can, support that *only* such a capacity can lend. Let's call this the Epistemic Solution. As we might put it, on this view, each subject is something like an expert on herself. To have to rely on the say-so of another is to forgo the expertise that one possesses. B's situation is then analogous to other cases of an expert deferring to the say so of a non-expert. B's deference to A is problematic for the same reason it would be problematic for a mechanic to take my advice about how to fix my car. It might be the right thing to do in the circumstances, but given that he is the expert, it would surely be better for him to rely on himself.

I think we should reject the Epistemic Solution because this asymmetry, however we understand it, cannot explain the inferiority of knowledge of oneself by testimony. That is because the inferiority persists even when symmetry is regained.

Suppose that instead of a conversation with his colleague, B is discussing his plans for the break with an omniscient (and honest) God.[18] The case is otherwise the same as that involving A and B. B describes his plans and manifests his anger. This God, being omniscient,

[18] I owe this example to Ulf Hlobil.

knows that B is angry. B is unaware of his anger, but, reasonably enough, accepts the God's testimony. Now even if self-knowledge is an epistemically privileged source of knowledge, it seems safe to assume that it cannot yield beliefs that are *better* justified than those relying on the testimony of the omniscient God. At best, self-knowledge yields beliefs that are as well justified as those that rely on the testimony of the God.[19] That means that if the Epistemic Solution is right, then there ought to be nothing problematic about the case involving B and the omniscient God. Remember, on this interpretation, what goes wrong with B in the original case is that he accepts testimony when a better source of justification is or should be available to him. But if the God's testimony provides as much justification as self-knowledge, then it ought not to matter on which source B bases his self-belief.

However, I contend that the presence of the omniscient God does not alter the case in any significant way. Even though the God surely knows that B is angry and is an infallible source of testimony, there is still something problematic about the fact that B has to learn of his mind from another. If this is right, then the Epistemic Solution does not offer a viable explanation of the inferiority of knowledge of oneself by testimony.

---

[19] One might argue that because it relies upon the fallible belief that he is speaking with an infallible God, B's testimonial belief is epistemically less secure. Of course, our self-beliefs can be false (for example, when they are based on observation or inference) and so there is fallibility in both cases. However, one might insist that in the case of self-knowledge we possess fallible access to infallible grounds, while in the testimonial case our grounds are themselves fallible. Thus, the difference between testimony and self-knowledge lies in the kind of grounds in each case. While this would represent a significant difference, I reject the interpretation of the case. The God's testimony is infallible; one's belief that one is receiving testimony from an infallible god is fallible. So here too we have a case of fallible access to infallible grounds. (I thank Kieran Setiya for alerting me to this possibility.)

1.4 THE SHORTCOMINGS OF TESTIMONY

We can't explain what goes wrong with B in terms of the epistemic merits of self-knowledge. But perhaps we can explain it in terms of the demerits of testimony. For surely there is something less than ideal about second-hand knowledge. Is B's problem simply an instance of that? Call this the testimony solution. I'll argue that this solutions fails to explain what is intuitively problematic about B's situation.

One might think that testimony is inferior because the hearer is dependent upon the speaker's reasons. On an attractive view of testimony it transmits but does not generate justification. The bare fact that Jones tells me that p doesn't provide me with justification for believing p. Rather, I am justified in believing what Jones says only if he himself is justified, and so my justification depends upon Jones' reasons, about which I am ignorant. However, as I explain in more detail below, B's situation is not improved by gaining knowledge of the testifier's reasons. Therefore, ignorance of the testifier's reasons cannot be the source of his problem

On another view, the problem with accepting testimony is that, precisely because the listener is ignorant of the speaker's reasons, the listener lacks *understanding* of what she believes. This idea has been developed in the context of the recent debate about moral testimony. Alison Hills has argued forcefully that there are reasons against accepting moral testimony on the ground that it fails to transmit moral *understanding*, which is required for morally worthy action and a morally good character (Hills 2009). Understanding some proposition requires both knowing the reasons for it – knowing why it is the case – and grasping how those reasons support that proposition. It is the latter that cannot be transmitted by testimony. Jones can tell me both that p and that q supports that p, and I can come to know these things from him. But I do not

thereby grasp how it is that q supports that p. Thus, absent further reflection and reasoning, I lack understanding of the truth that p.

Notice that the point that testimony doesn't transmit understanding generalizes beyond the moral case. If my grandmother tells me to plant garlic in late autumn because the growing season is short, I can gain knowledge, but I do not deepen my understanding of gardening. (The testimony might lead me to reflect and gain this understanding myself, but it is not transmitted by the testimony.) Hills' point is that because of the importance of specifically *moral* understanding there are reasons to refrain from accepting moral testimony that don't apply to other cases. This suggests an explanation of the problem with B: understanding is not transmitted by testimony, so B lacks understanding about his anger. This is attractive because it is natural to think that the aim of therapy is not merely self-knowledge but some kind of *self-understanding*, which, like moral understanding and unlike garlic understanding, is an important element in a well-lived life. Therefore, because of the importance of self-understanding, there are strong reasons against gaining self-knowledge on the say-so of others.

However, there is an ambiguity in the phrase "understanding about his anger." As we saw, understanding requires both knowledge of reasons and a grasp of the relation between those reasons and the proposition understood. The ambiguity is this. In understanding his anger, does B grasp the reasons *for believing that he is angry* or the reasons *for being angry*? Taking the analogy with the other cases of testimony seriously would suggest that it is the former. Suppose I learn by testimony that some leaves change color in the fall. In order to understand this I need to know why they change and to grasp the connection between the reason and the proposition. Obviously, in this case I learn of a reason for its being the case that leaves change color, not a reason *to change color*. Likewise, to supplement B's testimonial knowledge with understanding

would involve a kind of theoretical or interpretive grasp of his own psychology, the kind A presumably possesses.

However, as Freud makes clear, the problem with B is not a lack of *that* kind of understanding. His problem is the absence of something uniquely first-personal, something A cannot possess. B's relation to his anger is deformed. This deformation is not corrected by theoretical knowledge, what Freud calls mere "information" about himself and "the sense of his symptom." This is shown by the fact that other forms of knowledge of oneself are in as bad a shape as B's testimonial knowledge. Suppose B interrupts his description of his plans for the break to offer the following self-diagnosis: "You know, A, looking back over my behavior the last few holidays spent with family suggests that I get angry at these times and I bet that I am angry now." Suppose that B marshals a convincing amount of evidence in support of his hypothesis, revealing that he not only knows that he is angry, but he also understands that he is. It seems clear that B's knowledge of himself is lacking in the same way as his testimonial knowledge of himself. If it is not clear imagine that B is so constituted that every day at 5pm he gets angry. Knowing this, and noticing that it is 5pm, B infers that he is angry. In the right circumstances, this can count as knowledge and it may very well manifest understanding. But, intuitively, something is still awry with B. Given that he possesses the kind of understanding that testimony cannot transmit but his problem persists, it follows that his problem cannot consist in a lack of understanding, and so cannot be traced back to the demerits of testimonial knowledge in general. The Testimony Solution fails.

1.5 RATIONAL SIGNIFICANCE

I have argued that knowledge of oneself by testimony is inferior to self-knowledge and that this inferiority cannot be explained by appeal to epistemic differences between the two. In this section I present my solution to the Puzzle about Testimony.

The Testimony Solution seemed plausible because it seems right to say that the special value of self-knowledge is that it affords us self-understanding. But there is an ambiguity in the idea of self-understanding. In understanding one's anger one can grasp either reasons for believing one is angry or reasons for being angry. The Testimony Solution can appeal only to the former. My suggestion here will be that self-knowledge involves the latter kind of understanding. As we might put it, possessing self-knowledge involves both knowing a state of mind and *consciously occupying the perspective* of that state. For example, suppose that you are at a diner, catch a glimpse of a slice of key lime pie in a rotating jewel case, and think "I want some pie!" This thought expresses both self-knowledge and also one's attraction to the pie. Similarly, often, when I judge that I believe that p I at the same time take it that it is true that p. (This kind of understanding might involve knowledge of reasons, as in the case of belief, though it needn't, as the case of desire makes clear.) The problem with mere knowledge of oneself, testimonial or otherwise, then, is that it cannot place the subject in the right kind of relation to the *objects* of the mental states thus known. By contrast, when one possesses self-knowledge one stands in the right kind of relation to those objects, a relation that enables one to reason about them, and so to reason with the state known. Returning to our example, if B possessed self-knowledge then he would have a particular outlook on the object of his anger and would be positioned to act accordingly. But this is not the case when he possesses mere testimonial knowledge of himself. This is clear from the fact that it would not be subjectively rational for B to act on his anger upon

gaining testimonial knowledge. That is, it would not make sense for him to reason as follows: "well, A, you must be right, I'm angry at someone. I suppose I'll start shouting." On the other hand, if B were to possess self-knowledge it would be perfectly sensible (if, perhaps, not otherwise appropriate) for him to shout given his self-knowledge.

The suggestion is that self-knowledge enables a subject to see the world through or in terms of her mental state. Seeing it this way enables her to reason and act in certain ways, ways that are appropriate or sensible given that outlook. If this is right, then we should accept the following:

> *Rational Significance:* A first-personal self-belief that "I am in M" can play the roles in reasoning typically played by M.

Rational Significance captures the intuitive idea that, from the first-person perspective, there is an intimate connection between the self-ascription of a mental state and occupying a particular stance on the world.

It is important to emphasize that Rational Significance holds of first-personal self-*beliefs* and not just self-knowledge. That is, we needn't assume that the self-belief is true in order for it to be rationally significant. First-personal self-beliefs are not rationally significant in virtue of being *knowledge* but in virtue of the kind of mental state they are, understood in terms of the role they play in our mental economy, a role played by self-beliefs. This point is somewhat delicate, for I will argue that first-personal self-beliefs are constituted by the states represented in their contents and so guaranteed to be true. This means that first-personal self-beliefs cannot be false and so, plausibly, cannot fail to be knowledge. Thus, if I am right, there are no mere first-

personal self-beliefs. But there is at least a notional distinction to be made, which is required by the argument.[20]

The principal reason for accepting Rational Significance is that it solves the Problem about Testimony. Self-knowledge is valuable to us because of the role it plays in reasoning. When I know that I am angry first-personally I am therefore positioned to reason with my anger, to act on it self-consciously. Knowledge of oneself does not afford this possibility, at best it enables a subject to reason *about* her mental state. Returning to our example, on the basis of A's testimony B can draw conclusions from the fact that he is angry, practical and theoretical, but he cannot reason from anger.

And the problem with lacking this ability should also be apparent. It helps here to reflect on the idea that the intentional attitudes of rational subjects can be understood as commitments made by the subject.[21] For example, to believe that p is to be committed to its being true that p; to intend to φ is to be committed to φ-ing; to be angry at X is to take it that X has harmed or offended one in some way.[22] The idea here is that, typically, in possessing a mental state a subject takes a stance on how the world is or how she will behave, and she is thereby responsible for defending that stance in the face of scrutiny. If this is right, then knowledge of oneself is

---

[20] John McDowell has suggested to me that a Constitutivist can grant that our capacity for first-personal self-knowledge is fallible, occasionally yielding false self-beliefs. It is tempting to think that this is not so. According to Constitutivism, the way of possessing first-personal self-knowledge is simply being in a first-order state. So it's hard to see how that capacity could be exercised otherwise than by being in the relevant state, hence hard to see how the capacity could be fallible. However, I have come around to suspect that this may be the wrong way to put things, for two reasons. First, for reasons discussed further in chapter two, on a Constitutivist view, it is best to think of self-knowledge as a *modification* of our capacities for first-order belief formation, a way of forming beliefs (and other states). If that's right, then it is possible that defective exercises of the capacity to form beliefs will fail to yield a belief, but may yield a mistaken self-belief. An example: if there are "mock" or empty thoughts then one can take oneself to believe that "*that* chair is broken" when one believes no such thing since there is no such thought to think. Less controversially, one might think that in cases of self-deception one's capacity for first-personal self-knowledge is exercised defectively. Although I present the strong Infallibilist line here, the arguments can equally be made in other terms.

[21] See Collins 1987, Brandom 1994, and Bilgrami 2006 for developments of this idea.

[22] While it is controversial whether emotions and desires involve judgment or belief, it is plausible that they involve some kind of normative outlook on the world. See Benbaji 2013.

inferior because it involves being *alienated* from the normative perspective embodied by the state known. One knows *that* one has undertaken a commitment of some kind without endorsing that commitment. To be in a position where one must rely on knowledge of oneself is to be in a kind of divided state. No such alienation or division occurs in self-knowledge: self-knowledge involves a unified outlook.

Further support for Rational Significance can be found in some of Wittgenstein's discussions of Moore's paradox. Consider the following. In ordinary conversation a self-ascription of belief can function as an assertion of the proposition believed. One way, among many, of telling you that it is raining in Cleveland is to say that I believe that it is.[23] To explain this phenomenon, Wittgenstein considers an analogy with a description of a picture. Describing a photograph or painting can also be a way of describing the scene depicted. However, in that case it is necessary to consider whether the photograph is reliable. As Wittgenstein points out, if the analogy was a good one, "I should have to be able to say: "I believe that it's raining, and my belief is trustworthy, so I trust it" (Wittgenstein 1980, §483). But this doesn't make sense.

> Do you say, e.g. "I believe it, and as I am reliable, it will presumably be so"? That would be like saying: "I believe it – therefore I believe it."
>
> Ibid.

In the case of the photograph there is, first, recognition of what is depicted and then a consideration of its accuracy. But, typically and perhaps in the default case, no such gap opens up in the case of belief. Outside of extraordinary circumstances, to recognize that one believes that p is already to be committed to the reliability of the belief. This normative perspective is built into the kind of self-knowledge one possesses in these cases.

[23] I discuss some of the pragmatic issues that arise here below.

Another reason to accept Rational Significance is that it provides a compelling explanation of Freud's observation. The idea here is that, in at least some cases, therapy aims at a form of self-knowledge that enables the patient to take rational control of features of her mental life that had eluded her. But then the knowledge it aims at must be rationally significant: it must itself put the patient in a position to reason with the mental state in question. Knowledge of oneself by testimony is not rationally significant in this way; that is why the analyst cannot effect the cure by offering an expert report on the patient's mind. The value of gaining rationally significant self-knowledge in the therapeutic context can be understood in different ways. First, insofar as it leads to the alleviation of suffering, it is of instrumental value. Second, it may be thought to be intrinsically valuable to have one's behavior determined by mental states within one's rational, self-conscious control as opposed to states that are alien to that perspective. Third, perhaps, as Stuart Hampshire has argued, such self-knowledge is valuable because it is a necessary condition on freedom (Hampshire 1975).

I will now consider two worries one might have about Rational Significance.

Rational Significance claims that a first-personal self-belief can play the roles in reasoning typically played by the first-order state represented in its content. What exactly are these roles? First, it clearly cannot be the case that self-beliefs and first-order states can play all of the same roles in reasoning. Given that self-beliefs and first-order states have different *contents* there are at least some roles in reasoning that one plays and the other cannot. (Indeed, there must be at least one: from the first-order state to the belief that one is in it.) However, it is plausible that first-personal self-beliefs can play the roles in reasoning typically played by first-order states. For example, arguably desires are self-authenticating in the sense that, under normal

circumstances, desiring to φ itself renders φ-ing subjectively rational (Schafer 2013, contra Scanlon 1998). Put another way, desire constitutively involves being disposed to act. Returning to an earlier example, if Rational Significance is true then your self-belief that you want to eat a slice of pie renders getting some subjectively rational: given that you have this belief it makes sense for you to eat the pie. And that does seem to be what's going on in the example. Likewise, if I first-personally believe that I believe that p, I am thereby in a position to reason to new beliefs supported by the proposition that p. Thus, I might think, "I believe that p, and q follows from p, so q." Of course, this is not always the case. When I learn from social psychology that when asked to choose between two identical objects I am biased in favor of the one on the right, it is not subjectively rational for me to reason in the following way: "I believe that objects on the right are better, so I'll choose this one." Rational Significance can explain this: because the knowledge is based on testimony, it does not involve consciously occupying the perspective of the state. From the perspective of my knowledge of myself the bias does not seem reasonable.

Now one might worry that it is never rational to reason from a self-ascription to a further first-order state. But I think this worry arises as a result of being misled by the pragmatics of self-ascriptions. Typically a person will assert "I believe that p" precisely to *withhold* commitment to the truth of p or the value of eating pie. As it is often put, asserting "I believe that p" is usually a hedged or qualified assertion that p. This might incline one to think that self-ascriptions *never* fully embody the commitment constitutive of the first-order state ascribed and so lead one to conclude that it is never subjectively rational to act on self-knowledge. However, from the fact that self-ascriptions are typically hedged assertions it does not follow that they always are. Rational Significance holds only that there is a way of knowing one's mental states that involves consciously occupying the perspective of the state. For a subject with this self-

knowledge there is a unity between what she knows herself to believe (or desire or intend or fear.) and what she takes the extra-mental facts to be. Given the typical pragmatic function of self-ascriptions it would be misleading for such a subject to express her opinion about p by reporting on her belief. She would more clearly express her commitment by simply saying "p." This explains why rationally significant self-knowledge is typically not expressed in speech. However, again, this is not always so; not all self-ascriptions are hedged assertions. An example would be a profession of faith like the Nicene Creed. These often involve self-ascriptions of belief ("I believe in one God, the Father Almighty, Maker of heaven and earth, and of all things visible and invisible").[24] But there is no question of the self-ascription functioning as either a hedged assertion or a mere description of one's mind. Of course, professions of faith are pragmatically complex, but it seems clear that self-ascriptions of this kind manifest a wholehearted commitment to the embedded content. And it would be subjectively rational for a subject to reason from such a commitment.

Now that I've presented my proposed solution to the Puzzle about Testimony we can consider another non-epistemic solution. In a discussion of Freud's observation quoted earlier David Finkelstein suggests that self-knowledge is valuable because it enables a subject to express the state known by avowing it (Finkelstein 2003, 121). If B knew that he was angry then his avowal "I am angry!" would be an expression or manifestation of that anger. By contrast, a report of one's testimonial knowledge of oneself is not an expression of anger.[25] The problem with B is

[24] I'm assuming that the "belief in" language of the creed entails belief that.
[25] I am not offering an account of expression here. See Finkelstein 2003, Bar-On 2006, and Green 2007 for details, as well as chapter three below.

that he is prevented from giving expression to his anger. The value of self-expression accounts for the intuitive advantage of self-knowledge.

I deny neither that there is an intimate connection between our capacities for self-knowledge and self-expression nor that this distinguishes self-knowledge from knowledge of oneself. However, the idea of expression does not offer a real alternative to Rational Significance, but is instead an instance of the more general thesis.

One can express a state of mind intentionally or not. Both avowing anger and getting red in the face can express anger, but ordinarily the latter is not intentional. It seems clear that the absence of self-knowledge is not likely to interfere with non-intentional expressions of a mental state. Even though he is not aware of his anger, in his discussion with A, B may very well get red in the face, grind his teeth, or manifest his anger in other ways. Thus, I take it that the absence of self-knowledge only inhibits one's ability to intentionally express one's state of mind. To express one's anger intentionally is to do something in the knowledge that so acting is a way of expressing the state of mind one knows oneself to possess.[26] If this is right then intentional expression is simply a case of reasoning self-consciously with a mental state, in this case, reasoning practically about what to do given one's state of mind. And this means that the expressivist solution falls within the scope of the solution I have offered.

In this section I've argued that Rational Significance solves the Puzzle about Testimony. The problem with having to rely on knowledge of oneself is that it places one in an alienated relation

[26] Some have rejected this claim. Dorit Bar-On argues that it is possible to express a mental state by means of an intentional action without intentionally expressing that state. Thus, throwing a rock through a window is intentional and expresses anger, but in throwing the rock I may not have as my goal the expression of my anger. Such cases are of course possible, but irrelevant to the point at hand, where the focus is precisely on cases in which one's goal is to express one's state of mind. See chapter three below for further discussion.

to one's own mind. Self-knowledge, by contrast, involves a kind of unity between one's stance on oneself and one's stance on the world. We should accept this view because it can solve our puzzle, but it is also intuitively plausible as is clear from reflection on examples. Finally, it is able to explain the observations of Wittgenstein and Freud about the way in which we relate to our own minds. To accept Rational Significance is to recognize that self-knowledge involves knowing one's mind while at the same time occupying a perspective on the objects of one's mental states. I have interpreted this as follows: the self-belief that one is in M is partially constituted by a disposition to reason or act in a way subjectively rational given that one is in M. In the next section I'll argue that Authority follows from Rational Significance, so understood.

