

INTENTIONALITY AND SELF-KNOWLEDGE

By

CASEY WOODLING

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

© 2011 Casey Woodling

ACKNOWLEDGMENTS

I thank Professor John Biro for shepherding this project to completion. Without his philosophical guidance and support, this project would likely not have been completed. I also thank the members of my dissertation committee, Professors Robert D'Amico, Charles Gattone, and Gene Witmer, for their helpful comments and insightful suggestions on the content of this work.

I thank my former fellow graduate students and my professors at the University of Florida for creating a stimulating and challenging intellectual environment in the department. I would be remiss not to thank Virginia Dampier, Program Assistant in the Philosophy Department, for the crucial and kind assistance she provided at key points in this project.

I would also like to thank my entire family for all their support during my time in graduate school. Last, I must thank my wife, Emily, who was a constant source of inspiration and joy to me during this project.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS	3
ABSTRACT	7
CHAPTER	
1 INTRODUCTION	9
A Sketch of the Problem	9
Aspects of the Cartesian Theory of Mind Worth Preserving	10
Why the Debate has seemed Intractable	14
The Basic Program of the Essay	20
2 SELF-KNOWLEDGE OF INTENTIONAL CONTENT	26
Notions of Self-Knowledge	27
The Transparency of Intentional Content	27
The Asymmetry of Access	29
The Transparency Condition	31
First-Person Access to Intentional Content	38
First-Person Authority	42
How these Notions are Related	46
Clarifying the Picture	47
A Transcendental Argument about Justification	51
Engagement and rationality	54
Why rational agents must be viewed as having justified second-order beliefs about first-order intentional states	58
Why speakers must be viewed as rational agents by interpreters if there is successful communication.	60
A Transcendental Argument about First-Person Access	62
Explaining Self-Knowledge	65
3 EXTERNALISMS	68
Thinking about Twin Earth	68
Thinking about Arthritis	69
Understanding why the Arguments have been so Popular	72
Natural Kind Externalism	73
Social Externalism	82
How many Externalisms?	85
Singular Thought Externalism	85
Transcendental Externalism	87

4	QUESTIONS OF COMPATIBILITY	91
	The Incompatibility of Social Externalism and First-Person Access to Intentional Content.....	91
	The Incompatibility of all forms of Content Externalism and First-Person Access to Intentional Content	97
	An Argument for the Compatibility of Transcendental Externalism and First-Person Access to Intentional Content.....	101
	Addressing the Larger Debate	102
	Boghossian's Incompatibilist Argument	102
	Challenges to Boghossian's argument.....	104
	How to understand the importance of Boghossian's argument.....	107
	Other Incompatibilist Arguments.....	113
	Phenomenal Properties and Privileged, First-Person Access to Intentional Content.....	113
	Content Externalism and <i>A Priori</i> Access to the External World.....	114
	Internalism and Externalism about Justification.....	115
5	RETHINKING CONTENT EXTERNALISM	119
	Rethinking Social Externalism	119
	The Limits of the Disquotation Principle	120
	Rethinking Natural Kind Externalism	127
	Zemach's Response.....	128
	Crane's Response	130
	Taking stock	134
	Rethinking Singular Thought Externalism	136
	Arguments for Singular Thoughts being Individuated Externalistically	137
	From <i>de re</i> concepts to content externalism	138
	From neo-Russellian singular thoughts to content externalism.....	142
	<i>De re</i> and <i>de dicto</i> ascriptions.....	147
	Are there Singular Thoughts?.....	149
	Diagnosing the Arguments.....	150
	Content Externalism, Semantic Externalism	152
6	LIVING WITH CONTENT INTERNALISM.....	155
	Objections to Content Internalism.....	156
	The Need for a Private Language to Describe Thoughts	156
	Paving the Route to Skepticism about the External World?	161
	Content internalism and skepticism about the external world	162
	Content externalism and our knowledge of the external world.....	165
	What about Physicalism?.....	167
	Approaches to Internalistic (Narrow) Content	171
	Phenomenal Intentionality	171
	Why might intentional contents depend on phenomenal properties?	174

How should we understand unconscious intentional states and phenomenal consciousness?.....	178
Are there any sound arguments for the connection principle?	180
Intentionality and Aspectual Shape	184
Aspectual shape that does not depend on phenomenal properties	186
Concluding Remarks.....	189
LIST OF REFERENCES	193
BIOGRAPHICAL SKETCH.....	200

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

INTENTIONALITY AND SELF-KNOWLEDGE

By

Casey Woodling

December 2011

Chair: John Biro
Major: Philosophy

This essay concerns an important intersection of the philosophy of mind and epistemology: the connection between intentionality and self-knowledge. "Intentionality" is a philosopher's term of art used to refer to a system's ability to represent something. In this essay, we will be concerned with the ability of minds to represent the world. The concept of self-knowledge is more familiar than the concept of intentionality; before any philosophy, we are all acquainted with the idea that we know our own minds in a way that others cannot. I do not have to hear what I say, read my words, observe my behavior, and make inferences based on that evidence to know what I think, although I must do these things to know what you think. Each of us knows what we think in a way that others do not.

This essay is concerned with finding a theory of intentionality that is compatible with self-knowledge. The theories of intentionality we will consider concern the degree to which the content of our intentional states depends on environmental factors. One theory, content externalism, says that the content of our intentional states depends on the environment in an important way. The other theory, content internalism, says that

the content of our intentional states does not depend on the environment in an important way.

In this essay I show that we have good reason to believe that content externalism is incompatible with a key aspect of self-knowledge: our ability to access our intentional contents in a first-person way. I also show that the arguments taken to support content externalism are not as strong as some have thought. Thus, we should adopt content internalism, because it is the only view of intentional content consistent with self-knowledge and because there are no good arguments that properly support of content externalism. What may have appeared to be a genuine philosophical puzzle—understanding how content externalism and self-knowledge are compatible—dissolves when we see that there are no good reasons to adopt content externalism.

CHAPTER 1 INTRODUCTION

A Sketch of the Problem

The debate over the compatibility of content externalism and self-knowledge has been around for at least 20 years. Content externalism began to gain acceptance with two papers that Tyler Burge published in the late seventies and early eighties, “Individualism and the Mental” (1979) and “Other Bodies” (1982). Not long after, papers began surfacing addressing the compatibility of content externalism and self-knowledge.¹ The debate has been lively ever since. The general worry has been that content externalism is incompatible with self-knowledge. It is not hard to see how these theses could seem to be incompatible. If, as the content externalist says, the content of our thoughts is shaped in a significant way by the environment, then there may be instances in which we must investigate aspects of the external environment to know what we think. Investigating the external environment in order to know our own minds seems contrary to the common-sense idea of self-knowledge: the idea that each of us knows our own mind best and in a way that no one else can. If I have to investigate the external environment to know what I think, I do not seem to know my mind in a way that you cannot; our ways of accessing my mind appear to be the same. As equitable as this may be, it is in clear conflict with our common-sense idea of self-knowledge. So, there seems to be a problem. The popular view of content externalism, which is supported by seemingly powerful philosophical arguments (such as those using Hilary Putnam’s Twin Earth thought experiment and the work of Tyler Burge), seems to be incompatible with a

¹ For examples see: (Davidson 1987), (Burge 1988) and (Boghossian 1989).

concept that plays a significant role in how we think of ourselves and our place in the world.

So far we have seen nothing strong enough to show us that we have a true philosophical puzzle; putative incompatibilities have turned out to be illusory. Many have found the arguments for content externalism convincing, and are not inclined to reject the view merely because it appears to be incompatible with self-knowledge. Maybe we simply need to revise our common-sense understanding of self-knowledge. Or, better yet, it may be possible to show that our common-sense concept of self-knowledge and content externalism are actually compatible, thus requiring no revision in our common-sense concept of self-knowledge. Either such maneuver is a form of compatibilism. Incompatibilism, then, is the view that the doctrines are incompatible. Someone may argue that content externalism forces us to reject the common-sense idea of self-knowledge, just as someone may argue that we must reject content externalism because it conflicts with self-knowledge. Either sort of incompatibilism can seem to give something up. Either we reject a common-sense concept that plays a foundational role in how we think of ourselves in relation to others and the external world or reject a doctrine that is supported by powerful philosophical arguments.

Aspects of the Cartesian Theory of Mind Worth Preserving

Sometimes incompatibilists who accept content externalism argue that living without our common-sense conception of self-knowledge requires releasing ourselves from the grip of a false picture of the mind, a picture that has been with us since at least Descartes. This view of the mind famously came under attack in Gilbert Ryle's *The Concept of Mind*, and since then has been seen by many to be a barrier to philosophical

progress. Here is Ryle's vivid description of the notion of self-knowledge that is part of Descartes's official theory of the mind as Ryle sees it.