1.6 AUTHORITY

The argument for Authority from Rational Significance is very straightforward. I argued that first-personal self-beliefs can play the roles in reasoning typical of first-order intentional attitudes. But a disposition to play those roles is sufficient for possession of the first-order state. Thus, from Rational Significance and the dispositional account of the intentional attitudes, we get the claim that a first-personal self-belief is constituted by a disposition that is sufficient for the possession of the first-order state that is represented in its content. A first-personal self-belief is metaphysically sufficient for possession of the first-order state, which means that we should accept Authority.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

Notice that one can accept Authority without endorsing the claim that self-knowledge is a metaphysically necessary condition on being in the relevant first-order states. That is, one can accept Authority without accepting what I have called Self-Intimation. Authority is a claim about the metaphysics of self-beliefs while Self-Intimation concerns the metaphysics of first-order states.

This argument for Authority crucially depends upon the claim that being disposed to reason in a way subjectively rational given that one is in a particular mental state is sufficient for possessing that mental state. Once we accept a broadly dispositional account of the intentional attitudes then we are committed to the idea that some cluster of dispositions is sufficient for possessing a given attitude. That is, possessing an attitude is a matter of one's dispositions matching to an appropriate degree what Eric Schwitzgebel calls a "stereotype" of the attitude, one grounded in folk psychology (Schwitzgebel 2013.) My contention is that rational dispositions are sufficient for the intentional attitudes of concept-wielding reasoners.

One might object to the claim that a rationally significant self-belief is sufficient for possession of a first-order state as follows. It is plausible that some intentional attitudes, particularly the emotions, have a distinctive phenomenology, such that a subject could not count as being, say, angry unless she was disposed to experience conscious episodes with a certain qualitative feel. If this is right then Authority would not follow from Rational Significance unless all self-beliefs were partially constituted by dispositions to experience the relevant phenomenology.

There are two possible responses to this objection. First, one might deny that phenomenology is a metaphysically necessary condition on any of the intentional attitudes. That is, one might insist that it is possible for a subject to be angry without being disposed to enjoy the

phenomenology we associate with anger. This is of course consistent with holding that there is a distinctive phenomenology that typically accompanies conscious episodes of anger, it only requires the possibility of anger without phenomenology. Once this possibility is granted we can claim that rationally significant self-beliefs are instances of it.

Second, one might hold that possessing a self-belief does involve being disposed to experience episodes with a distinctive phenomenology. This is simply a consequence of accepting both Authority and the claim that the relevant states have a distinctive phenomenology. The point about phenomenology is only an objection to the argument I have offered if there is reason to think that a subject can believe that she is in a mental state without being disposed to experience the relevant phenomenology. I contend that any proposed example of this will be a case of someone who lacks a genuinely first-personal self-belief, for example, someone whose self-ascription is based on observation, inference, or memory.

Another worry: if Rational Significance is true then two beliefs with the content "I am angry" can be constituted by different dispositions. Given that an attitude is individuated by its content and the kind of attitude it is, it would seem to follow that first-personal self-knowledge is a distinct kind of attitude.[27] Self-knowledge is a different kind of attitude than garden-variety knowledge of one's mind. Freud was right: "Knowledge is not always the same as knowledge: there are different sorts of knowledge, which are far from equivalent psychologically" (Freud 347). Although this sounds controversial, it is not. For if we accept that attitudes are individuated by their roles in reasoning, then we should expect there to be different kinds of knowledge. For example, according to Motivational Judgment Internalism a subject who judges "I ought to φ"

---

[27] I thank Michael Caie for putting things this way to me.

will thereby be motivated, to some degree, to φ.[28] Put in the dispositionalist terms used here: moral knowledge, as a mental state, is partially constituted by a motivational state. Given that ordinary knowledge lacks this connection to motivation, it would follow that moral knowledge is a different kind of attitude, a unique species of knowledge. Similar points may hold of other kinds of knowledge as well.

I have presented a novel puzzle about self-knowledge, the Puzzle about Testimony, and offered a solution to it. First-personal self-knowledge is especially valuable to us because to possess it is to consciously occupy the perspective of the state known and so to be in a position to reason in a way that would be subjectively rational if one was in that state. To fail to possess this kind of knowledge, to have to rely on testimonial or inferential knowledge of oneself, is problematic because it leaves one alienated from the perspective of one's own mind. Furthermore, I have argued that if we accept Rational Significance then we must also accept Authority. First-personal self-knowledge is partially constituted by the state represented in its content.

Earlier I mentioned that most philosophers working on self-knowledge focus on the question of how we can rationally arrive at self-beliefs with distinctive epistemic features. This strategy biases one in favor of a view on which self-knowledge is ordinary knowledge of oneself possessed by special means. By pursuing a different strategy, by attending to the distinctive features of self-knowledge as a state of mind, we have arrived at a different conception. On this view, rather than thinking of self-knowledge as a way of discovering a mental state, we should think of it as a *way of being in that state*, being in it self-consciously.

---

[28] This view has been understood in different ways, for example, restricting it to the virtuous or rational subject. I'm just stating a crude version of the view for the sake of demonstration.

2 SELF-KNOWLEDGE AS AN ABILITY

2.1 KNOWLEDGE AND SELF-KNOWLEDGE

As we have seen, Constitutivists explain first-personal self-knowledge by appeal to constitutive relations between first-order states and self-knowledge. Since these relations hold in both directions, the view can be broken down into two core theses.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.
>
> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.

The restriction to a subset of second-order beliefs leaves room for self-ascriptions that do not fall under Authority, such as those based on self-observation, inference, or testimony. Self-Intimation makes a similar restriction and so avoids the obviously false claim that all beliefs are self-intimating. While it is controversial which beliefs are self-intimating, I won't take that question up here since, again, our focus will be on Authority.

Philosophers sympathetic with the Constitutivist view have interpreted these claims along two lines, which Dorit Bar-On has labeled "grammatical" and "ontological" versions of

Constitutivism.[29] According to the grammatical view, championed by Crispin Wright (1989, 1991, and 1998) and Richard Rorty (1979), Authority and Self-Intimation are brute facts about our psychological concepts and social practices that admit of no explanation. It is simply a constitutive feature of our mental concepts that under ordinary circumstances their first-person use is incorrigible and others must defer to a subject's self-ascription. As Wright puts it, this deference "enters primitively into the conditions of identification" of a subject's mental states (Wright 1991, 142).

By contrast, ontological versions of Constitutivism hold that these claims can be explained by appeal to metaphysical relations between classes of first and second order states.[30] In a discussion of Self-Intimation Shoemaker writes,

> to the extent that a subject is rational, and possessed of the relevant concepts (most importantly, the concept of belief), believing that p brings with it the cognitive dispositions that an explicit belief that one has that belief would bring, and so brings with it at least tacit belief that one has it.
>
> Shoemaker 1996, 241

According to the ontological view, then, Authority and Self-Intimation are truths about the dispositions that constitute a class of first and second-order states.

This paper offers a novel argument for an ontological version of Authority, following the strategy suggested by Shoemaker.[31] I will argue that there is a form of second-order belief constituted in part by dispositions possession of which is sufficient for possession of the first-order belief represented in its content. Rather than thinking of first-person self-knowledge as a

[29] Bar-On 2009. In a similar spirit, Zimmerman 2006 distinguishes between "anti-realist" and "realist" versions of Constitutivism.

[30] Ontological versions of Constitutivism are defended by O'Shaughnessy 2000, Zimmerman 2006, Rödl 2007, and Boyle 2011.

[31] For the remainder of the paper, "Constitutivism," "Self-Intimation," and "Authority" will refer to the ontological versions.

form of epistemic access to one's mind, we should think of it as a way of accessing extra-mental facts. As Tyler Burge puts the idea: "the first and second-order perspectives are the same point of view" (Burge 1996, 110). As I will put it, some beliefs that one believes that p are *ways of believing that p.*

The argument involves an application of a view about knowledge recently advanced by John Hyman to the case of self-knowledge of belief. According to Hyman, knowledge is the ability to have one's thoughts and behavior guided by the facts, so that when one knows that p, the fact that p can serve as one's reason for action. It would follow that self-knowledge of belief is the ability to be guided by the fact that one believes that p and to have this fact serve as one's reason for action. However, there is an ambiguity in these expressions. Sorting it out will reveal that there are two kinds of self-knowledge of belief, one of which is as Authority claims. I'll suggest that we should think of first-personal self-knowledge as a distinct kind of attitude, different from the ordinary knowledge of belief possessed by observation, inference from behavior, or testimony. More specifically, I will argue that first-personal self-knowledge should be understood as a distinct kind of ability: the ability to act for reasons self-consciously or in awareness of one's reasons *as reasons*. If this is right, then first-personal self-knowledge is best understood as a way of being guided by extra-mental facts, and since to be so guided is sufficient for possessing the first-order belief, first-personal self-knowledge is constituted in part by first-order beliefs.

I proceed as follows. In the next section I set out Authority in more detail, distinguishing it from other views in the literature. In Part Two I outline Hyman's account of knowledge. Then, in Part Three I explain what role beliefs play in action explanation if we accept Hyman's view. In Part Four I argue that we should recognize two distinct kinds of self-knowledge of belief,

which cannot be captured by a direct application of Hyman's view. I then amend Hyman's view for the case at hand and explain how doing so supports Authority. I conclude by considering objections.

2.2 AUTHORITY

The argument of this paper relies on modest assumptions about the nature of belief and other propositional attitudes, which it might help to rehearse. I assume that beliefs are clusters of dispositions to behave and reason in a variety of patterned ways.[32] This can be understood in broadly functionalist terms, though it needn't. On this view, possessing some conjunction of the following dispositions is sufficient for believing that p: to act in ways intelligible given that p, to affirm that p in speech and thought, to use p as a premise in practical and theoretical reasoning, to believe some of the logical consequences of p, to consciously judge that p, to entertain p in inner speech, to revise beliefs inconsistent with p, and the like. Now this view, even in broad outline, allows us the following thought: if, in virtue of being in a mental state M1, a subject possesses dispositions x, y, and z and x, y, and z are constitutive of being in a mental state M2, then M1 is partially constituted by M2. For example, suppose that to desire to φ is simply to be disposed to φ. Given that an intention to φ also involves such a disposition, it would follow that intentions are partially constituted by desires, so that if one intends to φ one thereby also desires to φ. (I am not endorsing this view of desire, which is obviously oversimplified, just setting it out as an example.)

[32] See Baker 1995 and Schwitzgebel 2013 for defenses of this idea. The alternative view is that possessing an intentional attitude consists in possessing a representational state stored internally and atomistically. See Fodor 1987.

This suggests a strategy for establishing Authority for belief: show that a class of first-personal self-beliefs are constituted by dispositions sufficient for possession of the first-order beliefs represented in their contents. This is the strategy pursued in the following.

I will now briefly contrast Authority with some other views in the literature.

First, it is important to distinguish Authority from Self-Intimation. Authority is a claim about the metaphysics of standing self-beliefs while Self-Intimation is a thesis about first-order states. Authority claims that some of our second-order states are constituted in part by first-order states. Self-Intimation claims that some first-order states are partially constituted by self-knowledge. One can accept Authority while rejecting Self-Intimation.

Because it is a thesis about the metaphysics of *standing* second-order states, Authority must be distinguished from the view that there is a self-verifying method for *arriving at* second-order beliefs. Alex Byrne has recently defended such a view.[33] According to Byrne, self-ascriptions of belief and other states are arrived at by following an epistemic rule that codifies a transition from a judgment or other mental state to a self-ascription. In the case of belief that rule is the following:

BEL: p, so I believe that p.

As Byrne notes, such a rule is necessarily self-verifying and that means that any subject who follows it will possess a second-order belief guaranteed to be true (Byrne 2005). However, this claim is significantly weaker than Authority, since, as we might put it, Byrne's view forges a link between a first-order state and an event of coming to possess a second-order belief. Byrne makes no claim about the truth of a standing self-belief formed by following Bel after the rule has been

[33] See Byrne 2005, 2011, as well as Setiya 2011.

implemented. It is consistent with Byrne's view that a subject follows Bel at t1, forming a self-belief that is guaranteed to be true, but which is false at t2, because the subject retains her second-order belief while losing the first-order state. We should expect this since Byrne conceives of the relation between a self-ascription of belief that p and the judgment or beliefthat p as an instance of the basing relation. Given the common assumptions that the basing relation is a causal relation and that causal relata are independent existences, Byrne's thesis predicts that it is possible to retain a second-order belief initially formed by following Bel while losing the first-order state. Thus, Byrne's view entails that while a self-belief formed by following Bel is guaranteed to be true when formed, no such guarantee holds at any later point.

By contrast, Authority is not restricted to events of coming to believe. It holds that at least some of our second-order beliefs are guaranteed to be true because of the dispositions that constitute them. Thus, there is a crucial difference between the way that Authority and self-verifying views like Byrne's conceive of the relation between first and second-order states. Authority explains self-knowledge in terms of its *constitution* as a mental state rather than its *etiology*.

One final point of clarification is in order. As I mentioned earlier, Authority is restricted to first-personal self-knowledge. However, this phrase may be taken to refer to any self-knowledge that is formed by a uniquely first-personal capacity: a way of knowing my mind available only to me. It is possible that there is more than one such capacity, yet Authority needn't apply to all. Thus, perhaps I can know that I believe that p on the basis of phenomenal qualities associated with that belief.[34] That would be a uniquely first-personal method for which Constitutivism does not hold. The Constitutivist can accept this possibility while insisting that

---

[34] See Pitt 2004.

some of our self-beliefs are as Authority claims: true in virtue of connections between first and second-order states. Thus, Authority is a claim about a certain class of self-beliefs. It is one of the principal tasks of this paper to isolate this class and to reveal its significance. In order to do that I will first set out John Hyman's account of knowledge.

2.3 HYMAN ON KNOWLEDGE

In a series of papers John Hyman has developed a conception of knowledge based on the idea, found in Wittgenstein and Ryle, that knowledge is a kind of ability.[35] The alternative view, dominant in the literature, is that knowledge is a species of belief, justified true belief plus whatever avoids Gettier cases. According to Hyman, one motivation behind the abilities approach is the failure of post-Gettier epistemology to discover the further necessary condition on knowledge.[36] For the abilities approach, like the knowledge-first approach, proceeds by identifying knowledge with a particular kind of mental achievement and then explaining belief as a falling away from this achievement, rather than proceeding in the opposite direction. Thus, Hyman's thesis involves two claims: that knowledge is an ability and that mere belief is an imperfect instance of knowledge. I will assume both here.

But what kind of ability is knowledge? It can't be the ability to perform specific types of actions, like brushing one's teeth or going to the park, since one can perform these on the basis of mere justified beliefs. Instead, Hyman suggests, knowledge is the ability to perform actions a certain way, knowledgably. Hyman compares this to other "adverbial tendencies" such as

[35] Hyman 2000, 2001, and 2006, Wittgenstein 1953, §150 and Ryle 1949, Ch. 2.
[36] Hyman 2000, 435-436.

reluctance and enthusiasm, which are not tendencies to perform specific actions but rather tendencies to perform actions a specific way, enthusiastically or reluctantly (Hyman 2001, 178-179). According to Hyman, to do something knowledgeably or with the knowledge that p is to act on the basis of the fact that p. Thus, knowing that p is the ability to do things with the fact that p as one's reason for action (in a suitably broad sense of "do" that includes mental acts like wanting, believing, and worrying.) Furthermore, if the fact that p is among one's reasons for action then one knows that p. Thus, "a person knows that p if and only if… the fact that p can be among her reasons for performing a certain kind of action or for refraining from performing a certain kind of action…" (Hyman 2006, 894). Hyman argues for the strong claim that this biconditional captures the nature of knowledge. However, for our purposes we need only accept the weaker claim that the biconditional captures a truth about knowledge.[37]

Hyman's thesis has considerable appeal. In acting intentionally we aim to achieve our goals in a way that is suitably sensitive to our environment. Using Hyman's metaphor, we hope to be guided by the facts in the pursuit of our goals. The idea of being guided is to be contrasted with brute determination; it is meant to capture the idea that intentional action requires an intelligent sensitivity to one's environment. By "intelligent sensitivity" I mean simply that a fact which guides one is a reason for action, and reasons for action must be represented by the agent herself. A reason for action is the reason *for which* the agent does something, as opposed to the broader class of reasons why she does it (Neta 2009). Epilepsy can be a reason why though not typically a reason for which one acts. Intentional action requires a certain kind of ability, an intelligent sensitivity to the facts. Hyman identifies this ability with knowledge.

---

[37] See Williamson 2000 for criticisms of Hyman's biconditional as articulating the nature of knowledge. Unger 1979, ch. 3 defends one direction of the biconditional, arguing that if one knows that p then there is some action that one can perform for the reason that p.

An advantage of Hyman's account is that it allows us to say that animals without the concepts or capacities required for self-knowledge can possess knowledge. A woodpecker bores into a tree because, by tapping it, it can hear that it is hollow, which indicates the presence of larvae and insects. The woodpecker bores into the tree because it is hollow; the fact that the tree is hollow is the woodpecker's reason for boring into it. Thus, the woodpecker knows that the tree is hollow. While the woodpecker can plausibly represent that the tree is hollow (or something like that), it cannot represent itself acting for the reason that the tree is hollow. But the absence of self-knowledge is no impediment to knowledge; these are different cognitive skills. (This will matter later.)

According to Hyman, the crucial difference between knowing that p and merely justifiably believing that p is that in the latter case the fact that p cannot serve as one's reason for action. In such a case one can only act for the reason that one believes that p. Early in the season no one can know whether the Yankees will make the playoffs. Suppose that, susceptible to cheap marketing gimmicks, I typically buy season tickets in order to get an early opportunity to buy playoff tickets. This year I refrain from buying season tickets because I have good reason to think the team's pitching will not hold up. It would be wrong to say here that I refrained from buying the tickets because the Yankees will fail to make the playoffs, even if they will. My behavior is not guided by the *facts* but only by my beliefs.[38] On the other hand, if I make the trip to the ballpark in order to see today's game and I know that there is a game on today, then my behavior is guided not only by my beliefs about the game, but by the facts themselves. The fact that there is a game today is my reason for going.

---

[38] Strictly speaking it is not guided *only* by my beliefs since it can also be guided by other facts known. For example, I might know that some of their pitchers have a history of injury. The point is that relative to their making the playoffs my action can only be guided by a belief and not the fact itself.

2.4 REASONS AND BELIEFS

Hyman's view relies upon the claim that reasons are facts, most often extra-mental facts, such as the fact that a game is on today.[39] This is controversial. Two rival views are that reasons are beliefs and that reasons are propositions, which may be true or false.[40] In this section I'll sketch an argument for the view that reasons are facts. My goal is not to give a full defense of this view, but rather to give a sense of the terrain and motivate Hyman's view enough for the purpose of considering the consequences of it for our understanding of self-knowledge. We must also consider the role of belief in action explanation, for even if reasons are facts, citing an agent's belief can explain her action. This will lead to a puzzle about how to understand self-knowledge of belief that will occupy us for the rest of the chapter.

We should accept that reasons are facts because reasons play two roles and only facts can play them both. First, reasons must be able to serve as premises in practical and theoretical reasoning. This means both that a reason is what I consider in reasoning toward action or belief and that different people can share it.[41] When I deliberate about whether to go to the movies I consider whether doing so will be entertaining, not whether I believe that doing so will be entertaining. Likewise, if you and I both go to the movies, that I believe that it will be entertaining is not the reason we share, outside of rare circumstances. Therefore, that beliefs usually do not serve as premises gives us good reason to think they cannot be reasons. However, both facts and propositions can function as premises, so this does not yet show that reasons are facts.

---

[39] I am assuming that facts are true propositions, but the argument I offer can be accepted even if one prefers a different conception of facts.
[40] Davidson 1980 holds that reasons are belief-desire pairs. Dancy 2000 holds that they are propositions.
[41] Hyman 2001 and Setiya 2014.

However, propositions, true or false, cannot play a second role that reasons play. In addition to serving as premises in reasoning, reasons explain action (Davidson 1980.) But explanation is factive: if p explains q then it must be the case that p. So if my reason for acting is that p, then it must be the case that p. False propositions cannot explain anything. Therefore, false propositions cannot be reasons. If reasons for action must be facts, propositions, or beliefs, and each of the latter two cannot play one of the roles essential to being a reason, then we must conclude that reasons for action are facts.

Now one worry here is that we often do act with false beliefs. Suppose that I need eggs and take as a premise in my reasoning that there are eggs at the corner store. When I walk to the corner store I am, in some sense, acting on the basis of the proposition that there are eggs at the store. But if there aren't, then I am acting on the basis of a false belief. If reasons explain action, it would seem to follow that my reason must be my belief, since it is the only thing available to explain what I am up to. Furthermore, even in the good case, when there is a fact to serve as a reason, it is still true that I believe that there are eggs at the store. Since the belief is the common factor in both explanations it is tempting to think that it is the reason.

I have argued that, since reasons both explain action and serve as premises, reasons must be facts. This view has trouble making sense of the bad case, where there is no fact to explain one's action, and yet, one's behavior can be given a rational explanation. Some philosophers think that this case should convince us that it is wrong to hold that reasons both explain action and function as premises. For example, Wayne Davis argues that beliefs and extra-mental facts play different roles in action: the facts can serve as premises but do not explain action, while beliefs typically explain an action without serving as premises (2005). However, it is very attractive to think that the very same consideration that serves as one's premise in reasoning can

also explain one's action.[42] When asked why one acted as one did, one will cite the premises of one's practical reasoning. But this would be pointless if what one appealed to didn't also, in the good case, explain one's action. Since it is not our primary concern here, I will assume the argument is a good one and that reasons are facts. Nevertheless, the above considerations show that beliefs play an important role in the explanation of action. As Hyman puts it: "it is an axiom that beliefs can explain actions" (Hyman 2001, 182). So what is the role of belief in action explanation if it is not that of a reason?

One view, recently defended by Eric Marcus, is that reasons explanations cite extra-mental facts but implicitly rely on an appeal to a capacity to be moved by one's recognition of those facts (Marcus 2012, 107-114). The role of belief in action explanation is to characterize the capacity by means of which the worldly fact is able to cause intentional action. On this view, the reason-giving force of citing a belief is derivative of the fact believed. In the good case citing my belief that there are eggs at the corner store is merely a way of noting that I know that there are eggs there and that fact is the reason why I walk to the store. That is, in this case, to say "he did A because he believes that p" is an indirect way of saying "he did A because of p," making explicit that the kind of reason in question is the reason for which he acted, his reason. In the bad case, too, citing the belief functions to cite the capacity exercised in one's action, though here the exercise of that capacity was defective. As Marcus puts it, an explanation that cites a false belief "invokes the ability underlying acting-for-a-reason in such a way as to make it possible to rationally explain the action even when the agent's reasons are unsound" (Marcus ibid. 106).

To say that knowledge is the ability to be guided by the facts, then, is to say that it is the successful exercise of an ability to act on the basis of one's representations of the facts. As I

[42] See Marcus 2012, ch. 2 for further defense of this idea.

mentioned earlier, on Hyman's view we should think of belief as an imperfect case of knowledge. We can now explain what that means. Mere belief and knowledge are exercises of the same capacity: the capacity to be guided by the facts, or to act on the basis of one's representation of the facts. Knowledge is the successful exercise of that capacity; mere belief is its unsuccessful exercise. Now even if one has a mere belief one will be disposed to do the same things one would do if one possessed knowledge, for example to perform certain actions or make various transitions in thought. In cases of mere belief, though, one's actions are not guided by the facts.

This view of the relation between reasons, belief, and intentional action requires more defense than I can give it here. But our primary goal is to consider its consequences for our understanding of self-knowledge of belief, to which we now turn.

2.5 ACTING FOR THE REASON THAT I BELIEVE THAT P

In the *Philosophical Investigations* Wittgenstein claims both that knowledge is akin to an ability (PI §150) and that "it cannot be said of me at all (except perhaps as a joke) that I *know* I am in pain" (PI §246). However, as Hyman notes, if the abilities view is correct then it is false that one cannot be said to know that one is in pain (Hyman 2001, 188). That I am in pain is a fact that can serve as a reason for action: it can be the reason why I reach for my knee, reach for the medicine cabinet, or reach for the telephone. Hyman's view about knowledge applies straightforwardly to the case of phenomenal self-knowledge since, whatever else is special about it, we respond to

our pains and sensations as facts like any other. That I am in pain can serve as a reason for action in just the same way that the fact that I have high blood pressure can.

However, things become more complicated when we consider the case of belief. On Hyman's analysis, self-knowledge of belief is the ability to act intentionally because of the fact that one believes that p. As we have seen, whenever one acts intentionally because of the fact that p, one believes that p. And that belief is explanatory of one's action. Agents with the concepts of belief and the first person are able to offer these explanations themselves. Thus, one of the roles of self-knowledge of belief that p is to explain one's acting for the reason that p. In this section I'll argue that a direct application of Hyman's view has trouble explaining these cases and that we can only account for them by recognizing that self-knowledge in action explanation is an ability distinct from the one at play in garden-variety knowledge, including knowledge of belief by inference, observation, or testimony.

Consider the following.

Case 1. I avoid a friend because she is out to get me. When I reflect on my action, I take it that I am acting for a genuine reason: I take myself to know that my friend is out to get me. Here I have reason to do a variety of things, such as to call the police, to avoid the friend, and so on. Now if someone asks me why I am avoiding my friend I will most likely say, "I am avoiding her because she is out to get me." But I can also say (perhaps misleadingly, but truthfully) "I am avoiding her because I believe that she is out to get me," indicating that my action is intentional and undertaken for what I take to be a reason.

Case 2. I believe that my friend is out to get me. Reading in a medical journal, I learn that such a belief is likely a delusion of persecution. I might also come to the unfortunate realization that the work of deliberation cannot dislodge this belief and that instead, at least for the time

being, this is a fact about myself with which I will have to find a way to live (Wollheim 2003). Knowing that I have this belief, I have a reason to act in certain ways. For example, I have reason to call a psychiatrist. In such a case the reason for which I call the psychiatrist is that I believe that my friend is out to get me.

Both cases involve self-knowledge of belief. But the self-knowledge plays a different role in the explanation of the intentional actions that it enables one to perform. In the latter case self-knowledge is much like ordinary empirical knowledge, an ability that allows one to act on the basis of the fact known. In cases like this one relates to one's beliefs as worldly facts that can serve as premises in reasoning. It just turns out that the worldly fact is a fact about one's mind. But in case 1 this is not so. There self-knowledge does not prepare one to act on the basis of a fact about oneself, or anyways not primarily. There one does not relate to the fact that one believes that p as a potential premise in reasoning. Rather, this kind of self-knowledge enables one to act on the basis of the fact known. But if, per Hyman's thesis, we individuate instances of knowledge that p in terms of the reason for which that knowledge allows one to act intentionally, then it would seem to follow that the self-knowledge in case 1 is simply an instance of knowledge that p, knowledge of what is believed. For in that case my self-knowledge enables me to act for the very same reason that my first-order knowledge would. This is a problem. For it is clear that knowing that p and knowing that I believe that p are not the same state of mind, since there are cases of knowing that p that are not self-conscious, for example, the woodpecker's knowledge that the tree is hollow. There must be more to self-knowledge in case 1 than mere first-order knowledge.

Hyman's view seems to have the unfortunate consequence of failing to explain the difference between self-knowledge of belief that p and knowledge that p in Case 1. To avoid it, I

think we need to modify Hyman's view for the case at hand. And I want to suggest that Hyman's own strategy can help here. The suggestion is that in cases like 1 above "self-knowledge" should itself be understood as an adverbial modifier. To possess self-knowledge is not merely to be able to do something knowledgeably, in the light of or by being guided by a fact about oneself. Rather, in these cases one acts *self-consciously*, that is, in the recognition of some fact *as one's reason for action.* The difference between self-knowledge in cases 1 and 2 is that while the latter is simply first-order knowledge of one's state of mind, the former is a modification or way of knowing an extra-mental fact. To act self-consciously involves knowing the fact that is one's reason *and* knowing it as one's reason for action. Notice that this is different from the unfortunate consequence of applying Hyman's view. That consequence was that self-knowledge that p could be identified with first-order knowledge that p. What I am suggesting is that we think of a variety of self-knowledge of belief as a way of knowing or believing that p.

If knowledge is the ability to be guided by the facts, self-knowledge (in case 1) is the ability to be guided by the facts while being aware of being so guided or simply believing that one is.[43] As Hyman points out, to know that p one must be aware that p and must be able to act for the reason that p. As we saw, animals lacking the capacities for self-knowledge or even practical reasoning are capable of this. By contrast, self-conscious creatures possess a further ability: the ability to have one's behavior guided by reasons that one *takes* to be reasons. Both my dog and I can know that there is food behind the cupboard; in some sense we are both aware of the food's location and can act on this knowledge, say, by approaching the cupboard. But I can do something further: I can act on the basis of my representation of the fact known *as my reason* for so acting. Possession of the concepts of belief, reason, and the first person enable one

[43] Unless otherwise noted, "self-knowledge" will refer to the kind possessed in case 1 and cases like it.

to reason about what one has reason to do and to act on the basis of one's assessment of those reasons. Acting self-consciously, then, involves acting for the reason that p, believing that p is a reason, and acting *because* one believes that p is a reason for action. That is, in self-conscious action one's beliefs about one's reasons are explanatory of one's actions. If I go to the store because there are eggs there, and this is self-conscious action, then my going to the store is explained not only by my knowing that there are eggs there, but also by my representing the fact that there are eggs as a reason to go. Since, plausibly, woodpeckers and other animals lack the concepts of a reason, belief, and the first person, they cannot act self-consciously.[44]

In addition to offering an attractive explanation of the distinction between two kinds of self-knowledge, this view is plausible in light of the role of belief in action explanation discussed earlier. As we saw, the reason-giving force of a belief is derivative of the fact believed. Self-knowledge in these cases could then be characterized as knowledge of a belief *as playing* this role in action explanation.

If self-knowledge is the kind of ability I have described then we should accept Authority.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

Start with the thought that the kind of self-knowledge possessed in case 1 is first-personal self-knowledge, the kind we possess without having to rely on self-observation or inference from behavior. Cases like case 2 involve knowledge of one's mind from an outsider's perspective, which is typically the result of observation, inference from behavior, or testimony. If this is right,

[44] While some have argued that other animals such as rhesus monkeys and apes are capable of metacognition (Hampton 2001, Call 2010, and Tomasello 2014, among others) there is good reason to be skeptical of the interpretation of the experiments, involving "opting-out" tasks, that are thought to suggest this (see Carruthers and Ritchie 2012).

then we should think of first-personal self-knowledge as the ability to be guided by the facts self-consciously.

Now Authority claims that if one has a first-personal self-belief it just follows that one possesses the first-order belief represented in its content. The idea we have arrived at is a version of this thesis. Self-knowledge is the ability to be guided by the facts self-consciously. Knowledge is the ability to be guided by the facts. If one knows that one believes that p then one is able to act for the reason that p (or in the bad case, to act for the reason that one believes that p). By possessing self-knowledge one thereby possesses the ability constitutive of knowing that p (or merely believing it). It follows, then, that the possession of first-personal self-knowledge of belief entails the possession of the first-order belief known.

Notice that this does not depend upon the factivity of knowledge. To see why, start with the idea of a first-personal belief that one believes that p, ignoring whether it is true or false. If this belief is the ability as described, then a subject who possesses it will be disposed to act either because of the fact that p (in the good case) or as if it was a fact that p, that is, because she believes that p. In both cases it just follows that the subject has the first-order belief too, given that she will be disposed to exercise her ability to act for the reason that p, though perhaps defectively. The ability to act for reasons self-consciously is itself an ability to act for reasons; if you like, it is a modification of that ability, a way of acting for reasons. Thus, possession of a first-personal self-belief is sufficient for first-order belief, and so we must accept Authority.

2.6 OBJECTIONS

I have argued that self-knowledge in cases 1 and 2 differs in the type of attitude involved. First-personal self-knowledge is a distinct ability from ordinary knowledge. But one might reject the claim that the difference between cases 1 and 2 is a difference in attitude types, insisting instead that a difference in content is in play. One might think that cases like case 1 involve self-knowledge of knowledge whereas cases like case 2 involve knowledge of mere belief.[45]

However, this objection doesn't work, for there are cases like case 2 that involve self-knowledge of knowledge, and cases like case 1 where one only possesses knowledge of belief. Start with the first. Suppose that I know that I know that Ty Cobb has the highest career batting average in major league history. As it turns out, people who know this are at high risk of contracting the flu. Suppose I get a flu shot. The fact that I know that Ty Cobb has the highest career batting average is my reason for getting the shot. But this explanation of my action does not implicitly highlight that I am getting the flu shot because Ty Cobb has the highest career batting average. That is not a reason to get the shot. My reason for action is not *what* I know about Ty Cobb, but *that* I know it. This is analogous to the case of the delusion of persecution. An example of the second case: suppose that I go to the park thinking there is a game on when there isn't. If I do this self-consciously, then I will take myself to be acting because of the fact that there is a game, and so, I will know that I believe this. Although I lack first and second-order knowledge, I possess self-knowledge of belief. In this case I mistakenly take myself to know there's a game on. This case is analogous to case 1 since when I explain my action by saying "I am going to the park because there is a game on" I take my reason to be *what* I believe and not

[45] I thank Karl Schafer for raising this objection.

that I believe it. What is distinctive about case 1, and cases like it, is not the attitude known, but that in reasoning about one's mental state one also reasons with the content of that state. One can do this with both knowledge of knowledge and knowledge of belief, although in the latter case one will mistakenly think the content of one's belief is a fact.

Another way of holding that the difference between the cases consists in the content and not the attitude type is as follows. One might deny that "belief" means the same thing in both cases. The thought here is that in case 1 when I say "I am avoiding my friend because I believe that p", "I believe that p" is only a hedged assertion that p. It is only in case 2 that I offer a genuine description of my state of mind. Thus, one might think that there is no difference in the self-knowledge present in the two cases since case 1 isn't a genuine case of self-knowledge at all (Urmson 1952).

It is certainly true that a self-ascription of belief often functions as a hedged assertion of the content of that belief. However, it is unclear why a hedged assertion that p cannot also function as an unqualified description of one's state of mind. It is *prima facie* implausible to think that psychological verbs have two radically distinct senses, one which contributes information about one's mind and another that modifies a description of the world. As Dorit Bar-On puts it, this view violates the intuitive thought that there is "semantic continuity" between these different uses (Bar-On 2004, 232.) It seems more plausible to hold that self-ascriptions both describe one's state of mind and function as hedged assertions at the same time. They describe one's state of mind and conversationally imply facts about the world (Kauppinen 2010).

Finally, the argument on offer here is limited in a number of ways. First, it assumes the truth of Hyman's biconditional about knowledge, of which I have only offered a limited defense. Second, I have focused here exclusively on self-knowledge of belief. One might reasonably ask

whether the argument here could be extended to cover cases like desire, intention, or the emotions. Indeed, many philosophers hold that an account of self-knowledge ought to generalize to these cases (Bar-On 2004, Byrne 2005). Although it is unclear what the results would be, there is reason to think that there is a strategy to be pursued for these cases. The argument here focused on the role of knowledge and belief in reasoning. It is because of the derivative role played by beliefs in rationalizing action that first-personal self-knowledge of belief must be quite unlike mere knowledge of a fact about oneself. If the role of these other states in reasoning is similarly derivative, gaining its force from reasons, then perhaps the argument here applies mutatis mutandis to those states. However, determining this would require a separate investigation into the role of these states in practical reasoning.

3 EXPRESSION AND THE CONSTITUTION OF SELF-KNOWLEDGE

3.1 FIRST-PERSON AUTHORITY

Other things being equal, if you want to know what I believe, desire, or intend to do, I am the person you should ask. If you and I disagree about my intentions then, absent a special reason to think I am being disingenuous or am somehow self-deceived, you should defer to me. This "first-person authority" distinguishes self-knowledge of the attitudes and sensations from the subject's knowledge of her moods, character traits, and the state of her body. While we expect the subject to be better *informed* about these features of herself, it seems that she lacks the kind of authority about them she possesses over her attitudes and feelings. In the latter cases, other things being equal, what the subject says goes. An account of first-personal self-knowledge of the attitudes ought to be able to explain in virtue of what a subject possesses this kind of authority.

According to Neo-Expressivists, this authority can be explained by the fact that self-ascriptions in speech based on first-personal self-knowledge, so-called "avowals", express the states of mind ascribed. When I tell you that I am angry you must defer to me because by avowing my anger I put it on display for you. In this paper I argue that while the Neo-Expressivist view captures the social dimension of first-person authority – the deference an audience owes a speaker - it cannot account for the fact that the speaker is positioned to authoritatively issue an avowal without relying on evidence. However, reflection on the

connection between self-knowledge and self-expression can help here. The Neo-Expressivist claims that our capacity for self-knowledge is also a capacity for self-expression. I argue that we should interpret that claim as follows: first-personal self-belief is partially constituted by a disposition to express the state that is represented in its content. First-personally believing that I am angry involves being disposed to avow and thereby express anger. Since expression is a factive phenomenon, expressing that one is in a mental state entails being in that state. Therefore, first-personal self-belief is partially constituted by a disposition which guarantees the presence of the states that are represented in their contents. Perhaps surprisingly, the expressive character of avowals lends support to a view on which self-knowledge is explained by constitutive relations between mental states and our knowledge of them.

I proceed as follows. In the next section I distinguish two aspects of first-person authority and argue that Neo-Expressivism can only explain one. Then I outline the Constitutivist view I favor. Part four sketches an account of the expression of mental states, drawing on recent work by Mitchell Green, but departing from him in significant ways. After a few general remarks about factivity in part five I offer the argument for a core Constitutivist thesis in Part six. I conclude, in part seven, by considering an objection to Constitutivism recently raised by David Finkelstein, a Neo-Expressivist.

3.2 TWO ASPECTS OF FIRST-PERSON AUTHORITY

Typically, self-ascriptions of beliefs, desires, and other mental states issued from the first-person perspective - so-called “avowals” - possess a special kind of authority or privilege.[46] Traditionally, this authority has been understood in epistemic terms: my self-beliefs are better justified or more reliably produced than my beliefs about the minds of others or their beliefs about me. Thus, Alex Byrne writes that explaining first-personal privilege amounts to explaining why it is “easier to be right” about one’s own mind than the minds of others (Byrne 2005, 81). This view conceives of first-person authority on the model of expertise, self-expertise, if you like. After all, at a minimum, an expert is someone for whom it is easier to be right about a particular subject matter. So understood authority is compatible with a wide variety of views about self-knowledge. Even Gilbert Ryle, who denied that we have a uniquely first-personal source of self-knowledge, could explain epistemic privilege as the result of the greater familiarity the subject has with her own behavior (Ryle 1949). Once one denies that avowals’ security reaches to the peaks of Cartesian certainty, explaining first-person authority would not appear to be difficult.

Against the traditional view, a number of philosophers have argued that first-person authority must be understood in a more demanding way, one which rules out an explanation in terms of self-expertise. Here are three significant disanalogies between the authority of the self and the authority of an expert.[47] First, it is typically inappropriate to question or doubt avowals, yet it is common to doubt the reports of experts. Second, it is typically inappropriate to ask a

[46] See Wright 1998, McDowell 1998, Finkelstein 2003, Byrne 2005, and Bar-On 2006 among many others. I will use the terms “authority” and “privilege” interchangeably.
[47] See Hampshire 1979, Finkelstein 2003, Bar-On 2006, Boyle 2011.

subject how she knows that she is in a particular state of mind, yet this is common practice with experts. Third, an expert is someone who possesses the ability to recognize what counts as evidence for a particular claim. Yet, intuitively, self-ascriptions are not based on any evidence at all; they are "groundless," "baseless," or "silent."[48] If this is right, then first-person privilege is significantly unlike the kind of privilege enjoyed by an expert, and so would require a different kind of explanation.[49]

According to the Expressivist tradition, associated with the later Wittgenstein, first-person authority is explained by the fact that avowals express the states they ascribe. When I tell you "I'm angry!" I thereby express my anger, "venting" it or putting it on display for you. In addition to expressing my belief that I am angry I express the anger itself. Aside from its intuitive plausibility, an advantage of this view is that it seems well suited to explain the more demanding notion of authority. If in avowing a state of mind I somehow put it on display for you, it is inappropriate for you to question it. Doing so would be analogous to doubting someone who shouted "I am shouting loudly!" Since the speech act manifests the very state of affairs it represents, it is self-verifying and so immune from ordinary doubts. As we might put it, Expressivists replace the model of the expert report with the model of a display.

Now early defenders of this approach, most notably Wittgenstein on one reading, held that because they are expressive avowals do not count as genuine truth-evaluable assertions.[50]

---

[48] See Wright 1998, McDowell 1998, O'Shaughnessy 2000. Notice that "groundless" here means that self-ascriptions are not based on evidence; it does not mean that there is no explanation at all of how they count as knowledge. See Cassam 2009.

[49] Of course, someone who pushed this line needn't deny that there is an epistemic asymmetry between the first and third person perspectives. She need only hold that what is distinctive about first-person authority can't be cashed out in terms of degrees or kinds of epistemic support.

[50] Kenny 1973 and Jacobsen 1996 endorse this reading of Wittgenstein. Finkelstein 2003 compellingly argues that Wittgenstein's views in this area are more nuanced. I won't consider the exegetical question here, nor will I dwell on the problems facing the non-cognitivist view attributed to Wittgenstein, which I take to be well-established. See Bar-On 2004.

Thankfully, in recent years philosophers sympathetic to the expressivist explanation of first-person authority have divorced the view from non-cognitivism.[51] On this view, dubbed "Neo-Expressivism", an avowal gains its distinctive privilege by expressing the state of mind avowed, yet it is still an assertion, one which, in the right circumstances, counts as knowledge. Thus, the Neo-Expressivist avoids the difficulties facing her non-cognitivist predecessors while taking on board the explanatory advantages of the idea of expression.