What sort of knowledge can be secured of the workings of a mind? On the one side, according to the official theory [Descartes's view], a person has direct knowledge of the best imaginable kind of the workings of his own mind. Mental states and processes are (or are normally) conscious states and processes, and the consciousness which irradiates them can engender no illusions and leaves the door open for no doubts. A person's present thinkings, feelings and willings, his perceivings, rememberings and imaginings are intrinsically 'phosphorescent'; their existence and their nature are inevitably betrayed to their owner. The inner life is a stream of consciousness of such a sort that it would be absurd to suggest that the mind whose life is that stream might be unaware of what is passing down it.²

In contrast to this, Ryle notes that according to the official Cartesian doctrine our access to the minds of others is quite the opposite of the direct and unmediated access described above in such glowing terms.

On the other side, one person has no direct access of any sort to the events of the inner life of another. He cannot do better than make problematic inferences from the observed behavior of the other person's body to the states of mind which, by analogy from his own conduct, he supposes to be signalised by that behavior. Direct access to the workings of a mind is the privilege of that mind itself; in default of such privileged access, the workings of one mind are inevitably occult to everyone else.³

Descartes view, according to Ryle, seems to be that the perfect self-knowledge we have of our own minds is not to be had when it comes to our knowledge of the minds of others. Worse, it seems that on the Cartesian picture we not only lack perfect knowledge of other minds, but their intentional contents seem to be in principle opaque to us.

² (Ryle 1984), pages 13-14.

³ (Ryle 1984), page 14.

Ryle is surely right to critique such a view of self-knowledge and other-knowledge. As for self-knowledge, there are no doubt instances in which someone is mistaken about the intentional content of his own thoughts. Also, we should not limit the domain of the mental to merely what is occurrently conscious to a subject. I believe that Tallahassee is the capital of Florida even when I am not consciously entertaining the belief; intentional states such as beliefs are part of one's mind even when they are not conscious. As for other-knowledge, our access to the minds of others may not be direct, but it is certainly another step to say that the intentional contents of the minds of others are occult to everyone but their owners. There is no mystery in how we gain knowledge of the minds of others. We mark their words, listen to their speech, and observe their behavior. With enough information and some communication, we very frequently come to know the intentional contents of these minds. Such knowledge seems to be a necessary condition on communication.

We should follow Ryle when it comes to correcting for Cartesian excesses. However, there is some truth in the official Cartesian doctrine that is worth preserving. We should adopt the idea of the directness of self-knowledge. The key point here about directness is that we do not know the contents of our minds on the basis of any inference or evidence. In addition to self-knowledge, this essay concerns intentionality, so we shall focus on the intentional contents of minds, the contents of intentional states and events such as beliefs, desires, wishes, hopes, fears and the like. With regard to intentionality, another piece of the Cartesian picture worth preserving is Descartes's content internalism. Descartes's view of the independence of intentional content from the external environment comes across clearly in his *Meditations*. Descartes asks the

reader to imagine a scenario in which the nature of the world external to the thinker is radically different from what the thinker takes it to be. The scenario is one in which an individual is being deceived by an evil demon who is fabricating the thoughts of the thinker so as to give the appearance of mind-independent reality being as the thinker has always taken it to be, while in fact mind-independent reality is much different. Clearly, such a scenario requires that the contents of one's intentional states can remain the same even if the external environment undergoes radical change. If content externalism is true, then the situation is incoherent, because according to content externalism, the environment would have to be a certain way for the thoughts to be the way they are. Content externalism forecloses the possibility of having certain thoughts if the environment is not a certain way. To believe that water is vital for human life, for example, the content externalist holds that there must actually be water in the believer's environment. An evil demon could not fabricate the thought that water is vital to human life if the thinker was not actually related to water.

In sum, content internalism and direct self-knowledge are features of the Cartesian picture we must preserve in order to avoid revising our conception of self-knowledge, or so I shall argue in this essay. I should add that none of Descartes's substance dualism factors into my overall project, and it definitely does not need to be adopted for any of the forthcoming conclusions to go through; nor do we need to make any other metaphysical assumptions about what sorts of entities minds are. I intend for the major conclusions of this essay to be amenable to both physicalists and non-physicalists because of their lack of attendant metaphysical commitments.⁴

⁴ Regarding arguing against content externalism from a Cartesian framework, Katalin Farkas takes a similar, though much more developed, line in her recent book, *The Subject's Point of*

Why the Debate has seemed Intractable

It should surprise no one experienced in philosophy that the debate about the compatibility of content externalism and self-knowledge was not settled shortly after it began. There has been no shortage of papers on the question of the compatibility of self-knowledge and content externalism. The anthologies on the subject give a good indication of the perceived importance of this question. *Externalism and Self-Knowledge* (1998) and *Knowing our own Minds* (1998) were followed by *New Essays on Semantic Externalism and Self-Knowledge* (2003). More recently, there has been *The Externalist Challenge* (2004) and *Internalism and Externalism in Semantics and Epistemology* (2007), in addition to many monographs and articles on the subject. No philosophical consensus about the compatibility of these two theses emerges from this material. The debate has seemed to be intractable because of differing ideas of what is at stake, about what is incompatible with what.

The use of *a priori* knowledge as a proxy for self-knowledge by some philosophers and not others provides a particularly clear example of the differing ideas at stake in this debate. Michael McKinsey is one philosopher who has appealed to *a priori* knowledge in his characterization of what he calls "privileged access to content." Here is his definition of privileged access to intentional content.

Privileged access to content (PAC) Necessarily, for any person x , if x is thinking that p , then x can in principle know *a priori* that he himself or she herself is thinking that p .⁵

View, wherein she defends a Cartesian picture of the mind and uses it to argue against content externalism.

⁵ (McKinsey 2003), page 97.

McKinsey says, “By ‘a priori knowledge’, I mean knowledge that can be achieved just by thinking without perceptual observation or empirical investigation and without having to make any empirical assumptions.”⁶ McKinsey is of course not the only philosopher to have defined self-knowledge in terms of *a priori* knowledge (see the Boghossian quote at the top of Page 17 of this essay). I think we should be wary of such definitions, for they attempt to reduce (or equate) a type of knowledge we already have a good intuitive grasp on (self-knowledge) to a type of knowledge on which we have less of a grasp (*a priori* knowledge). Farkas (2008b) has argued, following Nuccetelli (1999), that we do not know our own minds by way of *a priori* knowledge. She writes,

A priori knowledge (that is, the kind of knowledge we have of logic, maths, and conceptual truths) is traditionally regarded as knowledge attained by the use of reason alone, and this description does not seem to apply to knowledge of our mental states (cf. Nuccetelli 1999) When I register that I feel a slight pain in my knee, the faculty I am using is different from the one used in establishing the correctness of *modus ponens*. One difference between introspection and a priori knowledge is precisely that introspection provides special access to its subject matter, while a priori does not.⁷

I agree with Farkas on this point. The asymmetry of access is the key distinction between self-knowledge and other-knowledge; one seems to have self-knowledge of those intentional contents to which one has a type of access that is lacked by all other individuals. Introducing *a priori* knowledge into the mix merely obscures what is most important.

In addition to differing opinions on whether self-knowledge is a form of *a priori* knowledge, differing epistemic phenomena have been set out as putatively incompatible with content externalism. “First-person access,” “first-person authority,” “self-

⁶ (McKinsey 2003), page 97.

⁷ (Farkas 2008b), page 25.

knowledge,” “privileged access” and “privileged self-knowledge” have all been said to be incompatible with content externalism. We cannot make progress on this debate until we first get a clear understanding of what phenomena these names pick out, and, second, get a clear idea of how these phenomena are related. It is only after we see how the epistemic notions of self-knowledge are connected that we will be in a position to assess whether self-knowledge is compatible or incompatible with content externalism. Testing the compatibility of content externalism against just one of the above notions in isolation from the others is not enough.⁸

Because it will hopefully be helpful to get a rough idea of what I mean by "self-knowledge" and its attendant notions, I shall now sketch the picture of self-knowledge I lay out in more detail in Chapter 2. In the picture I defend, the connection between self-knowledge and rationality factors prominently. In my view, it is not possible to fully understand self-knowledge until one sees its connection to rationality. In this essay, I argue that rational agents like us must have direct access to the intentional contents of their minds, access that cannot be mediated by internal or external factors. Another way to put the point of immediate access to intentional content is to draw on the notion of intentional content being transparent to a thinker. Paul Boghossian uses the idea of mental contents being transparent in “The Transparency of Mental Content,” and also points out that the transparency of mental content is presupposed by our conception of rationality. In this paper, he articulates the thesis of the transparency of mental content as follows.

⁸ Donald Davidson, for example, makes this mistake of testing the compatibility of content externalism with just one of the notions of self-knowledge, first-person authority (see (Davidson 1987)).

The thesis of the epistemic transparency of content may be usefully broken into two parts: (a) If two of a thinker's token thoughts possess the same content, then the thinker must be able to know a priori that they do; and (b) If two of a thinker's token thoughts possess distinct contents, then the thinker must be able to know a priori that they do. Call the first the thesis of the *transparency of sameness* and the second the thesis of the *transparency of difference*.⁹

Having just pointed out that it is best not to cast self-knowledge as *a priori* knowledge, the point obviously needs refashioning. Boghossian's point is that intentional content is transparent to a thinker just in case knowledge of that content does not require knowledge of external factors, "without," as he says, "the benefit of further empirical factors."¹⁰ We can understand transparency without understanding it in terms of *a priori* knowledge of intentional contents; we should understand it as direct access to content that is not mediated by other factors.

Having granted the transparency of intentional content, or direct and unmediated access to intentional content that each of us has to our own intentional content, the next step is to note that the way others access our intentional content is quite different. There is a difference in how an agent accesses his intentional content and how others access that content. The distinction can be stated as a distinction between first-person access and third-person access. First-person access to one's own intentional content is not mediated by knowledge of some other external or internal factor; the access is direct and unmediated. Third-person access to another's intentional content is always mediated by some other factor and thereby not direct. I cannot access your thoughts

⁹ (Boghossian 2008), page 162.

¹⁰ (Boghossian 2008), page 159.

without observational data or contextual information. I use this information to make inferences to the nature of your intentional content.

Related to the epistemic notions of the transparency of intentional content and first-person access is another notion that is central to an understanding of self-knowledge: first-person authority. We will see that understanding first-person authority requires first understanding not only the asymmetry of access to content seen in the difference in first-person and third-person access, but also understanding how this asymmetry plays out in the context of communication. It is important to see that the concept of first-person authority only has application in the context of communication. Understanding the connection between first-person access to intentional content and first-person authority, and how these notions are related to rationality, will help us to see how the various notions of self-knowledge are related. In short, we would not be rational if our intentional contents were not transparent to us. Furthermore, it is hard to see how our reports of our intentional contents could have authority over others' reports of those contents if we did not bear some epistemic advantage over others when it comes to our own intentional contents; in other words, first-person access to intentional content helps explain why our reports of our own intentional states have the first-person authority they do. Access and authority are distinct but closely related notions. Without such authority and transparency, it is impossible to conceptualize ourselves as rational agents.

A lack of clarity about self-knowledge is not the only source of the debate's difficulty. Just as philosophers have understood self-knowledge differently, so have understandings of content externalism differed. In a forthcoming paper "Mental Content: Externalism and Internalism," Brie Gertler goes so far as to argue that there is in fact so

much ambiguity in the use of the terms "internalism" and "externalism" that, for clarity's sake, these labels should be abandoned all together. While this seems too rash to me, it is not hard to see how one could be driven to such a view; there are many different articulations of externalism. (Indeed, as we will see, there are genuinely different types of externalism.) It does not help matters that the debate has been put variously as one between notions of wide content and narrow content, between individualism and anti-individualism, and between content externalism and content internalism. At points, one finds philosophers talking as if the real issue is whether or not thoughts are literally in the head, no doubt an offshoot of Putnam's famous declaration that "meanings ain't in the head." As we will see, the debate is not properly understood as being about the spatial location of thoughts, but about the logical dependence of thoughts on the environment. I shall understand internalism and externalism about intentional content as theses about whether or not thoughts supervene on ways an individual is independent of his external environment.¹¹ I articulate the theses using the notions of intrinsic and extrinsic properties. The properties of an individual that are what they are independent of his external environment are an individual's intrinsic properties. One's mass for example is an intrinsic property. One's weight is not; one's weight depends on the environment one is in. My mass is the same on Earth and the Moon, though my respective weights in these two environments differ. Weight, therefore, is an extrinsic property of an individual: it is a way an individual is that is not independent of the

¹¹ Supervenience is a logical notion having to do with a set of properties depending on another. If Xs depend on Ys, then Xs supervene on Ys. Here is a simple example: facts about who wins a soccer match supervene on facts about the number of goals scored by each side. Sometimes talk of supervenience is accompanied by talk of reduction (as in debates about the metaphysics of mind). Content externalism and internalism, however, are not theses that involve talk of reducing one type of entity or fact to another type of entity or fact.

external environment. So in terms of whether or not intentional content depends on the external environment, the content internalist says that thoughts are like mass, whereas the content externalist says they are like weight.

In sum, part of the reason that the debate has seemed to be intractable is because of the many theses in circulation. Once we are clear about what is at stake, we can see that content externalism is incompatible with the first-person access each of us has to only our own intentional content. Once we see this incompatibility, we can see that content externalism is incompatible with the phenomena that depend on such first-person access: first-person authority and rationality.

The Basic Program of the Essay

In this essay I defend incompatibilism. In Chapter 2, I offer a picture of self-knowledge and its attendant epistemic notions, showing their interrelations and showing that we must have first-person access to our intentional contents if we are rational. Having just sketched this picture, I will move on to the other chapters.

In Chapter 3, I outline various forms of externalism about mental phenomena. There are many things one can be externalist about: intentional content (content externalism), the background conditions necessary for thought (transcendental externalism), phenomenal properties (phenomenal externalism), and linguistic meaning (semantic externalism). In Chapter 3, I sketch arguments for content externalism and transcendental externalism, as these forms of externalism are most important to our inquiry. Arguments for content externalism have been motivated, respectively, by reflecting on Twin Earth, by looking at the semantic properties of the ascriptions of

intentional content attributions,¹² and by reflecting on the nature of singular thoughts (thoughts which are uniquely about an object). Many philosophers have come to accept content externalism on the basis of some or all of these arguments. I do my best to articulate why these arguments have become so widely accepted.

In Chapter 4, I give my argument for incompatibilism. I argue that content externalism, the view that the contents of intentional states are not merely a function of the intrinsic properties of their owners, is incompatible with first-person access to intentional content. Because first-person access to intentional content is required for rationality, content externalism is not the right view of the intentional content of rational agents such as ourselves. We need not reject all types of externalism, though. For example, first-person access to intentional content is compatible with transcendental externalism. Because transcendental externalism is compatible with first-person access to intentional content, we do not need to follow Descartes in thinking that an isolated thinker with intentionality is a true possibility.¹³ Indeed, if transcendental externalism is true, an isolated thinker like the individual in Descartes's evil demon scenario is not possible given the background conditions necessary for thought. I also note that first-person access to intentional content is compatible with semantic externalism, the view that meaning is not a function of any individual's intrinsic properties.

¹² Throughout this essay, I follow Burge in understanding an ascription to be the linguistic device that attributes intentional content to a thinker. The attribution is the act of attributing the content. He introduces this useful distinction in "Individualism and the Mental." He writes, "I speak of *attributing* an attitude, content, or notion and of *ascribing* a that-clause or other piece of language. Ascriptions are the linguistic analogs of attributions. This use of 'ascribe' is nonstandard, but convenient and easily assimilated" (Burge 1979), page 24.

¹³ Of course, even in the scenario with the evil demon fabricating the individual's thoughts, the individual is not isolated due to the evil demon's presence.

Even though some forms of externalism are compatible with first-person access to intentional content, some may still feel compelled to be externalists about intentional content based on the perceived strength of the arguments for this view. I have no doubt that someone of a thoroughgoing externalist bent may be uneasy about giving up content externalism after reading Chapter 4. The goal of Chapter 5, then, is to ease such a concern, by showing that the arguments for content externalism are not as strong as some have thought.

In Chapter 6 I discuss common objections to content internalism. These objections spring from three main types of concerns: linguistic, epistemic, and metaphysical. One general linguistic concern is that adopting content internalism requires adopting semantic internalism, the view that linguistic meaning is a function of an individual's intrinsic properties. Speaker meaning—the meaning that a speaker intends to convey by his use of a linguistic expression—may indeed be a function of an individual's intrinsic properties (assuming that intentions supervene on intrinsic properties); linguistic meaning, however, is what linguistic expressions mean independent of anyone's intentions. We have good reason to reject semantic internalism, for it is a thesis that reduces linguistic meaning to speaker meaning. Some have thought that a content internalist must adopt semantic internalism because the description of narrow intentional states requires a special language, one whose meanings supervene solely on an individual's intrinsic properties. Such a private language is hard—if not impossible—to conceptualize. We do not need to worry, however, about developing or trying to conceptualize a special language for describing narrow intentional states. Content internalism is perfectly compatible with semantic

externalism. There are those, though, who have taken content externalism and semantic externalism to be synonymous, and thus do not think of content externalism and semantic externalism as distinct doctrines.¹⁴ We should be wary of such a conflation, because these views concern different subject matters. Content internalism is a thesis about intentionality, and semantic externalism is a thesis about linguistic meaning.¹⁵ Whatever view one takes of the relationship between intentionality and linguistic meaning, it is surely undeniable that they are different phenomena. Even though there is much to be said of their relationship that is of great philosophical importance, we should not lose sight of the fact that these theses concern different subject matters. Commonplace examples make this point. We are all familiar with times when we self-ascribe an intentional state to ourselves and misspeak in doing so. Suppose that I enter a discussion on the relative personality characteristics of a family. And suppose that I find such interpersonal comparisons distasteful. If in the course of trying to show my disapproval of such interpersonal comparisons, I utter, “Comparisons are odorous,” I have misspoken in a way that can nevertheless reveal the content of my thought—that comparisons are odious—even if *that content* is mischaracterized by the linguistic (or semantic) content of the expression (“comparisons are odorous”) initially

¹⁴ Sanford Goldberg (2007) offers such an example in a recent paper, “Semantic Externalism, Epistemic Illusions,” where he uses Burge words to define semantic externalism .