However, once we grant that avowals manifest self-knowledge it is far from clear that appealing to expression can fully explain first-person authority. Plausibly, expression is capable of explaining the *social* dimension of self-knowledge. That is, if we think of first-person authority wholly in terms of the deference an audience owes a speaker then the idea of expression seems well-suited to the task, as the comparison with other self-verifying speech acts shows. However, there is another dimension to authority: in issuing an avowal, or simply entertaining a self-belief, a subject takes herself to possess a kind of authority that does not rely upon responding appropriately to evidence. It is part of the project of explaining first-person authority to account for the nonevidential character of self-knowledge.

The idea of an expressive avowal does not explain this. Awareness of oneself as expressing a state of mind by avowing it typically takes one of two forms: observational or nonobservational knowledge of intentional action. But observing oneself express a mental state could hardly explain one's authority in issuing it, since, presumably, one possesses the authority prior to or in issuing it. On the other hand, in order to know that one is intentionally expressing a state of mind one must already know that one is in that state. It would seem, then, that the idea of expression presupposes and cannot explain the perspective of the subject on her own mind that

---

[51] See Falvey 2000, Finkelstein 2003, and Bar-On 2004.

enables her to express herself. While the idea of expression can explain an avowal's authority, it cannot explain how the subject can take herself to possess this authority in issuing it.

In response, the Neo-Expressivist might deny that in order to intentionally express one's mind one must know of it, hence denying the second dimension of authority. Dorit Bar-On argues that it is possible to express a mental state by means of an intentional action without intentionally expressing that state (2010b, 56). Thus, throwing a rock through a window is intentional and can express anger, but in throwing the rock I may not have as my goal the expression of my anger, and so I needn't have any beliefs about my anger in order to throw the rock. There are such cases, of course, but they are irrelevant to the point at hand where the focus is on avowals, which are precisely cases in which one has the goal of expressing one's state of mind.[52] When I shout "I'm so angry!" my goal may be to inform you of my emotions or simply to vent my anger. But in order to intentionally act on either goal I must know that it is my goal, which means that I must know that I am in the relevant state of mind.[53]

The Neo-Expressivist may insist that placing avowals on a continuum with nonlinguistic expressive behavior like crying and grimacing can explain their nonevidential character. It is obvious that a child crying in pain doesn't have any *reasons* for doing so, if that means considerations that warrant her in crying. She needn't know that she is in pain in order to cry and thereby express pain. If she is young enough then she will only be aware of her pain and not of *herself* as being in pain. While lacking the first-person concept she is still capable of giving expression to her pain. Later, when she comes to possess language she can express her pain by

---

[52] A point made by Boyle 2010. See Bar-On 2010b for discussion.

[53] For this reason I find Bar-On's invocation of Rosalind Hursthouse's claim that expressive behavior is a counterexample to Davidson's view that a belief and desire pair must always be the primary reason for action unpersuasive for the point at hand (Bar-On 2010b 56, citing Hursthouse 1991.) Even if one grants Hursthouse's point, it is not a counterexample to the claim at issue, which is that when a subject φ's intentionally she does so in the knowledge of herself φ-ing for a reason, a thesis which Bar-On herself endorses at 2010b, 57. Once we grant this then we must accept that avowals are issued on the basis of self-knowledge of the state avowed.

avowing it. Yet, the Neo-Expressivist insists, these later achievements do not substantially alter her relation to her pain. The avowal is not issued on the basis of an assessment of the reasons for doing so: it is spontaneously wrung from the subject. As Wittgenstein puts it, "to use a word without a justification does not mean to use it without a right" (PI §289). Similarly, Bar-On writes that "[t]he distinctive perspective of an avowing person, I would say, is not an epistemic perspective at all" (2010, 56).

Now it may be the case that emphasizing the continuity between avowal behavior and other forms of expression is useful in revealing the nonevidential character of self-knowledge. And it is certainly true that avowals are typically issued spontaneously. Nevertheless, it doesn't follow that the avower lacks an epistemic perspective on herself in some minimal sense. As we have seen, this just follows from the fact that avowal behavior is intentional self-expression and so performed on the basis of self-knowledge. Again, that an avowal is not issued as the result of conscious reasoning does not mean that it is not issued knowledgeably. While the nonlinguistic child need not take herself to be entitled to cry, a language-wielding creature must take herself to be entitled to avow her pains and other mental states, given that these acts are intentional. Neo-Expressivists tend to accept this point but underplay its significance. Thus, Finkelstein writes that it simply doesn't matter whether we say that avowals manifest self-knowledge (2003, 151). Bar-On argues that her Neo-Expressivist view is compatible with non-deflationary accounts of self-knowledge (Bar-On 2004). However, we need something stronger than rendering the Neo-Expressivist account of first-person authority *compatible* with claims about self-knowledge. That is because that account presupposes that we possess a specific kind of self-knowledge: nonevidential and nonobservational knowledge that places one in a position to express a state by avowing it. In order to understand first-personal authority we must understand this source.

The remainder of this paper explores what the connection between self-knowledge and self-expression reveals about this source. Start with the idea of a first-personal self-belief. As I understand it, this is a belief with the content "I am in M" that draws on the same epistemic source as first-personal self-knowledge without the assumption of factivity. Put another way: a first-personal self-belief results from an exercise from the same capacity without the assumption of success. I will argue that we should interpret the Neo-Expressivist insight along the following lines: first-personal self-beliefs are marked by their connection to expression. As I will put it, our capacity for self-knowledge is also a capacity for self-expression. This is the best explanation of the fact that we cannot express a mental state on the basis of observational, inferential, or testimonial knowledge of our minds. If I tell you that you are angry, you are thereby in a position to ascribe that anger to yourself. But this self-ascription does not express your anger. First-personal self-beliefs are uniquely situated to play that role. Taking this insight of Expressivism seriously requires that we accept the following:

> *Expression:* a first-personal self-belief that "I am in M" is partially constituted by a disposition to express M by avowing it.

Expression is significant because the expressing relation is factive: a piece of behavior cannot express that one is in M unless one is actually in M. As I argue below, this applies equally to the disposition to express one's being in M. That means that we should also accept the following:

> *Factivity:* if one is disposed to express being in M, one is thereby in M.

From Factivity and Expression it follows that a first-personal self-belief that "I am in M" is partially constituted by a disposition possession of which guarantees the presence of M. And that, in turn, means that we should accept the following:

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

Authority claims that a first-personal self-belief is a metaphysically sufficient condition for possession of the first-order state represented in its content. It is a core commitment of Constitutivist accounts of self-knowledge on which self-knowledge is explained by constitutive relations between mental states and self-knowledge. Taking the expressive character of self-knowledge seriously lends support to a Constitutivist account of self-knowledge.

That the Neo-Expressivist insight supports Constitutivism might come as a surprise since Neo-Expressivism and Constitutivism are typically presented as rival theories. However, given the explanatory limits of the former its insights need to be combined with some other account of the source of self-knowledge in order to adequately explain first-person authority. Here I argue that Constitutivism is the proper account. This option has gone unnoticed by both camps. The expressive character of self-knowledge has gone underappreciated by Constitutivists, while Neo-Expressivists have raised objections to Constitutivism. In the conclusion of this paper I'll argue that one of these worries misfires and so suggest that Neo-Expressivists should happily endorse Constitutivism.

It is important to notice from the beginning that Expression and Authority are claims about first-personal self-beliefs and not self-knowledge. The point here is subtle. Self-ascriptions of first-personal self-knowledge alone express first-order mental states. Self-knowledge based on testimony, observation, or inference does not enable a subject to express the state known by avowing it. I take this to show that avowals are not expressive in virtue of manifesting knowledge. It is not the factivity of first-personal self-knowledge that explains Expression. Rather, it is the constitution of that self-knowledge as a mental state. I think we should understand this as a fact about our capacity for self-knowledge, the same capacity exercised in

possessing first-personal self-beliefs. First-personally believing that I am angry involves being disposed to express anger. Such a disposition guarantees that I am angry. This means that first-personal self-beliefs are guaranteed to be true and so, arguably, guaranteed to count as knowledge, as Authority claims. But it is important to get the order of explanation here right. Avowals are guaranteed to manifest self-knowledge because they are expressive of first-order states, not the other way around.

3.3 AUTHORITY AND CONSTITUTIVISM

As we have seen, self-knowledge possesses a special authority or privilege. Gilbert Ryle held that this privilege was merely the result of the greater familiarity one has with oneself and that self-knowledge is of the "sorts of things that I can find out about other people, and the methods of finding them out are much the same" (Ryle 1949, 149). Few today accept Ryle's claim.[54] Instead, it is widely held that each subject can know her mind in a way unavailable to anyone else. In Alex Byrne's terms, self-knowledge is arrived at in a way *peculiar* to the first-person perspective (Byrne 2005). A philosophical account of self-knowledge ought to explain both privilege and peculiarity. If we accept the demanding conception of privilege, as I've suggested we should, then explaining first-person privilege also requires explaining the nonevidential character of self-knowledge.

[54] But see Gopnik 1993 and Carruthers 2011.

The Constitutivist account is surprisingly simple: roughly, for a class of mental states a subject knows that she is in one of those states simply by being in it.[55] Since only the subject herself can know of a mental state by being in it, the Constitutivist explains the peculiarity of self-knowledge with ease. Since self-ascriptions are metaphysically grounded in the very facts they represent, they are not based on evidence. A subject's privilege consists in the fact that in issuing an avowal she speaks from the very state of mind she represents herself as possessing. Because having the relevant kind of self-belief guarantees its truth, these beliefs are infallible and so epistemically privileged as well.

The explanation offered by Constitutivism is notably different from others in the literature. As Matthew Boyle puts it, rather than postulating a special epistemic method or capacity, such as inner sense or a rule of inference, Constitutivists offer a *metaphysical* account that focuses on the nature of the mental states in question (Boyle 2010). The account is metaphysical because it explains self-knowledge by appeal to the constitution of a class of mental states rather than their etiology. More specifically, Constitutivists accept the following two theses.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.
>
> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.

The restriction to a subset of second-order beliefs leaves room for self-ascriptions that do not fall under Authority, such as those based on self-observation, inference, or testimony. Self-Intimation makes a similar restriction and so avoids the obviously false claim that all mental

---

[55] Versions of Constitutivism are defended by Heal 1994, Shoemaker 1996 and 2012, O'Shaughnessy 2000, Bilgrami 2006, Zimmerman 2006, Rödl 2007, Boyle 2011, and Coliva 2012.

states are self-intimating. While it is controversial which states are self-intimating, I won't take that question up yet since I'll be focusing on Authority.

3.4 EXPRESSION

Our goal is to understand the relation between self-expression and self-knowledge. But what is expression? After all, we speak of "expression" in a variety of contexts. A piece of music expresses sorrow. A sentence expresses a proposition. A grimace expresses pain. Although there is plausibly an intimate connection between artistic, semantic, and mentalistic expression, our concern lies solely with the last of these. In this section, drawing on recent work by Green (2007), I will sketch an account of expression. Some philosophers sympathetic to the Expressivist explanation of first-person authority have avoided offering an account in favor of relying on the intuitive idea that can be solicited from examples (Finkelstein 2002 and Bar-On 2004). However, there are two reasons for thinking we need more. First, I aim to use the account sketched here to support the claim that expression is a factive phenomenon. And I need a robust account of expression to motivate that claim. Second, it is not obvious that the intuitive idea generalizes to beliefs and intentions. That is, there is reason to be skeptical that avowals of beliefs and intentions put those states on display in the way that crying and blushing plausibly place one's sadness and embarrassment on display. That is a problem, since if the notion of expression doesn't cover these states then it cannot be appealed to in explaining first-person authority, for it is clear that we exercise authority over our beliefs and intentions. Thus, my goal is to articulate the notion of expression that applies to emotions, desires, pains, and the like, and

then argue that it applies to beliefs and intentions as well. I'll begin by sketching the account and then raise a worry about the latter cases before explaining how they are covered by it.

In addition to using it to combat the traditional, epistemological interpretation of first-person authority, philosophers from both the Continental and Analytic traditions have appealed to the idea of expression to undermine the starting point of the *epistemological* problem of other minds.[56] That problem begins with the thought that in an encounter with another we are confronted with behaviors on the basis of which we must infer her state of mind. Thus Sartre claims of the other person that it is "only the outer shell which I possess" (Sartre 1943, 511). The rationality of these inferences looks problematic since it is unclear how bodily movements could be taken to constitute evidence for the existence of such things as thoughts and feelings. However, our ordinary interactions suggest that inference is not our only source of epistemic access to other minds. We speak of seeing how angry someone is, of hearing the fear in his voice, or of witnessing how badly a child wants something. We describe children as learning how to "hide" their emotions, the implication being that their minds are previously open to view (Bar-On 2010a). Merleau-Ponty writes that "I do not see anger or a threatening attitude as a psychic fact hidden behind the gesture, I read anger in it. The gesture does not make me think of anger, it is anger itself" (1945, 184). John McDowell expresses a similar idea:

> … we should not jib at, or interpret away, the commonsense thought that... one can literally perceive, in another person's facial expression or his behaviour, that he is in pain, and not just infer that he is in pain from what one perceives.
>
> McDowell 1998b, 305

[56] This contrasts with the *conceptual* problem of other minds, which asks how we could so much as have the idea of mental items that can be known from both the first and third-person perspectives. See Cassam 2007 for a discussion of the difference between these problems.

The perceptual account of the epistemology of other minds emphasizes the connection between expression and perception.[57] Let us take that as our starting point.

It is worth emphasizing that the crucial claim behind the perceptual view is that one can gain *noninferential knowledge* of another's mind, knowledge achieved directly on the basis of observing behavior. Whether this knowledge is in all cases strictly speaking *perceptual* is less important. This is clear from the fact that the claim that we can literally perceive another's pain commits one to a view about the admissible contents of perception that should hang free from one's views about the epistemology of other minds.[58] If we can literally see another's pain then "pain" must be among the admissible contents of perception. While some have endorsed such a liberal view, holding that we can see such things as lemons, baseballs, and Toyota Corollas (Siegel 2010), others have rejected it. However, even if one endorses a restrictive view of the contents of perception one should still make room for non-perceptual, but noninferential knowledge. An example here is facial recognition. It seems wrong to say that my ability to recognize Peter by seeing his face confers inferential warrant. After all, I might not even know what color eyes or hair Peter has, yet, I can tell it is him just by looking at him. But if, strictly speaking, *Peter* is not an admissible content of perception, then one must grant that facial recognition is a non-perceptual source of noninferential knowledge. Thus, a philosopher of perception with a restricted view of the contents of perception can accept that one can noninferentially know that one is in a mental state by observing behavior.[59] Following others, I will continue to call the view "the perceptual view" and speak of observing a mental state by

---

[57] The perceptual view of knowledge of other minds is endorsed by Dretske 1969, Austin 1979, McDowell 1982, 1998b, and Cassam 2007.

[58] The question of whether we observe another's pain or that she is in pain will be addressed below.

[59] McDowell is an example of a philosopher who holds both a strict view of the contents of perception (2008) and a version of the perceptual view of other minds (1982). I thank him for emphasizing the importance of this option.

observing behavior. If one favors a restrictive view of the contents of perception then one should interpret these terms loosely. To remain neutral, I will on occasion use the term "witness" to refer to a source of nonobservational knowledge based on perception that may itself not be strictly speaking perceptual. Thus, even if we do not strictly speaking perceive another's mental states, the core of the perceptual view is the claim that we can directly witness them.

It is easy enough to see why one might resist the perceptual view. It is natural to think that the relation between a mental state and its expression is causal. Thus, my face loses color because I am nauseous; the look on my face expresses my nausea. This is perhaps analogous to other causal relations, such as the relation between how my face looks and the flu I have caught. It is certainly true that if you look at my face you can come to know that I have the flu; and we might even say that you can "see" this. But it seems clear that your knowledge depends upon inferring that I have the flu on the basis of the look of my face, even if you don't draw the inference consciously. You have to infer this because the flu is not on display for you.[60] And this holds quite generally of causal relations. If the only difference between a genuine expression of anger and the behavior of a talented actor is its *cause*, then it is hard to see how one can witness anger in behavior. Therefore, if the perceptual view is at all plausible then expressive behavior must stand in some other relation to mental states, a relation that allows for a more demanding sense in which one sees anger by seeing a facial expression.

In his book *Self-Expression*, Mitchell Green argues that we can perceive mental states in another's behavior because that behavior stands in a kind of *part-whole* relation to the state

---

[60] I assume that knowledge is inferential just in case one's entitlement possesses the structure of inference. Notice that this can be so even if one does not consciously reason with premises in order to arrive at knowledge. Some inferences are "easy" in that one can just "tell" the conclusion without any effort (Dogramaci 2012). Now in colloquial speech "see" is ambiguous between cases of easy inferential knowledge and genuine noninferential knowledge. From the fact that we say that one can "see" that another has the flu we cannot infer that the relevant bit of knowledge is noninferential.

(2007). This view requires accepting two claims: that emotions and other mental states are complexes with parts and that one can perceive a whole by perceiving one or more of its parts. The first claim is plausible enough on the assumption that these states are constituted by clusters of dispositions. The second claim will take some time to set out.

Drawing on work in evolutionary biology, Green argues that expressive behavior is a *signal* that *shows* a state of mind. A signal is "any act or structure which alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved" (Maynard Smith and Harper 2004, 3). Signals can be designed by natural or cultural evolution, allowing for idiosyncratic conventions, such as the way that Jones bites his lip when he's angry. A signal is an act that conveys the information it was, in one way or another, designed to convey. Notice that not all signaling is showing. Green claims that mere signaling yields evidence while showing yields propositional knowledge (Green 105). His point, I take it, is that if I show you that p then you can thereby come to know that p noninferentially without drawing on any background knowledge to serve as a premise.

However, the idea of a characteristic component is narrower than part-whole: my foot is a part of the Milky Way but I cannot see the latter by seeing the former (Green 86). Green, following Darwin and others, claims that expressive behaviors are "characteristic components" of the mental states they express (Green 86-88). Just as one can see a house by seeing one of its sides, so too can one see another's pain by hearing her cry "Ouch!" Whether X is a characteristic component of Y is "relative to an organism O and an ecological situation E" (ibid. 87).

Notice that accepting Green's view does not require denying that expressive behavior is caused by mental states. That is because a part can be caused by its whole. As Green points out, rain is both caused by and a part of a storm (Green ibid.). Likewise, David Lewis argued for

“piecemeal causation,” which occurs when parts of a cause are temporally posterior to the effects (Lewis 1986). An example is that the recession caused the Occupy Wall Street movement, even though the recession lasted longer than the movement; the movement was caused by a part of the recession. Likewise, a temporally extended whole can cause some of its parts. Thus, a bankruptcy early in the depression is both caused by and a part of an economic depression. The important point for Green’s view is that it is in virtue of the part-whole relation and not the causal relation that one can witness a state of mind by observing behavior.

Now one might worry that Green’s proposal faces the same problem facing the causal model (Stout 2010). I claimed that the reason that one cannot see or noninferentially know that I have the flu is that it is possible for me to exhibit symptoms of a flu without actually having a flu. But it is also possible to see a wall and roof that is not actually part of a house. It would seem to follow that I have to infer that there is a house on the basis of seeing one side.

However, there is a crucial difference between the cases. Because the wall is a part of the house, the house and the wall are not distinct existences, whereas the flu and my facial expression are. It is because the flu is a separate entity that one must infer its existence on the basis of my facial expression. The facial expression is a symptom of the flu. However, a wall of the house is not a symptom of there being a house nearby: it is a part of the house. Of course, that I see the facing side of a house does not mean that I can see the *other sides*, instead the claim is that I can thereby see *the house*. Because I cannot see the other sides I must infer that they are there. But I infer that there are other sides on the basis of the proposition that “there is a house”, which I know simply by seeing a part of it. Likewise, my anger is not a distinct existence from the expressive behavior since the behavior is a manifestation of a disposition constitutive of the

anger. While seeing the behavior does not involve witnessing the way anger *feels* it does involve witnessing *anger*.

A natural question to ask is: what is on display in expressive behavior? That is, when a suitably placed observer gains noninferential knowledge of a mental state by observing behavior, is she confronted with the state itself, the fact that the subject is in that state, or something else? Here Green distinguishes between showing-that and showing-α, showing a fact and showing an object (Green 2007, 48-49). An act shows a fact, p, or object, o, if it entails either that p or the presence of o. Intuitively, self-expression is a form of showing-α (Bar-On 2010a). Crying, shouting, and avowing do not merely enable one to learn of the mind of another, they put one in a position to witness her fear, anger, and the like. If, as I think, expression enables perception or observationally based noninferential knowledge, then it must be showing-α, for one cannot perceive that "x is F" unless one is presented with x and its Fness. I can show you that my house is red in any number of ways (e.g. by pointing to the paint stains on my jeans), but this showing cannot enable you to perceive, and possess noninferential knowledge, that the house is red unless I present you with the house itself.