Let Semantic Externalism (SE) be the thesis that “the mental natures of many of an individual’s mental states and events are dependent for their individuation on the individual’s social and physical environments” (Burge 1986b), page 697.

Such a definition leaves no room to distinguish between semantic externalism and content externalism.

¹⁵ More specifically, content internalism is a thesis about the degree to which intentional content depends on the external environment.

used to make the thought public. Such instances should be kept in mind, for they show that there is a need for a distinction between intentional content and semantic content. It is a simple point, but one that is sometimes easily forgotten.

In addition to these linguistic worries, I address some epistemic worries about content internalism. One epistemic worry is that the content internalist cannot respond properly to skepticism about the external world. I discuss this worry, and note that the content internalist has plenty of resources available to answer the skeptic. Furthermore, content externalism seems faced with its own set of epistemic worries. One such worry is the main focus of the essay: its incompatibility with first-person access to intentional content and hence rationality. The other is that content externalism seems to license the conclusion that we have *a priori* access to the external world.¹⁶ I conclude that the epistemic worries about content internalism should not concern us when seen in light of the serious epistemic concerns raised by content externalism.

In Chapter 6 I also address metaphysical concerns about adopting content internalism. Given the widespread interest in the metaphysics of mind, I rather briefly address how physicalism relates to content internalism and content externalism. I note that there appear to be some forms of physicalism that are incompatible with content externalism, though there are others that appear compatible.

Content internalism holds that a thinker's intrinsic properties are responsible for his intentional contents. In Chapter 6, I survey some candidate intrinsic properties on which intentional content might depend. It seems to me that much more work needs to be done here by philosophers to help us understand how intentional content could depend

¹⁶ See (Boghossian 1998) and (McKinsey 1991).

on intrinsic properties. There has been renewed interest in the connection between phenomenology and intentionality. If phenomenal properties are intrinsic, they may be a candidate property on which intentional content supervenes. I discuss this possible supervenience relation as well as the possibility that intentional content supervenes on intrinsic properties of brains. Though I do not ultimately endorse any of the candidate properties, I offer some reasons to be skeptical of the idea that all intentional content supervenes of phenomenal properties.

CHAPTER 2 SELF-KNOWLEDGE OF INTENTIONAL CONTENT

In this chapter and the rest of the essay, the focus shall be on knowledge of and access to the contents of intentional states, or, as they are sometimes called, the contents of propositional attitudes. Believing, desiring, wishing, hoping, intending and remembering are all examples of intentional states (or propositional attitudes) that have content. *That Tallahassee is the capital of Florida* is the content of my belief that Tallahassee is the capital of Florida. I will often refer to these contents as “intentional contents” or “the contents of intentional states.” I do not discuss our knowledge of other mental phenomena, such as sensations. We may indeed have special access to our own sensations or know them in a way that others cannot, but that is a topic for another time.

The overall goal of this chapter is twofold. The first goal is to sketch the logical relations between various notions of self-knowledge of intentional content: the transparency of intentional content, the asymmetry of access to intentional content, first-person access to intentional content and first-person authority with regard to intentional content. I take the most fundamental notion, the transparency of intentional content, and explain the other notions in relation to it. From the transparency of intentional content and the fact that a subject's access to his intentional content is different from the access of all others, first-person access to intentional content follows. The asymmetry of epistemic access expressed by the thesis of first-person access to intentional content is mirrored by an asymmetry in authority regarding intentional contents in the context of communication. At the end of the chapter I suggest that seeing how all of these notions are related forms part of an explanation of self-knowledge. I do not pretend that all the

features of the picture of self-knowledge on offer are very novel. In fact, not all should be; we should all have some vague idea about self-knowledge and related concepts before any philosophy.

Notions of Self-Knowledge

The Transparency of Intentional Content

I briefly discussed the idea of intentional content being transparent to its owner in Chapter 1, and suggested that this talk of intentional content being transparent captures the fact that one accesses one's thoughts directly, unmediated by evidence or inference. Because of its various ordinary language connotations, the use of the term "transparency" is supposed to make vivid this direct and unmediated access. We talk of objects being transparent and opaque in order to capture whether or not the object can be seen through. Many prefer transparent ocean water to the opaque version, for such water allows objects in it to be seen clearly. We talk of someone being transparent in his dealings, and in doing so convey the idea that he is not withholding important information or trying to dissemble but rather being open and honest. These related senses of "transparency" both convey the idea of direct access. When wading in transparent ocean water, we appear to directly perceive the objects underneath the water's surface. In our dealings with someone who is being transparent, we do not have to spend time puzzling over what he is really up to, wondering about his true motivation, or worrying how to interpret what he says in light of some suspect motivations he may harbor; rather, we take his words at face value, and more directly access the content of the information he intends to convey.

These ordinary language senses of "transparency" do not fully capture what we are after when we talk about the transparency of content to a thinker. After all, it is not

as if we see our thoughts, even in some metaphorical sense of seeing them with our mind's eye. And it is not the case that our thoughts are transparent to us because they are not withholding information from us, as in the case of someone who is being transparent, for this suggests that our thoughts have a life of their own, as it were, and that they can deal with us in a transparent or opaque fashion. To get closer to a philosophical understanding of transparency, let us revisit Boghossian's articulation of the notion of the transparency of intentional content.

The thesis of the epistemic transparency of content may be usefully broken into two parts: (a) If two of a thinker's token thoughts possess the same content, then the thinker must be able to know a priori that they do; and (b) If two of a thinker's token thoughts possess distinct contents, then the thinker must be able to know a priori that they do. Call the first the thesis of the *transparency of sameness* and the second the thesis of the *transparency of difference*.¹

We noted that there are good reasons to not use the notion of *a priori* knowledge as a proxy for notions of self-knowledge, so we need to translate Boghossian's articulation into different terms. I suggested in the introduction that we understand transparency in terms of access that is not dependent on knowledge of external factors. With this idea, the thesis of the transparency of intentional content can be thusly understood.

The thesis of the transparency of intentional content: Each thinker accesses the content of his intentional states directly and without the benefit of evidence.

From this basic thesis we can extrapolate the theses of the transparency of sameness and the transparency of difference that Boghossian mentions. We can understand the transparency of sameness as follows. If two token thoughts of a thinker have the same content, then the thinker must know so immediately and without the benefit of external factors. And we can understand the transparency of difference as follows. If two token

¹ (Boghossian 2008), page 162.

thoughts of a thinker have different content, then the thinker must know so immediately and without the benefit of external factors.

I should note that my own articulation is different from Boghossian's in an important way: it builds immediacy into the definition of transparency. I do not think that Boghossian would object to this, based on concerns he expresses in other places about knowing our thoughts on the basis of some type of internal evidence.² In Boghossian's articulation above, there is room for understanding self-knowledge in a way that employs a robust notion of introspection, wherein each of us knows the content of our thoughts based on evidence internal to the mind. There should be no worry about building immediacy into our definition. We shall hear more about this in the section called "The Transparency Condition," but I shall try to make an intuitive case for it at present. While we have surely all experience moments of introspection where we try to decide what we believe or how we are to act in a particular situation, we arrive at decisions to questions about what to believe and how to act on the basis of immediate access to the contents of other intentional states. It is not as if we grasp the content of intentional states based on some type of mental evidence that is internal to our minds. As evocative as talk of the mind's eye perceive its objects is, there is certainly nothing based on the phenomenology of accessing intentional content that lends support to the idea that this content is accessed by way of some type of internal evidence.

The Asymmetry of Access

We do not have direct access to the minds of others. The contents of the minds of others are accessible by us, of course, but they are not transparent to us. We need to

² See (Boghossian 2008), pages 143-145.

use evidence and make inferences to know the minds of others. We do not need to do this in our own case. Based on these facts, we can infer that there is an asymmetry of access to intentional content, which the following thesis is intended to capture.

The thesis of asymmetry of access to intentional content: One has a type of access to the contents of one's intentional states that is not had by others.

So far this says nothing about whether or not this access is privileged and should not be controversial. Even so, not everyone agrees that the access is asymmetrical. Gilbert Ryle, for one, argues that access to intentional content is symmetrical. If Ryle is right, I know my own mind in the way I know the minds of others: by considering behavioral evidence and knowledge of other relevant contextual factors. It is easy to see how this is complimentary to Ryle's behaviorism, but hard to see how it could be true.

If we grant the truth of the thesis of asymmetry of access to intentional content, then we can ask what is different about these two sorts of access. It need not follow that one is privileged and the other not. In "Transparency, Belief, and Intention," Alex Byrne acknowledges this when he draws a distinction between privileged and peculiar access.

Two features of self-knowledge make it of particular interest. The first is that, by and large, beliefs about one's mental states are more likely to amount to knowledge than one's corresponding beliefs about others' mental states—one has *privileged access* to one's mental states. The second is that one has a way of knowing about one's mental states that one cannot use to come to know about the mental states of others—one has *peculiar access* to one's mental states. The two features are independent, in the sense that neither entails the other. But they are connected: the kind of peculiar access that we enjoy presumably explains why we have privileged access. A satisfying theory of self-knowledge will illuminate this connection.³

³ (Byrne 2011), page 202.

It seems to me that the reason that first-person access to intentional content is peculiar is that each individual has a unique relationship to his own mental states. The uniqueness of the relationship is grounded in the fact that accessing them does not require any evidence. Each person is presented with the content of his intentional states in a direct and immediate way; each person's contents are transparent to him. However, third-person access always requires evidence and is thereby indirect and not immediate. Lack of evidence, then, may be what makes peculiar access privileged access.

It may sound odd to say that a subject's access to his thoughts is privileged because it does not require evidence and inference. Ordinarily, the person with more evidence is in the privileged epistemic position. This is not the case, however, for self-knowledge, and there is no paradox here if we reject the assumption that all domains of knowledge are domains that require evidence. We have very good reason to reject this assumption in the case of self-knowledge. To know the intentional contents of your mind, I must first use evidence and then make an inference using that evidence. To know the contents of my own mind, I typically do not need to do any of these things. Is this difference in the need for evidence what makes one's access to one's intentional contents privileged? Before we answer this question, we should discuss what has been called the transparency condition; the discussion will help us see why self-knowledge is not a domain that requires evidence. Also, understanding this condition helps us to see why peculiar access is privileged access.

The Transparency Condition

In discussing the transparency condition in *Authority and Estrangement*, Richard Moran quotes Roy Edgley.

My own present thinking, in contrast to the thinking of others, is transparent in the sense that I cannot distinguish the question "Do I think that P?" from a question in which there is no essential reference to myself or my belief, namely "Is it the case that P?" This does not of course mean that the correct answers to these two questions must be the same; only I cannot distinguish them, for in giving my answer to the question "Do I think that P?" I also give my answer, more or less tentative, to the question "Is it the case that P?"⁴

Moran also quotes a famous passage from Gareth Evans.

In making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?"⁵

Moran thinks that these writers are tracking a notion of transparency that "concerns a claim about how a set of questions is to be answered, what sorts of reasons are to be taken as relevant."⁶ The relevant claim, for Moran, is "that a first-person present-tense question about one's beliefs is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world."⁷ Questions about our beliefs are transparent to the reasons or evidence that we turn to in order to justify our beliefs about the external world and not evidence that we introspect.

One way to understand this talk of transparency is in terms of questions a subject can ask himself about the content of his intentional states. When someone asks himself "Do I believe p?" it is equivalent to asking himself "Is p true?" in the sense that the once the subject has determined whether p is true, he has answered the question of whether

⁴ (Edgley 1969), page 90.

⁵ (Evans 1982), page 225.

⁶ (Moran 2001), page 62.

⁷ (Moran 2001), page 62.

he believes p. This should really be no surprise since to believe p is to hold p true. Deciding whether one believes p and whether p is true is making a decision about the same thing.

What seems interesting about the transparency condition, as understood by Moran, is not the claim that the question about whether one believes p is equivalent to the question of whether one holds p true, for we already know that the questions about what one believes and what one holds true are synonymous due to the very meaning of "to believe" and "to hold true." What is interesting is Moran's thought that the transparency condition must be understood in terms of the evidence that a subject uses to answer questions about his belief that p and about whether p is true. This is what Moran presumably has in mind when he notes, "that a first-person present-tense question about one's beliefs is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world." When one is asked whether one believes that there will be a third world war, one does not turn inward to some internal evidence. Rather, one turns to the evidence that one would use to answer the question of whether there will be a third world war. To a subject, the question "Do you believe that there will be a third world war?" is transparent to the question "Will there be a third world war?" in the sense that all the relevant evidence for answering the latter is all the evidence one has for answering the former. This alone should show us that theories of introspection that have a subject relying on internal data to know his thoughts are mistaken. Moran's view is set against views that construe introspection as a perception-like faculty whereby the mind's eye examines its contents in a fashion similar to the way the body's eyes perceive the external world. The

transparency condition states that no such internal data is needed to know one's intentional contents. Whether one believes *p* is answered by the evidence and reasons for *p* itself and not some other evidence that only the thinker has access to.

So, Moran suggestion is that we understand the transparency condition as one about evidence (or reasons) given for different questions. Moran thinks that the same reasons can answer the question “Do you believe *p*?” and “Is *p* true?” This seems true for certain propositions, for example, propositions that concern the external world, such as the proposition that there will be a third world war. However, is the transparency condition, as Moran understands it, true of second-order intentional states, intentional states about other intentional states?

As Moran has articulated it, the transparency condition appears not to be true of second-order intentional states and the first-order intentional states they are of (even if it appears true of first-order intentional states and the facts about the world the first-order states are of). The reason for this was given in the section on asymmetry of access: we do not use evidence to come to know the contents of our intentional states. If it were true that the evidence and reasons I use to arrive at my first-order belief that there will be a third world war are the same as the evidence and reasons I use to arrive at my second-order belief that I believe that I believe that there will be a third world war, then my second-order beliefs would depend on evidence.⁸ However, when we imagine the types of reasons that one could give for one's second-order beliefs, the reasons that support the first-order belief are *not* the likely reasons. Indeed, there do not appear to

⁸ This point can be made about previously held beliefs in addition to beliefs one arrives at. The evidence or reasons used to arrive at a belief become the evidence and reasons a subject typically uses to support that belief.

be any reasons we can cite to support our first-order beliefs. Consider the following dialogue to better see the point.

A: Do you believe that there will be a third world war?

B: Yes, I do.

A: Why do you believe that?

B: Well, it seems that many countries in the Middle East and North Africa are experiencing transitions of power from autocratic regimes to undetermined forms of government. There is a good chance for great instability in the region. With a lack of such stability, it seems that there is a good chance for a third world war.

A: That's interesting. Let me ask you a slightly different question. Do you believe that you believe that there will be a third world war?

B: I guess so. I know that I believe that there will be a third world war.

A: Yes, but do you believe that you believe there will be a third world war?

B: I guess I do, but I hadn't really thought about it. That's an odd question.

A: Why do you believe that you believe there will be a third world war?

B: Not sure what to tell you, other than what I have already said about why I believe there will be a third world war.

Even though artificial, the dialogue shows that we are not in a position to give evidence or reasons for our second-order beliefs. The reason for this is that there is no such evidence.⁹ The best one can do, in a conversational context, when asked for evidence

⁹ Burge makes a nice point that the warrant for our second-order beliefs differs from justification as it is typically understood. Of our second-order beliefs about first-order intentional states, he writes,

An individual's epistemic warrant may consist in justification that the individual has a belief or other epistemic act or state. But it may also be an entitlement that consists in a status of operating in an appropriate way in accord with norms of reason, even when these norms cannot be articulated by the individual who has that status. We have an entitlement to certain perceptual beliefs or to certain logical inferences even though we may lack reasons or justifications for them. The entitlement could in principle presumably—though often only with extreme philosophical difficulty—be articulated by

or reasons for one's second-order beliefs is to cite the evidence and reasons for the first-order belief that is the object of the second-order belief. It is important to see that the support for one's first-order beliefs does not provide support for one's second-order beliefs, though. Citing the support for the first-order beliefs is merely the best one can provide *pragmatically* when asked to justify one's second-order belief, but we should be clear that this justification does not support the second-order beliefs.

If we understand the transparency condition as committed to Moran's idea that a "question about one's beliefs is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world," then the transparency condition does not hold between second-order intentional states and the first-order intentional states they are about. There are of course other possible articulations of the transparency condition. Even if the evidence and reasons for the first-level belief cannot be thought to count as evidence and reasons for the second-level belief, there is still something interesting about the relationship between the questions that can be asked about the first-order and second-order intentional states.

I think we can extend the transparency condition, properly understood, to first-order and second-order questions. Baron Reed offers an articulation that can serve this purpose.

Let us say that one question, *A*, is transparent to another, *B*, when the answer to *B* determines the answer to *A*. In a derivative way, the judgment that constitutes the answer to *A* is transparent to the judgment that constitutes the answer to *B*. In the case of self-knowledge, then, the

someone. But this articulation need not be part of the repertoire of the individual that has the entitlement. (Burge 1996), pages 241-242.