But if expression is showing-α, what is the object shown? There appear to be two options: the object is either the state itself or an event of one's undergoing it. Now it is hard to see what it could mean to say that expressive behavior presents pain itself. We do not say that I see her pain. Instead, what we witness is her undergoing the pain, which suggests that the object shown is an event: one's undergoing an occurrent episode. Expressive behaviors enable us to directly witness occurrent mental states because they are parts of the process of undergoing or suffering them.

I think it is plausible to hold that this account applies equally to nonintentional and intentional behaviors. Getting red in the face and screaming both put one's anger on display for others. And it applies to specifically linguistic behavior as well (Bar-On 2010a). An avowal of pain is also an aspect of one's undergoing a particular state of mind. To paraphrase Wittgenstein, avowals replace cries and inherit their expressive power. This point is usually made by means of a just-so story.[61] Imagine a hungry prelinguistic child. Her shouts and reaching behavior manifest her desire. Later she masters language and can shout "I want food!" The introduction of a conventional, linguistic behavior does not seem to change the fact that both behaviors are expressive. Thus, the account of expression in terms of characteristic components applies to avowals.

If this is the correct way to understand the idea that expressive behaviors are characteristic components then we can readily see why it is difficult to accommodate the case of belief and intention within it. While it is undeniable that we can express these states in speech, it is far from clear that avowals of them are characteristic components. After all, beliefs (and arguably intentions) are not states we undergo for a fixed duration (Geach 1957). While one might believe that p for a period of time (before forgetting or changing one's mind), one cannot be in the middle of doing so. Of course, one might think that avowals express judgments, which are occurrent episodes, rather than beliefs. However, judgments are not processes we undergo either. One cannot be in the middle of or near completing a judgment. And that seems to mean that avowals cannot be characteristic components of beliefs.

[61] See Bar-On's discussion of "Jenny" at p. 215 of her 2010a.

It is clear that states like belief and intention are not manifest in behavior in the way that pain, anger, and the like are, by manifesting what one is currently suffering or undergoing. Nevertheless, there is a kind of event of which avowals of belief and intention are parts. Arguably, these states, unlike pain and perhaps emotions, involve a subject undertaking some kind of *commitment*, to the truth of p or to the performance of some action φ (Brandom 1994, Bilgrami 2006). That means, at least, that I am required to defend the belief if it comes under scrutiny, to believe some of its consequences, and to see to it that the belief is well-justified when new evidence arises. By believing or intending I come under certain obligations to think and do various other things, as the circumstances dictate. This is relevant to our purposes because it allows us to appreciate that avowals of belief and intention express commitments as well. In ordinary circumstances, when I tell you "I believe that p" I have thereby taken on an obligation to defend that belief and to limit my future assertions to what is consistent with p. If this is right then avowals of belief and intention are acts expressing commitments, which is to say that they are *performatives* (Coliva 2012). (Of course, it is *avowals* of belief and intention that are performatives, not believing and intending as such.) This places these avowals in a category with promises and renders the idea that avowals of these states are characteristic components unproblematic. A suitably placed observer can observe my believing that p in the same way that one can observe my promising to φ: by hearing me undertake the relevant commitment. When I tell you "I promise to drive you to the airport" you gain noninferential knowledge of a commitment I have undertaken. You need not infer that I made the commitment on the basis of what I say: the commitment is manifest in my speech behavior. My overt promise is not a report of an inner episode the existence of which is hidden from you, it is constitutive of promising. The same holds for avowals of belief and intention: they express mental states by

being performances of the commitments constitutive of those states. When I tell you I believe that p, you needn't infer that I do; my so believing is manifest in my committing to the relevant truth by avowing the state.

The idea that avowals are performatives, acts of committing to the embedded contents, has been objected to on the grounds that it commits one to the unattractive view that psychological terms have two different *meanings*, one which conveys information about the mind and another that functions to qualify an assertion or other speech act.[62] While we must reject that view, the performative account is not committed to it. I am suggesting that one way of undertaking a commitment to its being true that p is to avow belief that p. (I am not claiming that self-ascriptions of belief *always* play this role.) This is consistent with holding that the term "belief" always refers to a state of mind. When I avow that "I believe that p" I describe my state of mind. Since what I describe is a state of being committed to the truth that p, my avowal also expresses that commitment (at least in ordinary circumstances).

One might grant that avowals of belief enable noninferential knowledge but insist that this knowledge is testimonial and not broadly perceptual. The reason why you don't have to infer that I believe that p when I avow the belief is because *I have told you*, and testimony yields noninferential knowledge.[63]

While it is certainly true that avowing a mental state can be a form of testimony, more is involved.[64] Typically, testifying is not a characteristic component of the fact testified. Telling you that it is raining is not a characteristic component of the rain. However, avowals of belief are

---

[62] Geach 1965, 260 raises the objection. The view is endorsed by Urmson 1952.

[63] Thanks to Karl Schafer for raising this objection.

[64] Michael Thompson 2012, on behalf of Wittgenstein and Anscombe, claims that avowals are not forms of testimony. However, this is too strong. At best, as speech acts their primary function is not testimonial. But that is of course consistent with holding that they also function as testimony. Compare: when I shout "I am shouting loudly!" I offer testimony. But the primary way in which you gain knowledge that I am shouting is not through my testimony, but by hearing me shout. Thanks to Kieran Setiya for pointing this out.

characteristic components. They are aspects of the state of belief, here, performances of the commitment constitutive of it. It is in virtue of being characteristic components that avowals of beliefs are expressions.

The insight of Neo-Expressivism is that first-personal self-beliefs are marked by their connection with expression. That could only be true if the states about which we intuitively possess privileged self-knowledge are capable of being expressed in behavior in the relevant sense. In this section, drawing on Green's work but departing from him, I offered a sketch of an account of expression that does apply to all of the relevant cases. On that account expressive behavior stands in a kind of part-whole relation to mental states, the relation of being a characteristic component. Expressive behavior, including avowal behavior, is not merely an indication of its causes, but a form of displaying or showing mental states. There are two forms in which this display can occur. Expressions of pain, emotions, and perhaps desires are parts of the process of a subject undergoing an occurrent conscious episode. Avowals of belief and intention are performatives; they are ways of undertaking the commitment constitutive of those states. It is worth emphasizing that the performative account only applies to mental states that can be conceived of as commitments on the part of the subject. It is a difficult question whether desires and emotions should be so conceived. We can leave that question open, since it is clear that these states can be expressed in another way though, by behavior that is a part of an occurrent undergoing of the subject.

3.5 FACTIVITY

The account sketched in the last section is significant because it goes some way toward vindicating a broadly perceptual epistemology of other minds. Another reason it is significant is that it supports the thesis I have called Factivity.

Factivity: if one is disposed to express being in M, one is thereby in M.

I explain why in this section.

John Hyman defines factivity as follows: "Let us say that a sentence-forming operator O on one or more declarative sentences is factive if, and only if, the statement "O_{S1} ... $_{SN}$" cannot be true unless the statements "$_{S1}$" … "$_{SN}$" are true" (Hyman 2011, 358). The mark of a factive operator, then, is that sentences of the form "O_{S1} but it is not the case that $_{S1}$" are uninterpretable. For example, knowledge is factive and so one cannot intelligibly assert that "Jones knows that p, but it is not the case that p." Similar points apply to verbs like "see," "perceive," and "explain." As Hyman points out, whether a verb or operator is factive is not a brute fact about language (ibid). Rather it is a property of a fact, act, or mental state possessed in virtue of the relation between it and the facts. Knowledge is a relation between a subject and the facts, perception is a relation between a subject and an object, explanation a relation between true propositions, and so on.

Now this might suggest that expression is not factive since it is intelligible to say of the actor that he looks angry. However, it is important to distinguish facial signatures and behaviors that we associate with the expression of a mental state from actual expressions of that state (Stout 2010). The actor looks like he is angry, but his behavior is not an expression of anger since he is not angry. We cannot say that the actor expresses his anger by doing such and such since the actor is not angry; he has no anger to express. The actor can, at best, act as if he was expressing

anger. If expression is factive then that must be because of an underlying relation between expressions and mental states. The view developed in the last section explains this. Behavior cannot express a state of mind one is not in because expressive behavior is a characteristic component of being in that state. Thus, if the subject lacks the mental state his behavior cannot be expressive.

Knowledge is factive, therefore one cannot know that p unless p. Expression is factive, so one cannot express that one is angry unless one is angry. This highlights a relation between behavior and mental states. But it also entails the existence of a relation between states, specifically, between being angry and being disposed to express one's anger. The factivity of expression entails that one cannot be disposed to express anger unless one is angry. For if one could be disposed to express a mental state M without being in M then, given that actually expressing M entails being in M, being in M would have to be among the stimulus conditions of the disposition. This would lead to the absurd conclusion that everyone capable of possessing it at all is disposed to express a given state of mind all the time. That means that we should accept Factivity: anyone disposed to express a mental state must be in that state.

Even if one is inclined to deny that expression is factive or to reject the account of expression offered in the last section one should recognize that Neo-Expressivists have to accept Factivity. If expressing a mental state by avowing it did not entail being in that state then an avowal would not conclusively establish or show to an audience that one was in that state. That is, if avowals merely indicated mental states then they would not be self-verifying and so it would be appropriate for onlookers to question and doubt them. That, in turn, would mean that expression could not explain the social dimension of first-person authority. So anyone reluctant to accept the conclusions of the last two sections can read the argument of this paper

conditionally, as articulating what follows once one accepts the claim that expression can explain the social dimension of first-person authority.

3.6 SELF-KNOWLEDGE AND SELF-EXPRESSION

From the foregoing we can conclude that a subject cannot express that she is in a mental state unless she is in that state. And we have seen that the idea of expression applies to all of the standard propositional attitudes. We need to see how this supports the claim that first-personal self-beliefs themselves involve being disposed to express a state of mind, the thesis I have called Expression.

> *Expression:* a first-personal self-belief that “I am in M” is partially constituted by a disposition to express M by avowing it.

The only self-ascriptions that express the state ascribed are those based on our capacity for first-personal self-knowledge. Put another way, avowals, the self-ascriptions that express first-order states, are manifestations of first-personal self-knowledge. Self-ascriptions that are not avowals report first-order states, but they do not express them. This is revealed by reflection upon cases. If I tell you that you are angry, you can gain knowledge of your mind. And on the basis of this knowledge you can self-ascribe anger. But, intuitively, this self-ascription does not express your anger. Likewise, looking in the mirror I may see myself blush and realize that I am embarrassed. But a self-ascription of embarrassment here (e.g. “Looks like I’m embarrassed.”) is not an expression of embarrassment. Finally, suppose I know that I get angry every day at 3 o’clock. Looking at the clock and noticing that it is 3:00 I infer I am angry and report on this. It seems clear that I do not, but by accident, express anger. These self-ascriptions are merely

reports; they remain at a remove from the states reported. If this is right then avowals can only be manifestations of first-personal self-beliefs.

This shows that first-personal self-belief is *necessary* for a disposition to express a mental state by avowing it. A subject can only express that she is angry by asserting "I am angry" if she first-personally believes that she is angry. However, I think that the above also gives us reason to accept that first-person self-belief is *sufficient* for a disposition to express by avowing. If it wasn't, then the latter disposition would require a further element, call it E. A subject could only be disposed to express by avowing if she possessed both a first-personal self-belief and E. But if this was the case then it would be mysterious why E could not be combined with other forms of self-knowledge to enable the subject to express her mind by avowing it. I contend, then, that a disposition to express being in a mental state is a constitutive feature of first-personal self-belief. If it were not then it would be mysterious why we could not express first-order states on the basis of testimonial, observational, or inferential knowledge of one's mind. As I put it earlier, our capacity for first-personal self-knowledge is also a capacity for self-expression.

One might grant that there is a constitutive connection between first-personal self-knowledge and expression, but deny that this tells us anything of significance about self-knowledge. That is, one might hold that *all* that first-personal self-beliefs have in common is that self-ascriptions based on them express the states ascribed.

However, this is merely a reiteration of the Neo-Expressivist view that we have already found wanting. A subject who is in a position to express a mental state by avowing it is only in such a position because she possesses a distinctive kind of self-knowledge: nonevidential self-

knowledge. Thus, it is not the case that the only thing that first-personal self-beliefs have in common is their connection to expression.

I conclude that the lesson we should draw from the Neo-Expressivist insight is that a disposition to express by avowing is a constitutive feature of first-personal self-beliefs. In the previous section we saw that the possession of such a disposition entails the existence of the state ascribed. And that means that first-personal self-beliefs are partially constituted by a disposition possession of which guarantees possession of the state ascribed by the belief. That is, first-personal self-beliefs are guaranteed to be true. We should accept Authority.

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.

Earlier I argued that there are two aspects of first-person authority. First, when a subject avows a mental state her audience owes her a kind of deference not owed the reports she might give of another's mind or reports on her mind based on observation, testimony, or inference from behavior. The claim that avowals express first-order states explains this. Avowing "I'm so angry!" is a characteristic component of my undergoing anger. So by avowing anger I put it on display for you, which is why it is inappropriate for you to question the avowal, assuming you are suitably placed to recognize what is going on before you. Second, we need to explain how it is that a subject can take herself to possess authority without relying on evidence of the kind others rely on in attributing states to her. Authority explains this. If Authority is true then first-personally believing that I am angry guarantees that I am angry. That is why the subject needn't rely on evidence but speaks with authority.[65]

---

[65] Tumulty 2010 argues that a Neo-Expressivist who accepts the account of expression in terms of characteristic components should also accept Constitutivism, though her argument, which relies on Robert Brandom's (1994) idea of "making explicit", differs significantly from the argument offered here.

3.7 A NEO-EXPRESSIVIST OBJECTION TO CONSTITUTIVISM

To conclude, I'll review an objection recently raised by David Finkelstein against Constitutivism. Responding to it will suggest that those rightfully impressed by the Neo-Expressivist insight should endorse Authority. It will also suggest that accepting Authority puts pressure on one to accept Self-Intimation.

According to Authority if I believe that I am angry in the first-personal way it just follows that I am, in fact, angry. It is tempting to understand this thesis as stating that believing that I am angry *makes it the case* that I am. Understood this way, Authority will surely raise some eyebrows. Finkelstein brings this out by claiming that the view seems unable to appreciate the different ways in which we are responsible for our beliefs, desires, emotions, and pains (Finkelstein 2002, 45-50). Under ordinary circumstances we bear a certain kind of responsibility for our beliefs. While it is difficult to characterize this form of responsibility in any detail, it is uncontroversial that we are responsible for what we believe or judge in a way that we are not responsible for our pains or even for the propositional attitudes that assail us, like desires and emotions. However, if one can make it the case that one is angry by believing one is, then it would seem to follow that one can be responsible for one's anger in *exactly* the same way one is responsible for the belief that gave rise to it. Since this is absurd we should reject Authority.

However, this objection only works if the Constitutivist holds that first-personal self-beliefs are, in Crispin Wright's term, *extension-determining*, that is, that self-beliefs give rise to first-order states (Wright 1998). While some Constitutivists have held this, we should reject it. Finkelstein's objection arises from asking the following reasonable question: how do infallible self-beliefs arise? If they are true in virtue of their constitution then they are not based on any

ground at all. But then it looks like we possess them either by volition or compulsion, rendering them epistemically irresponsible. However, we can avoid this worry by endorsing Self-Intimation. If Self-Intimation is true then first-personal self-beliefs are metaphysically grounded in first-order states. One believes that one is angry because one is angry. (Of course, this is no answer to the question "how do you know?" since answers to that question are supposed to cite evidence.) That one's self-belief is a feature of the state itself means that the state is not determined or caused by the self-belief. Therefore, it is not the case that the self-belief is a matter of volition. Nor is it blind compulsion, for it is epistemically responsible to believe that one is angry because one is angry.

When a pre-linguistic child cries she expresses sorrow, putting that state on display for suitably placed observers to witness. But once the child comes to possess language she can do something further: she can place her mental states on display for others by intentionally expressing them with an avowal. I have argued that the self-knowledge that enables her to do this must be conceived as partially constituted by the very state she avows. That our capacity for first-personal self-knowledge is a capacity for self-expression means that Authority is true. If the argument just sketched in the preceding paragraph is on the right track and accepting Authority gives one reason to endorse Self-Intimation, then that would show that the state avowed by the conceptual sophisticate is of a fundamentally different kind than that expressed by her pre-linguistic counterpart. For if the mental states expressed by avowals are self-intimating then rather than thinking of self-knowledge as a special kind of epistemic achievement, as the expertise model claims, we should think of it is a distinct kind of first-order state, a way of being sad: being sad *self-consciously.*

4 INFERENCE AS A SELF-CONSCIOUS ACTIVITY

4.1 INTRODUCTION

We possess a variety of ways of knowing our beliefs. These can be divided into two categories: those that are available to people other than the subject, such as observation, inference from behavior, and testimony, and those that are only available from the first-person perspective. I assume that there is at least one capacity that falls into the latter category, a way for the subject to know what she believes that is unavailable to anyone else. Call this first-personal self-knowledge.

Many philosophers writing on first-personal self-knowledge think of it as the product of a capacity that operates independently of the first-order states it detects. They assume that self-knowledge is like our knowledge of the external world in that the states known and our knowledge of them are, in Hume's terms, independent existences. Our capacity for self-knowledge is a way of detecting or discovering these states, just as perception enables us to detect features of our environment that exist independently of our perceiving them. Sydney Shoemaker refers to this assumption as "the Independence Condition" (Shoemaker 1996). By contrast the Constitutivist rejects the metaphysical independence of first-order states and self-knowledge. She holds that self-knowledge is a metaphysically necessary condition on a class of

first-order states. Some of our mental states are such that part of what it is to be in one of them is to know that one is. More specifically, the Constitutivist holds the following two claims:

> *Authority:* there is a class of mental states M such that, if one first-personally believes that one is in M, then one is in M.
>
> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.

Rather than explaining self-knowledge in terms of its etiology, the Constitutivist explains self-knowledge in terms of the constitution of first and second-order mental states.

Many philosophers find these theses incredible. That is because if Authority is true then some of our self-beliefs are *infallible*: if I first-personally believe that I am angry or believe it's raining it just follows that I do. And if Self-Intimation is true then it would seem that we are *omniscient* about some of our mental states. We cannot be in them without knowing that we are. The idea that we are infallible and omniscient about any of our mental states is thought absurd in most quarters, an idea that Timothy Williamson calls one of "the quaint relics of Cartesian epistemology" (Williamson 2000, 193). Williamson has argued that no non-trivial condition is "luminous" in the way required by Self-Intimation. Other philosophers have argued that the great extent to which subjects are routinely mistaken about their own minds, as revealed by results in social psychology, is strong evidence that nothing like Self-Intimation is true (Cassam 2015). Against this trend, I think that Constitutivism is neither quaint, nor a relic, nor empirically falsified. Indeed, Cartesian though it is, it is true, when suitably qualified.

Two quick qualifications about the scope of these claims will help to make a start at rendering them more attractive. First, it should not be thought that Authority holds of just any self-ascription of a mental state, any belief with the content "I am in M." As I understand it, Authority is restricted to our first-personal self-knowledge. Thus, if I believe that I am angry on

the basis of your testimony it does not follow that I am, in fact, angry, likewise with self-beliefs based on self-observation, inference from behavior, or other "third-personal" sources. That leaves the defender of Authority the task of identifying what unifies the class of first-personal self-beliefs other than their truth. Second, Self-Intimation does not hold of all mental states. There are many kinds of mental states of which, by their nature, the subject is ignorant. These are not counterexamples to Self-Intimation. That leaves its defenders the task of demarcating its scope: which mental states are self-intimating and why?

The goal of this paper is to make a start at this second task by focusing on one case: inferential beliefs or beliefs held on the basis of other beliefs. I will argue that inferential beliefs are self-intimating, or as I will also put it, our capacity for inference is self-conscious. If I believe that p on the ground that q then, thereby, I know this.

The argument is as follows. A necessary condition on rational inference - the inferences of creatures capable of inquiry - is that the subject take her premises to support her conclusion and that she believes the conclusion because of this. Following Paul Boghossian, I will call this "the Taking Condition" on inference. The basic idea is that when I believe that p on the ground that q, I believe that p because I take q as my reason to believe p. My own appreciation of the significance of reasons explains my believing as I do. The Taking Condition is attractive for reasons I will set out. However, it has faced a number of objections. To maintain the condition in the face of them, we must accept that what makes it the case that I take it that q is my reason for believing p is that I believe (indeed, know) that I believe p for the reason that q. That is, the source of the attitude of taking is self-knowledge. A first-personal self-belief about one's reasons is sufficient for the taking attitude. My thinking that q is my reason for believing p makes it the case that I do believe it for that reason. The Taking Condition is a necessary condition on

inference. Therefore, drawing an inference must be sufficient for possessing self-knowledge of inferential belief. Thus, we should accept Self-Intimation for inferential belief because self-knowledge plays an explanatory role in the acquisition and possession of these beliefs.

So long as one accepts the Independence Condition this will sound paradoxical. How could my knowing that I believe that p be part of the explanation of my so believing? Don't I have to believe that p before knowing that I do? How could my belief that I so believe explain my actually believing it? (Boghossian 2014) How could such a belief be rational? (Finkelstein 2002). These are important questions. They express the worry that constitutive self-knowledge, if there is such a thing, does not deserve to be called "knowledge." I will respond to some of these worries later in the paper. But my goal here is to motivate the idea that there is constitutive self-knowledge. It is a further question to which conceptions of knowledge that idea conforms and which rule it out.

The remainder of the paper proceeds as follows. The next two sections are preliminary. First, I discuss Self-Intimation in more detail, distinguishing it from a close rival view and responding to some objections to it. Then I explain how I understand inference, including why we should think of it as agential, and motivate the Taking Condition. In the following sections I set out an account of inference as a self-conscious capacity in response to objections to the Taking Condition. I conclude by considering objections.

A caveat: in recent years a number of philosophers have sought to provide an interpretation of the Taking Condition that avoids some of the objections I will survey. These include appealing to a primitive notion of rule-following (Boghossian 2014), dispositional accounts of reasoning

(Broome 2014), or conceiving the attitude of taking as an intuition (Dogramaci 2012, Chudnoff 2014). Since my primary concern is to consider a reason for accepting Self-Intimation, I will not consider these alternative views here except in passing.

4.2 SELF-INTIMATION

Self-Intimation takes inspiration from Kant's claim that "the 'I think' must be able to accompany all my representations" (B 131). It interprets this as the strong claim that self-knowledge is a necessary condition on some representations. There are two points that need to be made about attaching this 'I think,' one psychological, the other epistemological. First, the Constitutivist does not hold that first-order states have self-reference built into their intentional contents, so that we should say that in being angry, desiring something, or believing something, I am at the same time and in the same way also thinking about myself. Such a claim is clearly psychologically unrealistic. Being in a first-order state is not sufficient for forming an explicit, conscious representation of oneself. Second, according to Constitutivism, forming an explicit self-representation, that is, consciously thinking about oneself or attaching the 'I think', does not involve gaining involve new knowledge. Attaching the 'I think' is merely a matter of bringing to consciousness knowledge one possesses in virtue of being in a first-order state. Since knowledge entails belief, Self-Intimation holds that being in a first-order state is sufficient for possessing a second-order belief, though this belief will likely be tacit or implicit, that is to say, its content is not thereby an object of conscious attention. A related point: while some philosophers hold that both the presence and absence of a mental state are self-intimating (Shoemaker 1996, ch 3), I aim to defend the weaker claim that the presence of some mental states is self-intimating.

Many philosophers think that self-knowledge must be explained by appeal to the very capacities that give rise to first-order states. Alex Byrne calls such accounts "economical" (Byrne 2005). This contrasts with accounts of self-knowledge that appeal to capacities like inner sense or mind-reading, capacities not involved in the formation of first-order states. Such views are "extravagant" (ibid.) Now those sympathetic with an economical view of self-knowledge accept that being in a first-order state is sufficient to provide one with justification for a second-order belief. But they disagree about the form this justification takes. According to what I'll call the Propositional view, being in a first-order state is sufficient to provide one with propositional justification for self-knowledge.[66] In order to take advantage of that justification and possess self-knowledge, something further is required, such as a trivial inference or attention to an episode in phenomenal consciousness.[67] By contrast, the Constitutivist, endorsing Self-Intimation, holds that being in a first-order state is sufficient to possess a doxastically justified second-order belief.[68]

Declan Smithies has argued that only the propositional view is plausible since one can be in a mental state but fail to recognize this due to conceptual poverty, inattention, or misclassification (ibid. 265). Creatures without the capacity for self-knowledge, like infants and

---

[66] I assume that a subject possesses propositional justification to believe that p just in case she possesses evidence or reasons that lend objective epistemic support of one form or another to the proposition that p. One can possess this kind of justification without appreciating its significance and without basing the belief that p upon it. Thus, if I enjoy a veridical perception that p, that provides me with propositional justification. Falsely believing that I am hallucinating, I cannot rationally take advantage of that justification. By contrast, to possess doxastic justification for the belief that p is to believe that p on the basis of the relevant evidence or reasons. For discussion see Leite 2004.

[67] Byrne 2005, 2011, Setiya 2011, Fernandez 2014, and Valaris 2014 all endorse the claim that self-knowledge is the result of a trivial inference. Smithies 2012 and Peacocke 2014 endorse the idea that it requires attention to phenomenally conscious episodes, perhaps a constitutive form of acquaintance.

[68] This is perhaps a bit misleading since typically one possesses doxastic justification by basing a belief on reasons or evidence. But constitutive self-knowledge is not epistemically based on anything at all, as I emphasize below. Nevertheless, since constitutive self-knowledge is knowledge, I take it that it is, in some sense, doxastically justified, though not on the basis of grounds or reasons. This relates to the worry mentioned earlier, that constitutive self-knowledge isn't really knowledge, a worry I am putting aside for the time being.
Smithies 2012 draws a similar distinction between epistemic and doxastic versions of Constitutivism. I prefer to reserve the label "Constitutivism" for views that endorse Self-Intimation.

dogs, can be in a mental state without believing it. And even creatures who have this capacity can fail to be aware of conscious episodes like pain. Smithies considers the case of "an athlete (who) incurs a painful injury while playing football but does not notice the pain until the game is over" (ibid.) Similarly, people who are primed to think they will be touched with a hot object but are touched by a cold one believe, if only for a moment, that they feel hot (ibid.). Given these limits on our introspective capacities, Smithies argues, we must reject Self-Intimation.

Let's start with the cases of inattention and misclassification. I think that Smithies' point is well taken in the case of self-knowledge of phenomenally conscious mental episodes like pain and feelings of hot and cold. Believing that one is in pain requires some kind of attention to the phenomenal character of one's pain. This attention is fallible in the same ways as attention generally. Therefore, if there is a constitutive connection between phenomenal episodes and self-knowledge it must be as the propositional view holds.

However, Smithies' counterexamples are not as compelling when considering our knowledge of the intentional attitudes. Someone sympathetic to Self-Intimation can avoid them by denying that self-knowledge of these attitudes depends upon attention. As I understand it, it is precisely the point of Constitutivism to deny that self-knowledge relies upon any discriminatory capacities at all. On this view self-knowledge is "groundless", "baseless," or "silent" (McDowell 1998, Wright 1998, and O'Shaughnessy 2000). It does not depend for its doxastic justification upon conscious attention directed at mental states, rather, it is metaphysically grounded in those states. That discriminatory capacities are fallible has no bearing on the truth of Self-Intimation (Berker 2004). To respond to the objection, then, the Constitutivist must draw a distinction

between knowledge of phenomenal states, which depends upon attention to items in phenomenal consciousness, and knowledge of intentional attitudes like belief, which does not.[69]

That leaves us with the worry about conceptual poverty. Must the Constitutivist deny that animals incapable of self-knowledge have beliefs? I don't think so. Instead, the Constitutivist needs to hold that the states that are self-intimating are different in kind from the beliefs of creatures without the capacity for self-knowledge.[70] This is, no doubt, controversial, but here is one reason to accept it. I will argue that we must accept that inferential beliefs are self-intimating in order to maintain the Taking Condition on inference. That leaves room for the thought that beliefs that are not subject to that condition are not self-intimating. And I think it is plausible to think that the reasoning of creatures without self-knowledge is like that.[71] When an infant responds to the sound of his mother's voice and forms a belief about her location, he treats the sound as evidence, if you like, but he needn't represent the sound as evidence or "take" it as a reason in any demanding sense.[72]

A related worry concerns how the defender of Self-Intimation can make sense of self-deception. Suppose that, when asked, I tell you that I believe that my wife is an excellent baker. Yet observing my behavior over time you notice that I always have a reason not to eat anything she bakes, I'm out the door when the oven goes on, and so on. In short, it is clear that I do not find her to be an excellent baker, indeed, I believe that she is a poor one. My false self-ascription is motivated by a concern for something other than the truth. Now if Self-Intimation is true, and I

---

[69] See Boyle 2009 for an argument for drawing this distinction. Notice that Smithies would not recognize this distinction as significant since he conceives of beliefs as individuated by their phenomenal character (Smithies 2012, following Pitt 2004). I find this view of belief highly implausible, but cannot take up the issue here.
[70] Versions of this idea are endorsed by McDowell 1994, Marcus 2012, ch 3, and Boyle, forthcoming. See also discussions in chapter 2 and the conclusion of chapter 3 below.
[71] Contra Boghossian, who takes it as a condition of adequacy on an interpretation of the Taking Condition that it apply to creatures without self-consciousness or concepts.
[72] Of course, that leaves us wondering about the unity of reasoning in creatures without self-knowledge. But that is beyond the scope of this paper. I am merely marking a commitment of the view I defend.

believe that my wife is a poor baker, then I know this, and so cannot, it would seem, simultaneously believe that she is an excellent one. It seems that for self-deception to work we precisely need to be ignorant of some our beliefs, at least temporarily. And that makes cases like the above seem puzzling. How is self-deception possible if we always know what we believe?

In response, I think we should restrict Self-Intimation to a subset of our beliefs, the conscious or, in Shoemaker's terms, "available" beliefs (Shoemaker 2012). As Shoemaker puts it, "a belief is 'available' if the subject is 'poised' to assent to its content if the question of whether it is true arises, to use it as a premise in her reasoning, and to be guided by it in her behavior" (ibid. 245).[73] The idea is that at a given time a subject is poised to assent to and use in reasoning a limited number of the many things she believes. This allows us the following interpretation of the example: the motivational forces that lead to my self-deceived self-ascription also repress or otherwise render inaccessible my belief that my wife is a poor baker. I believe that she is a poor baker, but, because this belief is not accessible, I do not know this. What is impossible is for my belief about her baking to be accessible while, at the same time, I have a motivated false self-ascription.[74] That would involve consciously holding contradictory beliefs at the same time, or, if this is different, being disposed to make Moore-paradoxical assertions. Thus, Self-Intimation is consistent with the existence of self-deception and false self-ascriptions so long as these involve some form of division or partitioning within the mind (Davidson 1986).

---

[73] It is important to emphasize that the idea of availability does not involve phenomenal consciousness. A belief can be available even if it is not phenomenally conscious (Shoemaker ibid.).

[74] Shoemaker makes the stronger claim that there cannot be false self-ascriptions of belief (ibid. 248-249) since such a self-ascription would not be constituted by the relevant first-order state, violating Authority. By contrast, I would say that false self-ascriptions are possible but that they must be based on third-personal sources like observation, inference from behavior, or something else. I am inclined to think this is exactly what happens in self-deception: a subject self-ascribes a belief because of a pleasing image of himself or some other motivational factor. Such a self-ascription is not based on the first-order state itself, and so does not fall within the scope of Authority. Remember here the claim made earlier that Authority and Self-Intimation are restricted to a subset of mental states.

While our topic is not self-deception, three things are worth noting about this proposal. First, in order to accept that self-deception involves division, we needn't follow Davidson in thinking of the division as a literal partitioning of two autonomous minds. We need only make room for motivational factors masking one belief and leading one to hold a contradictory or conflicting one. Second, while self-deception requires at least temporary ignorance of one's beliefs, it also seems to require some sort of familiarity with that belief. When someone overcomes self-deception it is natural to say that, in some sense, she knew all along what she had hidden from herself. Plausibly this familiarity is what distinguishes self-deception from wishful thinking and mere ignorance. This might be taken to suggest that rather than revealing the limits of self-knowledge, the explanation of self-deception may need to appeal to a demanding view of self-knowledge.

The third note concerns the relationship between this qualification of Self-Intimation and inference. I am going to argue that Self-Intimation holds of inferential beliefs: a rational subject with the concept of belief who believes that p on the ground that q thereby knows that she believes that p on the ground that q. We should accept this because Self-Intimation can explain the attitude of taking it that q supports p, which is necessary for rational inference. Now if Self-Intimation is restricted to available beliefs, then it cannot explain a necessary condition on inferential beliefs unless the restriction applies to them as well. If an inferential belief subject to the Taking Condition can be unavailable or unconscious, then Self-Intimation cannot explain the Taking Condition on inference. But this is not so. The Taking Condition holds only of instances of believing p on the ground that q where both beliefs are available in the sense outlined here. To see why, I turn now to a discussion of that condition.

4.3 THE TAKING CONDITION ON INFERENCE

By “inference” I mean the formation of beliefs by reasoning from other beliefs, the kind of reasoning that, when all goes well, yields knowledge. As I understand it, inference is personal-level, conscious, and employs contents involving concepts grasped by the subject. In this way it contrasts with “fast”, sub-personal, automatic information processing of the kind that underlies many psychological systems. The words “automatic” or “fast” are a bit misleading in this context since they suggest that all the reasoning that is properly called “inference” requires conscious attention, effort, and time. That is not so. Much personal level reasoning is effortless and does not require conscious attention. For example, seeing the puddles on the street I immediately conclude that it is raining. Though effortless and quick, this falls under the heading of personal-level reasoning.

In a recent paper, Paul Boghossian has argued that we should accept the following as a necessary condition for a mental act to count as theoretical inference:[75]

> *Taking Condition:* Inferring necessarily involves the thinker taking his premises to support his conclusion and drawing his conclusion because of that fact.
>
> Boghossian 2014, 5

Suppose that I believe that it is raining because there are puddles on the street. The Taking Condition holds that in order for this to count as genuine inference it must be the case that I am able to access the premise, recognize the support it lends the conclusion, and believe the conclusion because of this recognition.

[75] Plausibly, something like the Taking Condition holds of practical inference as well, but I will not consider that here.

I will assume that the “taking” in question is a doxastic attitude. One might ask, what is the content of this attitude? The Taking Condition holds of all inferences, deductive as well as non-deductive. To capture this generality, I will say that what one takes is simply that “q supports that p”, where this is intended to leave room for any particular form of epistemic support. Notice that the content here is not a conditional. Our topic is not the conditional, nor is it knowledge of logical truths, nor the basis of our understanding of the logical connectives. Furthermore, the attitude of taking must be distinguished from beliefs that “q supports that p” where commitment to belief in q or its truth are withheld, such as in hypothetical reasoning. Our focus is inference as a way of knowing.

One reason for accepting the Taking Condition is that it can explain the sense in which inference is an activity performed by the subject. I think there is a strong intuition in favor of this idea, but as we will see, the point is delicate. First, the intuition. Here is Boghossian:

> In the relevant sense, reasoning is something we do, not just something that happens to us. And it is something we do, not just something that is done by sub-personal bits of us. And it is something that we do with an aim—that of figuring out what follows or is supported by other things one believes.
>
> Boghossian 2014, 5

Inference, or reasoning, is something we do because it is an act that has a certain aim, the aim of gaining knowledge or truth, and these aims are, in some sense, the aim of the thinker as a whole rather than some part of her. Another way in to the intuition is by way of the idea that an inference is something for which the subject is responsible (Moran 2002, Hieronymi 2011, Boyle 2012). Thus, if I believe that p on the ground that q, I am responsible for believing the conclusion and responsible for treating q as a reason to believe p. But responsibility entails some form of agency. Therefore, our practices of holding people responsible for their beliefs requires making room for some notion of cognitive agency.

Now the delicate point. It helps here to contrast dynamic processes like counting and throwing a football from states like believing that it is raining or being eligible to vote. The difference between these categories is marked grammatically by the applicability of perfective and progressive aspect (Setiya 2013 and Comrie 1976). Processes are expressed by verbs that admit of both perfective and progressive aspect: we can say both that "Jones counted the people in the room" and "Jones was in the midst of counting the people" or "Jones threw the football" and "Jones was throwing the football." Verbs in the second category do not admit this distinction. While it is permissible to say that "Jones came to believe that it was raining" and "Jones became eligible to vote", we cannot say that Jones is in the midst of but has not yet completed these tasks. And this grammatical distinction reflects a metaphysical difference between processes and states.

Now the claim that inference is an activity might incline one to think of it as a *process*, as an event that unfolds over a period of time, like counting or throwing a football. However, we should resist this for two reasons. First, the grammar suggests that drawing an inference or arriving at a conclusion itself is not a process. One cannot be in the middle of drawing a conclusion, though one might be in the middle of considering the significance of evidence. Contemplating the evidential weight of accepted premises may be a process, but drawing a conclusion is not. We should not identify inference with the processes that precede it. As Galen Strawson puts it, these processes and actions are "entirely *prefatory*" and "essentially – merely – *catalytic*" (Strawson 2003, 231). Thus, while inference is typically preceded by various cognitive processes, it should not be identified with those processes.

Second, even if there is a dynamic process of inferring, the Taking Condition is not restricted to events of coming to believe. Suppose I believe that p on the ground that q. I have

believed it for some time on that ground. As I understand it the Taking Condition applies to this static belief. If I truly believe that p on the ground that q then I must take it that q supports p, and this attitude of taking explains my believing as I do. Thus, if the Taking Condition is a necessary condition on inference, and it applies to more than dynamic processes, then the phenomena we mean to capture with talk of "inference" extend beyond the dynamic.

It is natural to ask in what sense inference is active if it is not a process. An unsatisfactory answer: inference is active because the beliefs formed by it are based on reasons, and any attitude that is based on reasons is active (Raz 1997). I say this answer is unsatisfactory not because it is false but because it is too deflationary. Even skeptics about mental agency like Setiya 2013 and Strawson 2003 can accept this. Indeed, it's hard to see how anyone could deny it. On the other extreme, some philosophers hold that inference is active in precisely the same way that intentional actions are (Wright 2014). This view strikes me as implausible given the clear disanalogies between the cases. An obvious one: drawing an inference is not the result of forming an intention to do so. Once I recognize the evidential force of the premise my hand is forced; there is no intention mediating my assessment of the evidence and my drawing the inference.

On a more moderate view, inference exhibits a *sui generis* form of agency (Hieronymi 2011, Moran 2002, 2012, Boyle 2012). This view can be developed in a number of different ways, but the basic idea is that inference must be thought of as agential due to features that go beyond mere reason-responsiveness. We have to think of inference as an act performed by the subject because the products of inference are *attributable* to the subject herself. This is because they are possessed in virtue of the subject's own assessment of their credentials. And that is why the Taking Condition can be thought to articulate a deflationary, though not trivial, sense of

cognitive agency. I bring it about that I believe that p on the ground that q when I believe that p because I take it that that is what I should believe: my sense of how things should be with my beliefs is explanatory of my believing as I do. (Whether this activity is in any more demanding sense agential, for example, if it is in any sense *free*, is a further matter.) Given that many attitudes are based on reasons but are not based on the subject's appreciation of those reasons, I think there's real teeth to this notion of cognitive agency. Thus, one reason to accept the Taking Condition is that it can explain the intuition that inference is agential without requiring us to understand this in a trivial or overly demanding way.

Although my presentation of this view of cognitive agency has been brief, there are two things to note about it. First, if inference is agential, even in this somewhat minimal sense, then, assuming that agency is a causal concept, inference is causal. This holds even if we deny that it is a causal process, as I recommended. Second, as I mentioned, the Taking Condition applies to static beliefs as well as events of coming to believe. I take that to mean that, insofar as it is a necessary condition on inference, then the phenomena we have in mind include both states of believing for a reason and events of inferring. Accordingly, I will use the phrases "inferential belief" and "belief for a reason" interchangeably. Of course, the latter phrase can be understood as referring to a more general phenomenon, such as when one believes that p on the basis of testimony, perceptual experience, or memory. While it is plausible to think that an analogue of the Taking Condition holds of these as well, I won't take up that issue here.[76] My focus is on beliefs based on other beliefs. So my use of "belief for a reason" will be restricted to these cases.

[76] If an analogue of the Taking Condition holds for beliefs based on perceptual experiences or intuitions generally then it would seem that the attitude of taking cannot be understood as an intuition, on pain of regress.

In this section I've explained what I mean by inference and motivated the Taking Condition. I suggested that the condition offers an attractive account of the way in which inference is an activity performed by the subject. Attractive though the Taking Condition is, it has come under attack. In the remainder of the paper I'll set out the objections and then argue that we can retain the condition by thinking of inference as a self-conscious capacity. One's taking it that q supports that p is explained by the fact that one believes (and, indeed, knows) that q is one's reason for believing p.