This quote clearly endorses the idea that the warrant for self-knowledge need not be something to which the subject has access.

transparency condition is met when the answer to the question *whether to believe that p* is determined by the answer to the question *whether p*.¹⁰

This articulation does not require that the evidence or reasons given by the subject to support his first-order intentional states also support his second-order states. It therefore gives us a way to understand an important feature of questions about one's intentional states that is not limited in the way that Moran's articulation of the transparency condition is. Second-order questions (e.g., "Do you believe that you believe there will be a third world war?") are transparent to first-order questions (e.g., "Do you believe that there will be a third world war?") in the sense that the answer to the first-order question determines the answer to the second-order question. The first-order and second-order questions do not violate the transparency condition so understood because there is no additional evidence that one uses to move from belief that *p* to belief that one believes that *p*. One does not make a conscious inference here. In this sense of transparency, then, we can say that first-order and second-order questions, corresponding to first-order and second-order intentional states, are transparent in the sense that the answer to the first-order question determines the answer to the second-order question. The transparency condition, then, is something that can hold between both questions about second-order intentional states and first-order intentional states and questions about first-order intentional states and corresponding facts about the world. The condition holds in so far as questions of a given level about the content of the states or the facts themselves are answered by questions of another level.

This transparency of questions (at least of first-order and second-order questions) is related to another type of transparency that we mentioned earlier, the

¹⁰ (Reed 2010), page 169.

transparency of intentional content. The intentional contents of one's mind are transparent to one in the sense that they do not depend on any evidence. The reason that there is a transparency between questions, for example, the reason that the answer to first-order questions determines the answer to second-order questions, is that there is no independent evidence required to answer the second-order question. The intentional contents of one's mind are transparent to one precisely because there is no evidence in addition to whether p needed to determine whether one believes p , and there is no evidence in addition to whether one believes p to determine whether one believes that one believes p .

First-Person Access to Intentional Content

In "Transparency, Belief, Intention," Alex Byrne offers an account of self-knowledge, which he calls "the transparency account," that he thinks explains both peculiar and privileged access. Seeing why his account fails will help us understand the proper role the transparency condition plays in understanding privileged access. Using the work of Andre Gallois, Byrne explains that following the doxastic schema is a reliable guide to self-knowledge. The doxastic schema has the following form.

$$\frac{p}{\text{I believe that } p}$$

Byrne calls moving from p (where ' p ' stands for any proposition) to the conclusion that I believe that p "the transparency inference." The idea is that one does not need to inspect anything internal to know what one believes. Rather, in the fashion captured by Evan's quote, one looks outward to the world and then comes to know what one believes. Byrne rehearses some worries about the strength of this inference, and ultimately concludes that the inference is strongly self-verifying. That it is strongly self-

verifying explains privileged access. Byrne is surely right that no one other than the subject referred to by “I” in the schema can make the inference from p to the conclusion about the subject’s beliefs. Suppose Byrne thinks it over and decides that there will be a third world war. From this he can infer, using the doxastic schema, that he believes that there will be a third world war. It is clear that I cannot perform similar inference about his beliefs. I cannot note the fact that there will be a third world war and then infer that because there will be a third world war, Byrne believes there will be a third world war. No one else but the subject can make the inference; hence the subject is in a privileged position. Peculiar access is explained because the method expressed by the doxastic schema, in Byrne’s words, “only works in one’s own case.”¹¹

Byrne notes that the transparency inference is an inference from world to mind. It should strike us as odd that we must first know something about the world to know something about our minds. Surely there are questions such as the question of whether or not there will be a third world war that we do not already have an answer to, but require decision on our part. However, for many beliefs, it seems odd to say that our knowledge of them rests on an inference from a fact about the external world. Consider my belief that Tallahassee is the capital of Florida. It certainly seems false to say that I infer my having this belief from the fact that Tallahassee is the capital of Florida. It seems that my knowledge of this belief does not rest on any inference at all.

The problem with Byrne’s proposal is, in short, his focus on inference. The point about transparency is best expressed by focusing on questions being transparent to each other in the sense described earlier. It is not that I infer that I believe that p from p;

¹¹ (Byrne 2011), page 207.

in reflecting on the question of whether or not I believe p , there is nothing other than the truth of p that I need to decide. If I have already made up my mind, then reflecting on the truth or falsity of p is all I need to reflect on. Reflecting on it, either deciding what I believe or recalling my belief, is not to infer it from a fact about the world. This problem with Byrne's approach is summed up well by Matthew Boyle in "Transparent Self-Knowledge."

Instead of thinking of the subject as making an inference from P to I believe P , he can think of the subject as taking a different sort of step, from believing P to reflectively judging (i.e. consciously thinking to himself): I believe P . The step, in other words, will not be an inferential transition between contents, but a coming to explicit acknowledgement of a condition of which one is already tacitly aware.¹²

Boyle's idea is in line with the understanding of the transparency condition for which I have been advocating. Reflecting on the questions about first-order and second-order intentional states shows us that answering questions about first-order states is answering questions about second-order states. The point of transparency is that there is nothing epistemic mediating a subject's knowledge of his first-order states. If one holds the transparency condition, and rejects Byrne's account of inference, then one can conclude that once a subject reflects on p , where p is a proposition about the world, this reflection is all that is needed to know that the subject both believes p and believes that he believes p . All these ascending states are had for free, as it were, because the questions that can be asked about the states related to p meet the transparency condition.

Contrary to Byrne, then, our access to the contents of our own minds is privileged not because we are in a position to make an inference no one else can make,

¹² (Boyle 2011), page 227.

but because we do not need to make an inference that everyone else must make. Each person is in a position to access non-inferentially and without the benefit of evidence what others can only access on the basis of inference and evidence.

What the transparency condition shows us is that reflecting on whether *p* gives a subject immediate access to whether he believes *p*, just as reflecting on whether he believes that *p* gives him immediate access to the question of whether he believes that he believes that *p*. Consider questions we can ask about whether I believe that there will be a third world war.

(a) Will there be a third world war?

(b) Do I believe that there will be a third world war?

(c) Does Casey believe that there will be a third world war?

When I want to know the answer to (b) all I need to do is answer (a). When someone else wants to know the answer to (c), he must do more than answer (a). His questions about the content of my intentional states are not transparent to questions such as (a). To answer (c) he must use evidence and make inferences.

Reflecting on the transparency condition puts us in a position to state the thesis of first-person access to intentional content.

The thesis of first-person access to intentional content: Each subject has first-person access to his own thoughts because each subject accesses his contents non-inferentially and without the benefit of evidence, whereas others access those contents by way of inference.

The motivation for this thesis begins with recognition of the truth of the thesis of the transparency of intentional content and the truth of the thesis of asymmetry of access to intentional content. As Byrne's work shows, it can seem that our intentional contents are known on the basis of inference due to the fact that reflecting on *p* is all that is needed

to determine that one believes p . However, it is a mistake to think that we infer the contents of our beliefs from facts about the world. We typically go about our affairs in a first-order mode. When we have cause for reflection, as we do now, we see that there is nothing more that we need to decide other than p to know whether or not we believe p , whether 'p' is filled in by a fact about the world or a fact about our minds. Once we see this, we can see that we have first-person access to our intentional contents, for we access them non-inferentially and without evidence, whereas others cannot likewise access them.

First-Person Authority

The idea of first-person access is closely tied to idea of first-person authority, so much so that these respective terms are sometimes used interchangeably. One can see why this is so. If our access to our own intentional content were of a privileged, first-person sort, then it would seem to follow that we are in a position of authority with regard to this content. But what does this notion of first-person authority really come to? What does it add to the foregoing theses of self-knowledge? To properly situate the concept of first-person authority alongside these notions, we must examine how the idea of authority functions in an epistemic context. We will see that the concept of first-person authority has application only in the context of communication.

We are all quite familiar with the idea of someone being an expert or authority on a given domain. Jeffrey Sachs, for example, is an authority on international development. He has contributed to the scholarly debate and possesses a great deal of theoretical and practical knowledge about many aspects of international development. Imagine that he were on a panel with lay people concerning international development. When asked to explain the fundamental struggles that developing nations face, we

would look to Professor Sachs for the authoritative answer. Someone on the panel might have read his books and be quite familiar with his work, but his knowledge of the subject matter is not likely to be as deep and broad as Professor Sachs's.

We might also say that Professor Sachs's assertions about the matter of international development carry a certain authority. When he states that governance is poor in Africa because Africa is materially poor, then the assertion carries something with it, as opposed to when that same assertion is made by someone who is not an expert on the subject matter. What is this something? What does an assertion of an expert have that a non-expert's lacks? It seems natural to say that an expert's assertions have a sense of authority that a non-expert's lack. Additionally, the lay panelists would naturally defer to Professor Sachs's on questions concerning the subject matter of international development. It is not that he could not be wrong, but when the subject matter is international development he is assumed to know the answer given his mastery of the subject matter. This authority of utterances and assumption of authority appear to be two features seen in our authority over our own minds.