4.4 THE ACCESS OBJECTION

I want now to look at an argument for rejecting the Taking Condition, one which focuses on non-deductive inference. Kieran Setiya asks us to imagine a scientist working on an experiment for some time who encounters a large amount of data before drawing a conclusion (Setiya 2011). Imagine a meteorologist who examines records, radar, and other data before concluding that "the storm will miss the region." We want to say that the meteorologist's conclusion is a paradigm of knowledge, but she is unlikely to know all of the evidence on which her belief rests. As Harman puts it: "[i]t is doubtful that anyone has ever fully specified an actual piece of inductive reasoning, since it is unlikely that anyone could specify the relevant total evidence in any actual case" (Harman 1970, 884, cited by Setiya 2011). Many of our beliefs depend on far more evidence than we could ever articulate or survey. On the interpretation of cases like this favored by Harman and Setiya, each piece of data about which the meteorologist forms a belief is a piece of evidence, which functions like a premise in an enormous inductive inference. Consciously believing the conclusion on the basis of the premises in a way conforming to the Taking

Condition would require access to all of the premises. But that much data is inaccessible. Therefore, drawing the inference can't require access to the premises: the Taking Condition is false. Call this the Access Objection, since it denies that we can access all of the premises of our inferences.

In response, we should distinguish two ways in which a bit of knowledge can rationally depend upon evidence or background knowledge. Sometimes evidence or background knowledge functions as a premise in inference. But it is also possible for that knowledge to function as an enabling condition on the ability to justifiably perform inferences on the basis of other premises. It is natural to describe the experiences and beliefs of the meteorologist as providing her with an inferential ability: the ability to tell when a storm will miss the region on the basis of her knowledge about such things as that the temperature is dropping. That ability allows her to conclude that the storm will miss the region on the basis of a limited number of premises that are accessible. Thus, given the background knowledge and expertise she possesses, she is in a position to draw inferences expressible like this: "the temperature is dropping, so the storm will miss the region." Of course, someone without her level of expertise and knowledge could not gain knowledge by means of that inference. But it doesn't follow from this that we have to conceive of her background knowledge as functioning like a premise in her reasoning. We needn't think of her expression of her inference as in any way abbreviated, though in order to be entitled to make it she must implicitly rely on her own competence to make inferences like it.

One might worry that the distinction between mere background beliefs that endow one with an inferential ability and beliefs that are properly premises is not clear enough to do the required work.[77] Suppose that I believe p on the ground that q. In the cases we are considering,

[77] I thank Kieran Setiya for raising this worry.

my ability to draw that inference depends upon my possessing background knowledge, say r. Now we might imagine two scenarios, in one r is "merely" background knowledge and in the other r functions as an explicit premise. It is tempting to say that the difference between these cases is of little epistemic significance. After all, in both cases the epistemic status of the belief that p depends upon r. And one might think that whether r functions explicitly as a premise depends only on pragmatic features of the situation, such as what is salient given the challenges I face from interlocutors.

My response to the access objection is to insist that from the fact that a belief that p depends for its positive epistemic status upon some further belief r, it does not follow that the kind of support lent by r is the same as that offered by a premise. Put another way, it does not follow that the subject inferred that p on the basis of r. I have insisted that there are two ways one belief can depend upon another, whereas the objection insists there is only one. One advantage of the view I am recommending is that it saves the appearances. It certainly seems that when I draw an inference I do so on the basis of a limited number of premises that I can access. It sounds odd to think that, secretly, I have performed an inference so large that it could not be represented by me. Furthermore, Harman is committed to an implausible view about how to individuate inferences. If all of the beliefs upon which my drawing the conclusion rests are premises, then you and I cannot draw the same inference unless we share all of the same beliefs. And it seems clear that two subjects can draw the same inference. Returning to our example, two meteorologists might draw the same conclusion from the fact that the temperatures are dropping. It is surely intuitive to think that they reason in the same way, drawing on the same evidence. Yet it is unlikely in the extreme that they share the exact same body of background knowledge. This is reason to prefer the view I'm recommending.

A final reason to prefer the view I'm recommending is that the idea it makes use of – that background knowledge can play a causal and enabling role in providing a subject with a capacity for knowledge – is independently attractive and might be indispensable for making sense of some cases of noninferential knowledge. I have in mind cases of noninferential knowledge possessed by perceptual means that, strictly speaking, aren't cases of perceptual knowledge. For example, suppose one endorses a restrictive view about the contents of perception on which the only things we strictly speaking perceive are the proper and common sensibles (McDowell 2008). On such a view when an umpire observes a pitch he cannot strictly speaking *see* that it is a strike. Nevertheless, it seems clear that the umpire's ability to tell strikes from balls yields noninferential knowledge. If the Taking Condition is true this must be so. For in these kinds of cases there is no evidence available for the subject to take as a premise. Much background knowledge is required to possess such an ability, but the knowledge doesn't function as a premise in an inference. This view allows us to maintain the Taking Condition, since that condition requires only that the subject be able to access the *premises* in her inference or the *reasons* for her belief. It does not require that she be able to access all of the enabling conditions on her ability to perform that inference or respond to those reasons.

I conclude that the Access Objection does not threaten the Taking Condition so long as we think of our capacity for inference as a sensitivity to the evidential import of an accessible number of one's available beliefs, a sensitivity shaped by background knowledge and understanding. Aside from its force as a response to the objection and its intuitive plausibility, this view opens the way for an attractive account of the taking attitude, which I will set out in the next sections after sketching two more objections to the Taking Condition.

4.5 CARROLL'S AND FUMERTON'S REGRESSES

In this section I will outline two further objections against the Taking Condition, both of which involve the threat of vicious regress.

Remember, the Taking Condition on inference is the following claim:

> *Taking Condition:* Inferring necessarily involves the thinker taking his premises to support his conclusion and drawing his conclusion because of that fact.

The Taking Condition can usefully be divided into two separate theses:

> *Taking:* if you believe that p on the ground that q, you take q to support that p.

> *Explanation:* if you believe that p on the ground that q, you believe that p *because* you take p to support that q.

The Taking Condition is just the conjunction of these claims, but it will help to be able to focus our attention on each separately.

By Explanation the attitude of taking must explain one's inferential belief. When I believe that p on the ground that q I take it that q supports that p. That means that I believe that q supports that p, which belief is among the causes of my belief that p. But this natural gloss on the Taking Condition leads to the following regress, which I'll call Carroll's Regress: [78]

Carroll's Regress

> *Taking:* If you believe that p on the ground that q, you take q to support that p.

> *Taking as Belief:* If you take q to show that p, you believe that q supports that p.

> *Beliefs are Premises:* If a belief is relevant to the rationality of an inference, its content is a premise of that inference.

> *C:* To believe that p on the ground that q is to have an infinite series of beliefs. Given *Explanation* they explain one another in an infinite series.

[78] I call this Carroll's Regress because of the premise "Beliefs are Premises" which seems to play a role in Carroll's paper (Carroll 1895). I'm not claiming that this is exactly what Carroll had in mind.

Beliefs as Premises would seem to follow from the idea that in order for the belief that p to cause the belief that q, the subject must antecedently possess the belief that q supports that p.

It is useful to think of the regress in terms of a procedure. Taking and Taking as Belief require that if some set of beliefs B is to rationally support the belief that p then there must be a further belief that B supports that p. The view that such beliefs must be premises requires us to "place" this belief about reasons in the set B. But this means that we have a *new* set of beliefs linked to the belief that p, which means that we must once again apply the Taking rule and "add" a further belief. On and on without end. Perhaps an infinite series of beliefs isn't impossible. But Explanation claims that these beliefs together explain why the subject believes the conclusion, which is bizarre.

Here is the second regress involving the Taking Condition, which I will call Fumerton's Regress.[79]

Fumerton's Regress

Taking: If you believe that p on the ground that q, you take q to support that p.

Taking as Belief: If you take q to support that p, you believe that q supports that p.

Need for Inference: If you are rational in believing that q supports that p, you must believe this on the basis of an inference from some other belief, r.

C: You must have made a prior inference in order to believe that p on the ground that q. Given Taking and Taking as Belief we can iterate. Inference is impossible.

Unlike Carroll's Regress, Fumerton's Regress does not depend on the causal role of taking in inference, as is clear from the fact that it does not appeal to Explanation. This regress is a regress of justification. Suppose one believes p on the ground that q. Then, by Taking and Taking as Belief, one must also believe "q supports that p." But the belief "q supports that p" must itself be

[79] So-called because of its role in Fumerton's discussion of Inferential Internalism in his 1995.

justified (at least subjectively). After all, if the belief that "q supports that p" doesn't need to be justified why does the inference? (Setiya 2013).[80] The assumption is that beliefs about justificatory relations cannot be justified noninferentially, so they must be justified by inference.[81] This means that one must also infer "q supports that p" on the basis of another belief, say, r. But, again, given Taking and Taking as Belief, one must also believe "r supports that 'q supports that p'", which itself stands in need of justification. And so on.

Many philosophers think that the Taking Condition is the culprit in these regresses and that we are forced to reject it.[82] Consider Carroll's Regress. It would be odd to accept the Taking Condition and reject Explanation. For the sake of argument, we have assumed Taking as Belief.[83] Finally, while Beliefs are Premises is suspicious, it looks like we are forced to accept it once we accept Taking as Belief and Explanation. Regarding Fumerton's Regress, it really is hard to imagine what power of insight could provide us with noninferential knowledge of relations of rational support. If this is right, then we have good reason to reject the Taking Condition. We must countenance justified stretches of reasoning and belief for a reason the rationality of which does not depend upon the subject's take on her reasons, so-called "blind reasoning" (Boghossian 2003).

I think this is the wrong way to respond to the regresses. Given the attractions of the Taking Condition we ought to seek out an interpretation of it that avoids the two regresses. I present my interpretation in the next section.

---

[80] As Setiya points out, one faces this regress worry even if one denies that taking is a doxastic attitude, as in Boghossian 2014 and Broome 2014. Setiya 2013.
[81] While some might countenance noninferential knowledge of general truths of probability or justificatory support, few would think we have noninferential knowledge of specific justificatory relations, such as that holding between the propositions that the streets are wet and that it has rained. See Fumerton 1995 for discussion of the Keynesian view of a priori knowledge of truths of probability.
[82] A sample: Johnston 1988, Boghossian 2003, Wedgwood 2006, Railton 2006, and Setiya 2011, 2013.
[83] Again, those who endorse an intuitional view of the Taking Condition would reject Taking as Belief. I think this view can't be right, for reasons I've indicated, but am not going to take it up here.

4.6 INFERENCE AS A SELF-CONSCIOUS ACTIVITY

Carroll's Regress crucially involves the claim I have called Explanation. If we are to retain the Taking Condition and avoid the regress we must reject Beliefs are Premises and replace it with a positive proposal for how taking it that q is a reason to believe p explains one's believing p. Fumerton's Regress is a regress of justification; it is generated by the assumption that beliefs about rational support between propositions in non-deductive inference must be justified inferentially. So if the regress is to be avoided there must be a noninferential source of such justification; we must reject Need for Inference and replace it with a positive proposal. In this section I'll outline the positive proposal.

Remember, Self-Intimation is the following thesis:

> *Self-Intimation*: there is a class of mental states M such that, if one is in M, one thereby knows that one is in M.

Suppose that Self-Intimation is true for inferential belief. A subject who believes that p on the ground that q thereby knows this. She does not know this as a result of a procedure followed *after* drawing the inference, rather she possesses the self-knowledge *in* or *by* drawing the inference. Believing that p on the ground that q is metaphysically sufficient for knowing that one believes that p on the ground that q. Now nothing in what follows will turn on one possessing self-*knowledge* in these cases, and so we can assume the weaker claim that inferential belief is sufficient for justified belief about one's belief and its grounds. Thus, believing that p on the ground that q is sufficient for believing that one so believes.[84]

---

[84] I make this qualification so that the view I recommend won't seem to solve the regress problems trivially. I'll argue that self-knowledge can constitute one's taking it that one believes p for the reason that q. By the factivity of

My proposal is that taking it that q supports that p consists in first-personally believing that one believes p on the ground that q. Call this claim Taking as Self-Belief:

Taking as Self-Belief (TSB): if you first-personally believe that q is your reason for believing p, then you take it that q supports that p.

It is important to emphasize that TSB is restricted to first-personal self-beliefs, the kind explained by Self-Intimation. It does not hold, absurdly, that any belief about one's reasons establishes that those are one's reasons.

The conjunction of TSB and Self-Intimation satisfies Taking. By TSB a self-belief is the source of the attitude of taking; by Self-Intimation we necessarily possess that self-belief whenever we possess an inferential belief. Therefore, whenever one believes that p on the ground that q, one takes it that q supports that p. And this proposal satisfies Explanation. Self-belief explains inference because it is a constitutive feature of one's capacity for inference. As we might put it, to say of the subject that she takes it that the premise supports the conclusion is just to say that she draws the inference self-consciously. Think again of the meteorologist. Upon noticing that the temperatures are dropping she concludes that the storm will miss the region. While she might explicitly think about the connection between these two propositions, this is not necessary.[85] It certainly seems to be the case that we often draw inferences without any conscious attention to propositions of that kind. Assuming that the taking attitude is not an unconscious or subpersonal state playing the role of a premise, in what sense does the meteorologist take her premise to support her conclusion? It is plausible that in cases like this her taking the premise as support is shown by the fact that if you asked her why she believes that the

---

knowledge it follows that if I know that I believe p on the ground that q then I do believe on that ground. But the view I'm defending isn't trivial, it's that some beliefs about one's grounds are guaranteed to be true (and so, presumably count as knowledge) because they constitute one's believing for those reasons. To avoid this confusion I'll speak in terms of self-belief.

[85] That is, her inference might be a case of what philosophers call "basic" inference (Wedgwood 2006 and Valaris 2012).

storm will miss the region she will say “because the temperatures are dropping.” That is, if you were to ask the subject why she believes the conclusion, she will cite the premises. That makes it plausible to think that TSB can account for the attitude of taking. But we need Self-Intimation in order to guarantee that in inference the subject possesses the relevant self-belief, since the Taking Condition is a necessary condition on inference. The conjunction of TSB and Self-Intimation satisfies both Taking and Explanation, and so satisfies the Taking Condition.

The proposal is an intelligible interpretation of the Taking Condition and is supported by reflection on examples like the meteorologist. It can also be motivated by an attractive picture of the role of the subject in justifying her beliefs. Earlier I mentioned Harman’s claim that nobody can make any inductive inference explicit since each depends upon an unwieldy amount of evidence. I argued that we aren’t forced to accept the picture of inference he assumes. We should reject it in order to maintain the Taking Condition. Another reason to reject it is that it depicts the subject in an alienated relationship to her own reasons. On Harman’s view when a subject justifies her beliefs by citing reasons she is attempting to offer a causal explanation of how she came to believe what she did, an explanation that is independent of her believing what she does. Adam Leite refers to this as “the Spectatorial Conception.” It holds that the facts that justify a belief “are in place independently and not directly affected by what goes on when the person attempts to justify the belief” (Leite 2004, 225). As Leite convincingly argues, that conception is mistaken because it assumes that the justifications we offer for our beliefs are merely indicative of our reasons. But it is clear that “we generally don’t treat what people say in defense of their beliefs as merely evidence about, or an indication of, the reasons for which they hold their beliefs, but rather as a direct expression of their rational activity” (ibid. 227). We should reject the Spectatorial Conception and recognize that beliefs about our reasons are not *guesses* but are,

typically, *constitutive* of believing for those reasons. This means that a belief about reasons can be sufficient for the attitude of taking, as TSB claims.

Leite also endorses the converse claim: “if someone is justified in believing as he does in virtue of basing his belief upon good reasons, then, in the absence of any special circumstances preventing him from doing so, he must be able to provide those reasons in defense of his belief” (Leite 238). It is plausible to think that anyone who rejects the Spectatorial Conception outright should also accept this claim. And Self-Intimation offers an explanation of the necessary connection between believing for reasons and being able to offer those reasons in defense of one’s belief: the subject is able to provide reasons in defense of his belief because he comes to possess that belief self-consciously, in the knowledge of his reasons as his reasons. So rejecting the Spectatorial Conception gives us reason to accept TSB and Self-Intimation.

Assuming that the proposal is a viable interpretation of the Taking Condition and is attractive on other grounds, how does it avoid the regresses?

Start with Carroll’s Regress. As I mentioned earlier, it is natural to think of taking as a doxastic attitude contemporaneous with drawing an inference. By Explanation and Beliefs are Premises we should then think of it as a belief with which one reasons in inference. But we can deny this and the proposal explains why. To say that a subject possesses self-knowledge in or by inferring does not mean that she reasons with her self-knowledge as a premise. Self-knowledge is not a contemporaneous attitude “inside” the inference, it is a way of characterizing the inference itself. And if that is right then we have found a place for the attitude of taking that is not that of a premise. We can deny Beliefs are Premises and avoid Carroll’s Regress. (Notice that this does not require us to reject Taking as Belief.)

The proposal also allows us to avoid Fumerton's Regress. By Self-Intimation and Taking as Belief, a subject who believes that p on the ground that q will justifiably believe or know that she believes that p on the ground that q. In the good case, this will involve knowing that q and coming to know that p by inference. In virtue of believing for a reason, a subject knows that q is her reason for believing p, she is self-conscious of her belief and its ground. But if she believes that p on the ground that q, and this is the exercise of a capacity to gain knowledge, it follows, in the good case, that q really is a reason to believe p. Thus, self-knowledge can be a source of justification for beliefs about rational relations between propositions. Since the subject possesses self-knowledge noninferentially, so too does she possess noninferential justification for the taking attitude.

One might think that the subject needs to possess justification that she is in the good case. If such justification had to be independent of drawing the inference this would be a problem for the proposal, for it would mean that our capacity for self-knowledge could not be the source of noninferential justification about the rational relations between propositions. However, one can deny that independent justification is needed. There are two possibilities here. First, one might appeal to a claim about self-trust along the following lines:

> *Self-Trust:* A rational creature possesses a default or a priori entitlement to believe that her belief-forming capacities are reliable.

If Self-Trust is true then a subject needn't possess justification for believing that she is in the good case. Of course, that does not mean that she is guaranteed to be in the good case. It just means that she needn't prove that she is. Second, one might think that our capacity for self-knowledge itself provides the relevant justification, though not independently of the goodness of the inference in question. On this view when the subject knows that she believes that p on the ground that q, she represents this as the exercise of a capacity to gain knowledge. In the good

case, she thereby knows that she possesses knowledge that p. We could call this view Disjunctivism about inferential knowledge. For our purposes it won't matter which of these one accepts, as both respond to the objection.

In this section I've outlined my proposal of interpreting the Taking Condition as the conjunction of Taking as Self-Belief and Self-Intimation. I've argued that the proposal is a viable interpretation of the condition, that it is supported by examples, and that it is supported by an attractive picture of the place of the subject in justifying her beliefs. Finally, I argued that it can allow us to avoid the regress worries surveyed in the last section. I'll conclude in the next section by considering two objections to this proposal.

4.7 OBJECTIONS

As mentioned earlier, Self-Intimation is a strong claim, which many philosophers will want to deny. In this paper I have defended it against objections and offered a reason to accept it. Self-Intimation, in conjunction with TSB, can allow us to maintain the Taking Condition in the face of regress worries. In this section I'll consider two objections, one directed at each of the conjuncts. First, one might think that our capacity for self-knowledge can salvage the Taking Condition even if Self-Intimation is false. Second, TSB is incompatible with widely held views about rational explanation.

The first objection is that what I have called the Propositional View can adequately explain the Taking Condition: the argument here gives us no reason to endorse the latter. I have argued that taking q as a reason to believe p consists in believing that q is one's reason for

believing p. The key move was to deny that in inference one reasons with the taking attitude. It is enough that it be true of the subject who draws an inference that, if asked, she would cite q as her reason. And one might interpret this as the claim that a subject who believes p on the ground that q must be *in a position* to believe by reflection alone that q is her reason for believing p. But the Propositional view can explain this. It holds that being in a first-order state puts one in a position to know that one is. Thus, believing p on the ground that q puts one in a position to know that q is one's reason for believing p (Setiya 2011). And the defender of the Propositional view can take advantage of the same claim about self-trust in order to explain how such a subject possesses noninferential justification for her belief about the support lent by her premises.

The problem with this alternative is that it cannot make sense of Explanation. By Explanation the taking attitude must explain one's drawing the conclusion one does. And that means that at the time one believes or comes to believe the conclusion one must also possess the taking attitude. But the Propositional view denies this. It holds only that when the subject draws an inference she is *in a position* to know of her reasons and so only *in a position* to possess the taking attitude. But an attitude one is in a position to possess cannot explain one's right now actually possessing a different attitude. If self-knowledge is the source of the taking attitude, which is necessarily present in inference, then self-knowledge must be possessed in virtue of possessing an inferential belief. Thus, self-knowledge is a metaphysically necessary condition on inferential belief, as Self-Intimation claims.

The defender of the Propositional view may respond by rejecting Explanation. She could do this by denying that inference is causal. On this view, TSB is true: believing that q is one's reason for believing p is sufficient for believing for that reason. Yet it is not true that one's belief

that q is a cause of one's believing that p. To believe that p on the ground that q just is to believe that q is one's reason for believing p (Lehrer 1971, Setiya 2013, Neta 2014).