In our everyday practices we often assume that others are authorities about what they think, about the intentional contents of their minds. Because of this assumption, we defer to them when the subject matter is what they think. In addition to this assumption of authority, it also seems that one's assertions about what one thinks carry a sense of authority that the related assertions of others lack. I think that both of these ideas are relevant to our conception of first-person authority. Each person is assumed to be an authority on his own mind, and assertions that each person makes

about his own mind are taken as authoritative. These ideas of first-person authority have been articulated by many philosophers. In "Privacy," A.J. Ayer, for instance, writes,

When it comes, on the other hand, to a person's knowledge of his present thoughts and feelings, then I do think that there are many cases in which we logically are obliged to give him the last word. Even if we allow it to be possible for others to become aware of his thoughts and feelings in the way he does, their knowledge of them will be subordinate to his. The accuracy of their reports will be checked by his, and where there is disagreement his verdict must prevail. Thus, even if one's mental states are not private in the sense that there is any single way in which, of necessity, they are detectable by oneself alone, they may still be private in yet another sense. One may be the final authority concerning their existence and their character.¹³

According to my terminology, Ayer says that even if we were to give up on first-person access to intentional content, we need not reject first-person authority. Ayer actually calls the thesis above a version of privileged access, though he appears to be talking about first-person authority and not privileged access. Ayer suggests that we are logically obliged to take the first person's word over the third person's word. One question we might ask here in trying to get a clearer idea of this thesis is: What does Ayer mean to contribute to this claim by adding the adverb "logically"? In speaking to the logic of statements that make up such first-person reports, Ayer writes, "The logic of these statements that a person makes about himself is such that if others were to contradict him we should not be entitled to say that they were right so long as he honestly maintained his stand against them."¹⁴ Ayer is saying not that it would be a logical contradiction to side with the interpreter in this matter, but that it would be a conceptual confusion to do so. We naturally defer to the first person, of course, but the

¹³ (Ayer 1963), page 68.

¹⁴ (Ayer 1963), page 73).

point is deeper: it would be a conceptual error to take the third person as authoritative about the first person's mental states.

In *The Significance of Consciousness*, Charles Siewert offers what he calls the "claim of first-person warrant," which strikes me as quite similar to Ayer's claims of about first-person authority.

I hold that one has a type of warrant for some of one's beliefs or claims, assertible using a first-person singular pronoun to attribute some experience (or attitude) to oneself, that differs from the type of warrant had (ordinarily, at least) for any beliefs or claims, whose assertion would constitute the attribution of some experience or attitude only to someone other than the speaker.¹⁵

One way to understand talk of warrant for some belief or other is as authority over or about that belief. It is clear from the quote that, in Siewert's view, the first person has a special sort of warrant for beliefs or claims about his own experiences or attitudes.

In "First Person Authority," Donald Davidson echoes Ayer's point that the first person is in a position of authority with regard to his own mind. He writes, "In psychoanalytic practice, recovery of authority over an attitude is often considered the only solid evidence that the attitude was there before being noninferentially appreciated by its holder."¹⁶ This is the same as Ayer's idea that we must also use the first person as a check in deciding on the nature of his or her attitudes. The notion of first-person authority that emerges more centrally in Davidson's work, though, differs from Ayer's. Davidson's notion has to do primarily with an assumption that he claims is built into the nature of interpretation.

¹⁵ (Siewert 1998), page 6.

¹⁶ (Davidson 2001), page 7.

There is a presumption—an unavoidable presumption built into the nature of interpretation—that the speaker usually knows what he means. So, there is a presumption that if he knows that he holds a sentence true, he knows what he believes.¹⁷

All three of these thinkers are tracking similar ideas. Ayer and Davidson are focused on how someone else must see the speaker's authority over his intentional contents, while Siewert focuses on the type of authority that one's own reports of one's intentional states carry. If these thinkers are right, we must not only assume that subjects know what they think, but we must also grant that their first-person reports of their intentional contents have a special authority. Let the following two theses capture these various aspects of first-person authority.

- **The thesis of the presumption of first-person authority:** One is assumed to be an authority about the content of one's own intentional contents.
- **The thesis of the authoritativeness of first-person reports:** One's reports about one's intentional contents are authoritative.

Having laid out the various notions of self-knowledge, we are now in a position to see how these notions of self-knowledge are related.

How these Notions are Related

Here are the five theses that capture the epistemic notions related to self-knowledge:

- **The thesis of the transparency of intentional content:** Each thinker accesses the content of his intentional states directly and without the benefit of evidence.
- **The thesis of the asymmetry of access to intentional content:** One has a type of access to the contents of one's intentional states that is not had by others.

¹⁷ (Davidson 2001), page 14.

- **The thesis of first-person access to intentional content:** Each subject has first-person access to his own thoughts because each subject accesses his contents non-inferentially and without the benefit of evidence, whereas others access those contents by way of inference.
- **The thesis of the presumption of first-person authority:** One is assumed to be an authority about the content of one's own intentional contents.
- **The thesis of the authoritativeness of first-person reports:** One's reports about one's own intentional contents are authoritative.

Although their connection will be discussed again later in this chapter, it is worth briefly sketching how these notions are related. The thesis of the transparency of intentional content is the most basic notion. From that and the fact that there is an asymmetry of access to intentional content, we can infer the thesis of first-person access to intentional content. Seeing this package of notions in the context of communication allows us to more easily see the truth of the presumption of first-person authority and authoritativeness of first-person reports.

Clarifying the Picture

While we have made some progress in understanding the different notions of self-knowledge, more needs to be said. After all, I have not even given a simple example of self-knowledge based on the tripartite conception of knowledge. We shall see that when we think of self-knowledge as true, justified belief some puzzling aspects emerge. The main puzzle may already be evident: if self-knowledge does not require evidence, then one wonders what justifies self-knowledge. If knowledge is true, justified belief, then it seems that to have knowledge of my intentional content that Tallahassee is the capital of Florida, for example, the following necessary and sufficient conditions must be met.

- (a) I believe that I believe that Tallahassee is the capital of Florida.
- (b) It is true that I believe that Tallahassee is the capital of Florida.

(c) I am justified in believing that I believe that Tallahassee is the capital of Florida.

Can all these conditions been met? Based on what was said in the previous section, it appears that there is no straightforward answer to the question of why I am justified in believing that I believe p. After all, we saw that I typically do not appeal to any evidence. What sort of story would I tell about why I believe that I believe that Tallahassee is the capital of Florida? As we noted in discussion of the transparency condition, if asked for the reasons we hold our second-order intentional states, we are at a loss to provide justification for them and can pragmatically do no better than to cite reasons that support the first-order intentional states the second-order states are of. On the face of it, we appear to be at a loss to say wherein the justification for our second-order intentional states lies.

Are our second-order beliefs ever justified? Are we forced to say that we know our first-order states by having true beliefs about them? Paul Boghossian discusses this problem in “Content and Self-Knowledge.” He argues that we cannot know our thoughts on the basis of inference and thereby must know them on the basis of nothing empirical. Boghossian goes on to point out that if this is true, there are three possible types of justification for direct, non-empirical knowledge of contingent propositions.

The warrant for such judgments derives from other sources: from the meanings of the concepts involved, or from the satisfaction of general conditions, or from the judgment-dependent character of the phenomenon being judged. Whatever the source, no observation, or inference based on observational premises, is required or relevant. These judgments, when known, constitute knowledge that is based on nothing empirical. In my terms, they are not cognitive achievements and are subject, therefore, to an insubstantial epistemology.¹⁸

¹⁸ (Boghossian 2008), page 153.

From which of these sources does our justification for self-knowledge come?

One of the types of warrant mentioned by Boghossian clearly will not work: the method based on the meaning of concepts. We are perhaps justified in believing that all bachelors are unmarried men not because of empirical investigation, but because of the content of the concepts involved in the proposition believed. One test for whether justification derives from such analyticity is whether the sentence that expresses the proposition believed can be translated into a tautology. "All bachelors are unmarried men" can be translated without loss into the sentence "All unmarried men are unmarried men." Can we perform a similar translation of a sentence that expresses an instance of self-knowledge? The sentence that expresses the proposition believed is: "I believe that I believe that Tallahassee is the capital of Florida." Does the warrant for believing that this sentence expresses a truth come by analyzing the content of the concepts involved? While there are no doubt various translations of "to believe," there appears to be no tautology on the horizon. Therefore, it appears that if we do not know our thoughts on the basis of empirical evidence, then our justification for our beliefs about our first-order intentional states does not lie in *a priori* reasoning about the conceptual constituents of the propositions we believe when we have beliefs about our first-order intentional states. Therefore, assuming that Boghossian's categorization is exhaustive, warrant for self-knowledge comes from "the satisfaction of certain general conditions, or from the judgment-dependent character of the phenomenon being judged."

In "Individualism and Self-Knowledge," Tyler Burge offers a picture of self-knowledge in which our warrant comes from "the judgment-dependent character of the phenomenon being judged." There is a rather serious problem with this picture, though.