I find this response unattractive since it requires that we give up on the idea that inference is agential. Regardless, it doesn't help with the problem. For even if one denies that inference is a causal phenomenon, an analogue of Explanation still holds. That is, if one accepts the Taking Condition then one should insist that when one believes that p on the ground that q one believes that p because one takes it that q supports that p. One can hold this even if one denies that the "because" in question is causal. The objection to the Propositional view does not assume the taking attitude causes one's belief in the conclusion, it only depends upon the former explaining the latter. And even if the explanation is not causal, it is still the case that when one believes that p on the ground that q one must thereby actually possess the taking attitude. This requirement is not met by being in a position to possess it. So the conjunction of the Propositional view and TSB is not a viable interpretation of the Taking Condition.

In his discussion of the Taking Condition Paul Boghossian considers a view like TSB, which he calls "the Counterfactual Proposal." According to Boghossian, the view cannot be correct because "it treats the property of being an inference as though it were a response-dependent property, which I believe it cannot be" (Boghossian 2014, 10). Roughly, a property F is response-dependent just in case a subject's (or group of subjects') dispositions to say or believe that an object is F, in a context, analytically entails that it is. A familiar example: on a response-dependent theory of color an object is red if and only if suitably placed observers are disposed to experience it as red. TSB claims that one believes that p on the ground that q if and only if one first-personally believes that q is one's reason for believing p. Therefore, TSB holds that whether

something is an inference depends upon the responses of the subject. So Boghossian is right that it conceives of inference as a response-dependent property. And Boghossian thinks inference cannot be a response-dependent property, because "saying that R was his reason can't be *constitutive* of R's being his reason" (Boghossian 2014, 10).

Here is one way into Boghossian's worry. Suppose that you have spent some time in therapy and are asked to reflect on some of your beliefs. In performing this task you come up with an interpretation, an attempt to explain how you came to believe what you do. This could appeal to rational or nonrational considerations. The important point is that the interpretation is something like your best guess of how you came to your current condition. Although you yourself are the subject of your interpretation, your activity is really no different than that which your therapist or friend might perform in coming up with a rival interpretation. And just as it would be absurd to think there was a constitutive connection between your therapist's interpretation of your reasons for belief and your actually believing for those reasons, so too is it absurd to say the same about your interpretation.

It should be clear, then, that Boghossian objects to TSB because he assumes what Leite calls "the Spectatorial Conception." He assumes that beliefs about one's reasons are always and everywhere guesses about causes. But if, following Leite, we reject that conception and accept that believing that q is one's reason can be constitutive of its being one's reason, then we will not be swayed by Boghossian's worry.

However, Boghossian's objection is an instance of a more general point. Inference, he says, is a causal concept, and causal concepts cannot be response-dependent (ibid. 10). The thought here, I take it, is that there cannot be a constitutive connection between the relata in a causal relation. Since belief for a reason is causal there cannot be a constitutive connection

between your beliefs about your reasons and your actual reasons. So we can understand Boghossian's objection in two ways. First, he assumes a picture of what beliefs about reasons are like. Beliefs about one's reasons are attempts to discover the causes of one's beliefs. As such, they are prone to error and so cannot be constitutive of believing for reasons. Second, he assumes a picture of causation on which there cannot be constitutive relations between causal relata. Rejecting the Spectatorial Conception will only help in dealing with the first version of this objection. It leaves the second in place.

To summarize, the following three claims are inconsistent:

> *Taking as Self-Belief (TSB):* if you first-personally believe that q is your reason for believing p, then you take it that q supports that p.
>
> *Explanation:* if you believe that p on the ground that q, you believe that p *because* you take it that q supports that p.
>
> *Causation is not Constitution (CNC):* there cannot be a constitutive connection between causal relata.

By TSB there is a constitutive connection between beliefs about reasons and beliefs held for those reasons. By Explanation that connection is causal. By CNC no connection can be both.

Boghossian's response is to reject TSB. As we have seen other philosophers would reject Explanation (Lehrer 1971, Setiya 2013, Neta 2013). However, given the attractions of the Taking Condition and the picture of inference as agential, it seems preferable to hold on to Explanation. So I think we should reject CNC.

We have already seen a sketch of how this might work. In response to the Access Objection I recommended conceiving of our capacity for inference as a causal sensitivity to the epistemic significance of an accessible number of beliefs, thanks to background knowledge and understanding. Appealing to the capacity offers a causal explanation of inferential belief. In response to the regresses I suggested we should qualify this to think of the relevant capacity as

self-conscious, as constitutively involving self-beliefs. What I want to say here is that, since the capacity is self-conscious, self-belief, and so the taking attitude, can be thought of as causally responsible for inferential belief as well. Because it characterizes the capacity that yields inferential beliefs, self-knowledge can be a cause of those beliefs, even though it constitutively entails them. This is, of course, a controversial and difficult topic. The sketch is not close to a conclusive response to the worry. But I do think it gives reason to take seriously the possibility that the explanation of belief by reasons is an exception to CNC.[86]

The goal of this chapter has been to argue that we can retain the Taking Condition in the face of objections so long as we embrace a Constitutivist account of self-knowledge of inferential belief. Self-knowledge can play a necessary explanatory role in rational inference. Now the force of this argument will depend a great deal on whether there are other, less contentious interpretations of the Taking Condition available. I have not taken up this issue here. Suppose, though, that there aren't any other contenders. Someone unsympathetic to the argument of this paper might take it as a *reductio* on the Taking Condition. However, giving up on this comes at a high cost: it requires giving up on the idea that the subject is herself involved in her own reasoning, a core feature of our self-conception as rational creatures (Boghossian 2014). So one way to take the argument of this paper would be the following: in order to hold onto this self-conception we must hold that the subject is *thoroughly* involved in her reasoning. There may be no middle ground between, on the one hand, accepting the Taking Condition and endorsing Constitutivism and, on the other, denying the condition and with it the idea of inference as the subject's own activity.

[86] See Marcus 2012.

BIBLIOGRAPHY

Austin, J.L. 1979. Other Minds. In his *Philosophical Papers.* Oxford: OUP.

Baker, L. 1995. *Explaining Attitudes: A Practical Approach to the Mind.* Cambridge: Cambridge University Press.

Bar-On, D. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford: OUP.

----- 2009. First-Person Authority: Dualism, Constitutivism, and Neo-Expressivism. *Erkenntnis*. 71: 53-71.

----- 2010a. Expressing as "Showing what's within": On Mitchell Green's *Self-Expression. Philosophical Books.* 51 (4): 212-227.

----- 2010b. Avowals: Expression, Security, and Knowledge: Reply to Matthew Boyle, David Rosenthal, and Maura Tumulty. *Acta Analytica*. 25: 47-63.

Benbaji, H. 2013. How is Recalcitrant Emotion Possible? *Australasian Journal of Philosophy.* 91 (3): 577-599.

Berker, S. 2008. Luminosity Regained. *Philosophers' Imprint.* 8 (2): 1-22.

Bilgrami, A. 2006. *Self-Knowledge and Resentment.* Cambridge, MA: Harvard University Press.

Boghossian, P. 1989. Content and Self-Knowledge. *Philosophical Topics.* 17: 5-26.

----- 2003. Blind Reasoning. *Proceedings of the Aristotelian Society, Supplementary Volume.* 77: 225-248.

----- 2014. What is Inference? *Philosophical Studies* 169 (1): 1-18.

Boyle, M. 2009. Two Kinds of Self-Knowledge. *Philosophy and Phenomenological Research*. 78 (1): 133-164.

----- 2010. Bar-On on Self-Knowledge and Expression. *Acta Analytica.* 25: 9-20.

----- 2011. Transparent Self-Knowledge. *Proceedings of the Aristotelian Society, Supplementary Volume.* 85 (1): 223-241.

----- 2012. Active Belief. *Canadian Journal of Philosophy.* 39 (1): 119-147.

----- forthcoming. Additive Theories of Rationality: A Critique. *European Journal of Philosophy.*

Brandom, R. 1994. *Making it Explicit.* Cambridge, MA: Harvard University Press.

Broome, J. 2014. Comments on Boghossian. *Philosophical Studies* 169: 19-25.

Burge, T. 1996. Our Entitlement to Self-Knowledge. *Proceedings of the Aristotelian Society.* 96 (1): 91-116.

Byrne, A. 2005. Introspection. *Philosophical Topics.* 33 (1): 79-104.

----- 2011. Transparency, Belief, Intention. *Proceedings of the Aristotelian Society, Supplementary Volume.* 85 (1)*:* 201-221.

Call, J. 2010. Do Apes Know that they can be Wrong? *Animal Cognition* 13: 689-700.

Carroll, L. 1895. What the Tortoise said to Achilles. *Mind* 4: 278-280.

Carruthers, P. 2011. *The Opacity of Mind*. Oxford: OUP.

Carruthers, P. and Ritchie, J. 2012. The Emergence of Metacognition: Affect and Uncertainty in Animals. In M. Beran et al. (Eds.) *Foundations of Metacognition*. Oxford: OUP. 211-237.

Cassam, Q. 2007. *The Possibility of Knowledge.* Oxford: OUP.

----- 2009. The Basis of Self-Knowledge. *Erkenntnis.* 71: 3-18.

----- 2015. *Self-Knowledge for Humans.* Oxford: OUP.

Chudnoff, E. 2014. The Rational Roles of Intuition. In A. Booth and D. Rowbottom (Eds.) *Intuitions*. Oxford: Oxford University Press. 9-35.

Coliva, A. 2012. One Variety of Self-Knowledge: Constitutivism as Constructivism. In A. Coliva (Ed.) *The Self and Self-Knowledge.* Oxford: OUP.

Collins, A. 1988. *The Nature of Mental Things.* South Bend, IN: Notre Dame University Press.

Comrie, B. 1976. *Aspect*. Cambridge: Cambridge University Press.

Dancy, J. 2000. *Practical Reality.* Oxford: OUP.

Davidson, D. 1980. *Essays on Actions and Events.* Oxford: OUP.

----- 1986. Deception and Division. In J. Elster (Ed.) *The Multiple Self.* Cambridge: Cambridge University Press.

Dogramaci, S. 2013. Intuitions for Inferences. *Philosophical Studies.* 165 (2):371-399.

Dretske, F. 1969. *Seeing and Knowing*. London: Routledge & Kegan Paul.

Falvey, Kevin. 2000. The Basis of First-Person Authority. *Philosophical Topics.* 28 (2): 69-99.

Fernandez, J. 2013. *Transparent Minds*. Oxford: OUP.

Finkelstein, D. 1999. On Self-Blindness and Inner Sense. *Philosophical Topics*. 26 (1): 105-119.

----- 2003. *Expression and the Inner.* Cambridge, MA: Harvard University Press.

Fodor, J. 1987. *Psychosemantics.* Cambridge, MA: MIT Press.

Frege, G. 1979. Logic. In *Posthumous Writings*. Chicago: University of Chicago Press.

Freud, S. 1966. *Introductory Lectures on Psycho-Analysis.* New York: W.W. Norton & Co.

Fumerton, R. 1995. *Metaepistemology and Skepticism.* London: Rowan & Littlefield.

Geach, P.T. 1957 *Mental Acts.* London: Routledge & Kegan Paul.

----- 1965. Assertion. *The Philosophical Review.* 74: 449-465.

Gopnik, A. 1993. How we Know our Minds: the Illusion of First-Person Knowledge of Intentionality. *Behavioral and Brain Sciences*. 16 (6): 1-15, 90-101.

Green, Mitchell. 2007. *Self-Expression.* Oxford: OUP.

Hampshire, S. 1979. Some Difficulties in Knowing. In T. Honderich and M. Burnyeat, (Eds.) *Philosophy as it is.* Harmondsworth: Penguin.

----- 1975. *Freedom of the Individual*. Princeton, NJ: Princeton University Press.

Hampton R. 2001. Rhesus Monkeys Know when they Remember. *Proceedings of the National Academy of Sciences of the United States of America*. 98 (9): 5359-5362.

Harman, G. 1974. *Thought.* Princeton, NJ: Princeton University Press.

Heal, 1994. Moore's Paradox: A Wittgensteinian Approach. *Mind*. 103: 5-24.

Hieronymi, P. 2011. Believing at Will. In D. Hunter (Ed.) *Belief and Agency*. Calgary: University of Calgary Press. 149–187

Hills, A. 2009. Moral Testimony and Moral Epistemology. *Ethics*. 120 (1): 94-127.

Hursthouse, R. 1991. Arational Actions. *The Journal of Philosophy.* 88 (2): 57-68.

Hyman, J. 1999. How Knowledge Works. *The Philosophical Quarterly.* 49 (197): 433-451.

----- 2001. Knowledge and Self-Knowledge. In S. Schroeder (Ed.) *Wittgenstein and Contemporary Philosophy of Mind.* London: Palgrave. 171-193.

----- 2006. Knowledge and Evidence. *Mind.* 115 (460): 891-916.

----- 2011. Acting for Reasons: Reply to Dancy. *Frontiers of Philosophy in China.* 6 (3): 358-368.

Jacobsen, R. 1996. Wittgenstein on Self-Knowledge and Self-Expression. *Philosophical Quarterly.* 46 (182): 12-30.

Johnston, M. 1988. Self-Deception and the Nature of Mind. In B. McLaughlin and A. Rorty (Eds.) *Perspectives on Self-Deception*. Berkeley, CA: University of California Press. 63-91.

Kant, I. 1998/1787. *Critique of Pure Reason.* P. Guyer and A. Wood (Eds.) Cambridge: Cambridge University Press.

Kauppinen, A. 2010. The Pragmatics of Transparent Belief Reports. *Analysis*. 70 (3): 438-446.

Kenny, A. 1973. *Wittgenstein.* Harmondsworth: Penguin.

Kind, A. 2003. Shoemaker, Self-Blindness and Moore's Paradox. *The Philosophical Quarterly*. 53 (210): 39-48.

Lehrer, K. 1971. How Reasons Give us Knowledge, or the Case of the Gypsy Lawyer. *Journal of Philosophy.* 68 (10): 311-313.

Leite, A. 2004. On Justifying and Being Justified. *Philosophical Issues* 14: 219–253.

Lewis, D. 1986. Postscript to Causation. In his *Collected Papers, Volume II.* Oxford: OUP. 172-213.

Marcus, E. 2012. *Rational Causation.* Cambridge, MA: Harvard University Press.

Maynard Smith, J. and D. Harper. 2004. *Animal Signs.* Oxford: OUP.

McDowell, J. 1982. Criteria, Defeasibility, and Knowledge. *Proceedings of the British Academy.* 68: 455-479.

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

----- 1998a. Response to Crispin Wright. In C. Wright, B.C. Smith, & C. Macdonald (Eds.) *Knowing our own Minds*. Oxford: Clarendon Press.

----- 1998b. On 'The Reality of the Past'. Reprinted in his *Meaning, Knowledge, and Reality.* Cambridge, MA: Harvard University Press. 295-313.

----- 2008. Avoiding the Myth of the Given. Reprinted in his *Having the World in View: Essays on Kant, Hegel, and Sellars.* Cambridge, MA: Harvard University Press. 256-272.

McGinn, C. 1982. *The Character of Mind*. Oxford: OUP.

Merleau-Ponty, M. 1945/2012. *Phenomenology of Perception.* Donald Landes (trans.) London: Routledge.

Moran, R. 2001. *Authority and Estrangement.* Princeton, NJ: Princeton University Press.

----- 2012. Self-Knowledge, ‘Transparency’, and the Forms of Activity. In D. Smithies and D. Stoljar (Eds.) *Introspection and Consciousness.* Oxford: OUP. 211-237.

Neta, R. 2013. What is an Inference? *Philosophical Issues* 23: 388–407.

Nisbett, R. and Wilson, T. 1977. Telling More than we Know: Verbal Reports on Mental Processes. *Psychological Review.* 84: 231-259.

O’Shaughnessy, B. 2000. *Consciousness and the World.* Oxford: OUP.

Peacocke, C. 2014. *The Mirror of the World: Subjects, Consciousness, and Self-Consciousness.* Oxford: OUP.

Pitt, D. 2004. The Phenomenology of Cognition: or, What is it Like to Think that p? *Philosophy and Phenomenological Research.*69 (1): 1-36.

Railton, P. 2006. How to Engage Reason: the Problem of Regress. In R. J. Wallace, P. Pettit, S. Scheffler, and M. Smith (Eds.) *Reason and Value*. Oxford: Oxford University Press. 176-201.

Raz, J. 1997. When we are Ourselves: The Active and the Passive. *Proceedings of the Aristotelian Society, Supplementary Volume.* 71: 211-227.

Ryle, G. 1949/2000. *The Concept of Mind.* Chicago: University of Chicago Press.

Rödl, S. 2007. *Self-Consciousness.* Cambridge, MA: Harvard University Press.

Rorty, R. 1979. *Philosophy and the Mirror of Nature.* Princeton, NJ: Princeton University Press.

Sartre, J.P. 1943/1956. *Being and Nothingness*. H.E. Barnes. (trans.) New York: Philosophical Library.

Scanlon, T. 1998. *What we Owe to Each Other.* Cambridge, MA: Harvard University Press.

Schafer, K. 2013. Perception and the Rational Force of Desire. *The Journal of Philosophy.* 110 (5): 258-281.

Schwitzgebel, E. 2013. A Dispositional Approach to the Attitudes: Thinking Outside the Belief Box. In N. Nottelman (Ed.) *New Essays on Belief.* New York: Palgrave MacMillan.

Setiya, K. 2007. *Reasons without Rationalism*. Princeton, NJ: Princeton University Press.

----- 2011. Knowledge of Intention. In *Essays on Anscombe's* Intention. A. Ford, J. Hornsby, and F. Stoutland (Eds.) Cambridge, MA: Harvard University Press. 170-197.

----- 2013. Epistemic Agency: Some Doubts. *Philosophical Issues.* 23 (1): 179-198.

----- 2014. What is a Reason to Act? *Philosophical Studies.* 167 (2): 221-235.

Shoemaker, S. 1996. *The First Person Perspective and Other Essays.* Cambridge: Cambridge University Press.

----- 2012. Self-Intimation and Second-Order Belief. In D. Smithies and D. Stoljar (Eds.) *Introspection and Consciousness.* Oxford: OUP.

Siegel, S. 2012. *The Contents of Visual Experience.* Oxford: OUP.

Siewert, C. 2003. Self-Knowledge and Rationality: Shoemaker on Self-Blindness. In *Privileged Access*. B. Gertler (Ed.) Burlington, VT: Ashgate.

Silins, N. 2012. Judgment as a Guide to Belief. In D. Smithies and D. Stoljar (Eds.) *Introspection and Consciousness.* Oxford: OUP.

Smithies, D. 2012. A Simple Theory of Introspection. In D. Smithies and D. Stoljar, (Eds.) *Introspection and Consciousness.* Oxford: OUP.

Stout, Rowland. 2010. Seeing the Anger in Someone's Face. *Proceedings of the Aristotelian Society Supplementary Volume*. 84: 29-43.

Strawson, G. 2003. Mental Ballistics *or* the Involuntariness of Spontaneity. *Proceedings of the Aristotelian Society* 103: 227–256.

Thompson, M. 2012. Anscombe's *Intention* and Practical Knowledge. In A. Ford, J. Hornsby, and F. Stoutland (Eds.) *Essays on Anscombe's* Intention. Cambridge, MA: Harvard University Press.

Tomasello, M. 2014. *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

Tumulty, M. 2010. Showing by Avowing. *Acta Analytica.* 25: 35-46.

Unger, P. 1979. *Ignorance: A Case for Skepticism.* Oxford: Clarendon Press.

Urmson, J. O. 1952. Parenthetical Verbs. *Mind* 61: 480–96.

Valaris, M. 2014. Reasoning and Regress. *Mind* 123: 101-127.

Wedgwood, R. 2006. The Normative Force of Reasoning. *Nous.* 40 (4): 660-686.

Williamson, T. 2000. *Knowledge and its Limits.* Oxford: OUP.

Wittgenstein, L. 1953/2009. *Philosophical Investigations.* Fourth Edition, Hacker & Schulte (trans.) Oxford: Blackwell.

----- 1980. *Remarks on the Philosophy of Psychology, Volume 1.* Chicago, IL: University of Chicago Press.

Wright, C. 1989. Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intentions. Reprinted in his *Rails to Infinity.* Cambridge, MA: Harvard University Press.

----- 1998. Self-Knowledge: The Wittgensteinian Legacy. In C. Wright, B.C. Smith, & C. Macdonald (Eds.) *Knowing our own Minds*. Oxford: Clarendon Press.

----- 2014. Comment on Paul Boghossian, "The Nature of Inference". *Philosophical Studies* 169: 27-37.

Zimmerman, A. 2005. Basic Self-Knowledge: Answering Peacocke's Criticisms of Constitutivism. *Philosophical Studies*. 128 (2): 337-379.