We do not need to form second-order beliefs for each particular first-order intentional state we have. Typically, outside of a philosophy seminar room or the psychoanalyst's office, we do not have much cause to reflect on whether we believe that we have such and such first-order intentional states. I may, no doubt, question why I believe that Madagascar's government is corrupt; ask myself why I desire to be more healthy; wonder about the basic source of my fear of riding in elevators; and ask myself why I hope that Obama will be reelected. All of this involves taking a reflective position with regard to my first-order intentional states; however, the very process of this reflection assumes that the first-order states are what I take them to be. Furthermore, I would still have these first-order states even when I do not make judgments about them. I believe that Madagascar's government is corrupt even when I am not forming the judgment that Madagascar's government is corrupt. Burge's view of self-knowledge (at least his view in "Individualism and Self-Knowledge") does not cover our standing intentional states. Surely our view of self-knowledge will cover both occurrent and standing intentional states.

The only option, then, appears to be that if our beliefs about our intentional states are justified, then they are justified by way of the satisfaction of general conditions. If this were true, it would explain why there is no justificatory story that is typically forthcoming from the first-person perspective. From the subject's point of view, there is nothing on which he is basing his belief that he believes what he believes. What justifies us in believing what we believe will come by way of the satisfaction of general conditions. The questions we must ask, then, take on a Kantian spirit.

- What are the general conditions that need to be satisfied such that the satisfaction of those conditions guarantees that subjects have justified beliefs about their first-order intentional states?
- What is it about the satisfaction of these general conditions that makes beliefs that one has about one's first-order intentional states justified beliefs?

A Transcendental Argument about Justification

What is it about the satisfaction of these general conditions that makes beliefs that one has about one's first-order intentional states justified beliefs? Of this sort of justification, Boghossian writes, "The truth of, and warrant for, the belief are secured, not by evidence, but by the satisfaction of certain very general conditions on experience. The thinker counts as knowing something thanks not to the possession of any evidence on his part, but simply courtesy of those general facts."¹⁹ Boghossian calls this type of knowledge "insubstantial." The idea here is like Kant's idea that knowing that the world contains substances is cognitively insubstantial. A thinker could not fail to be justified in believing that the world contains substances; the world having substances is a necessary condition on the very existence of agents with experiences and beliefs. From the fact that an agent experiences and believes anything, it is secured that there are substances. He is justified, then, not for reasons of his own, but because he cannot fail to be justified. Might the justification of self-knowledge be like this?

Because self-knowledge does not depend on evidence or inference, we have good reason to suspect that the justification of self-knowledge must be of a transcendental sort. When we think about the relationship between rationality, communication and self-knowledge, we see that certain conditions are required for certain phenomena. Here is a sketch of the argument that shall concern us. The first premise is that if we are to be

¹⁹ (Boghossian 2008), page 152.

viewed as rational agents, we must be viewed as having justified second-order beliefs. The second premise is that we must be viewed as rational agents if there is successful communication. The third premise is merely the claim that there is successful communication. From the second and third premise, we can conclude that we must be viewed as rational agents. From this conclusion, and the nature of how others must view us if we are to count as rational, it follows that our second-order beliefs are justified in virtue of how others must view us in a community of rational agents.

This argument appeals to the notion of the necessity of a rational agent being viewed in a certain way. A natural question to ask here is: viewed by whom? The above sketch does not state who must view the rational agent in this way. It could be others in that agent's community or it could be us—philosophers thinking about the structure of justification in a community of rational agents. There are two viewpoints from which we can ask questions about how we must view agents, about what assumptions must be made about them in the context of communication. In the previous sketch it is unclear which of these viewpoints I have in mind. I think that seeing things from both viewpoints is important, but it is the viewpoint of the theorist, the sort of viewpoint we take up now, that the argument must be run. I shall try to say why that is.

If we ignore the viewpoint of the theorist and focus just on the viewpoint of those in a rational community, we run into one main problem. If an agent's rationality and self-knowledge depends on being viewed by others as having those very features, then it appears that when one is not communicating with others, one may lose one's rationality and self-knowledge. Of course, these features of a subject should not at all depend on whether someone is actually interpreting that subject or not. It is obvious that a subject

can be rational and have self-knowledge even when he is not communicating with others.

Though we cannot run the argument from the viewpoint of the interpreter, taking up that viewpoint shows us that these are ways he must view the speaker if communication is to go off. He must view the speaker both as having justified second-order beliefs and as rational. Taking up this viewpoint of the interpreter is the easiest way for us to see how justification for self-knowledge arises from facts about communication. The next step is to take up the viewpoint of the theorist and see that as theorists we must also see subjects as having these attributes (being rational and having justified second-order beliefs) when we see subjects as agents among fellow rational agents in a community where there is communication. With that said, here is a summary of the argument.

- (P1) If we are to be viewed as rational agents, then we must be viewed as having justified second-order beliefs about our beliefs, which I shall call "self-reflexive second-order justified beliefs."²⁰
- (P2) Speakers must be viewed as rational agents by interpreters if there is successful communication.
- (P3) There is successful communication.
- (C1) Therefore, we must be viewed as rational agents.
- (C2) And so we must be viewed as having self-reflexive second-order justified beliefs in virtue of being viewed as rational agents.

In answer to the first question (What are the general conditions that need to be satisfied such that the satisfaction of those conditions guarantees that subjects have justified beliefs about their first-order intentional states?), then, it appears that the general

²⁰ In formulations of arguments such as the one above, "P" shall stand for "premise" and "C" shall stand for "conclusion." Thus, "P1" means premise 1 and "C1" means conclusion 1 and so on.

conditions that must be satisfied are conditions regarding communication in a community of rational agents. I think that the story here is somewhat complex. I will do my best to make it clear.

Many thinkers have pointed out that there is a connection between self-knowledge and rationality. Our present argument focuses on rationality as it relates to how interpreters must view speakers in terms of their rationality and in terms of whether or not their self-reflexive second-order beliefs are justified. If this argument is sound, then it secures some connection between self-knowledge and rationality, namely that the justification for self-knowledge is had automatically in virtue of the nature of communication. Before I discuss the support for (P1) and (P2), I want to discuss an idea that is shared by a number of philosophers working on self-knowledge and rationality, as it will be not only relevant to the argument at hand, but also relevant to another key argument, the argument for the conclusion that first-person access to intentional content is required for rationality.

Engagement and rationality

There are two ideas in the literature that I want to discuss before we get to the key arguments in this chapter. This stage setting will be helpful in bringing out a key point about how a subject must be related to his first-order intentional states, a notion that I shall call "engagement." The two notions I want to discuss are Richard Moran's idea of deliberative and theoretical stances that can be taken toward the content of intentional states and Sydney Shoemaker's notion of self-blindness. Let us discuss Moran's idea first.

In his book *Authority and Estrangement* Morgan discusses two types of stances a thinker can take toward his intentional states: a deliberative and theoretical stance. The

deliberative stance is the stance of practical reasoning when we ask ourselves what we should do, what we are to believe and so on. The theoretical stance is a stance we can take toward those same first-order states we use in practical reasoning. We take a theoretical stance to our first-order states when we ask ourselves what we are going to do and what we believe. In the deliberative stance we are directly engaged with our first-order intentional states. In the theoretical stance we examine data and make inferences based on such data. We clearly must take a theoretical stance with regard to the intentional states of others. While we do in fact take both stances toward ourselves, we are generally operating in the deliberative mode.

The distinction between these two stances is important for what Moran says about the connection between rationality and self-knowledge. In short, he thinks that the idea of a rational agent who takes only a theoretical stance to himself is an incoherent one. And this incoherence is supposed to show us something important about the relationship between an agent, his intentional states and his rational agency. Here is a lengthy quote from Moran.

The problem with the idea of generalizing the theoretical stance toward mental phenomena is that a person cannot treat his mental goings-on as just so much data or evidence about his state of mind all the way down, and still be credited with a mental life (including beliefs, judgments, etc.) to treat as data in the first-place. For any given mental presentation of mind, just as for any utterance, it may be true that I *can* treat it as data, something which gives me a more or less good indication of my genuine belief. But for there to be judgments or deliberation in the first place, I cannot adopt this point of view of my own mental life quite generally. At some point, I must cease attempting to infer from some occurrence to my belief; and instead *stake* myself, and relate to my mental life not as something of symptomatic value, but as my current commitment to how things are out there. And so, for this reason the abrogation of first-person authority is not made up for by improved theoretical access to myself.²¹

²¹ (Moran 2001), page 150.

This quote captures Moran's idea that a rational agent must be by and large engaged with the content of his intentional states. A deliberative stance toward his intentional contents must be taken up by a rational agent, even if he is able to take a theoretical stance to his intentional contents as well.

For what it is worth, Moran notes that this point about the need to adopt the deliberative stance is anti-Cartesian in certain respects. He notes that the Cartesian idea of a private theatre of ideas has the same flaws as a view of self-knowledge according to which an agent bears a theoretical stance to his intentional states. In both sorts of pictures, the agent is not properly engaged with his intentional contents. For Moran, this is a problem. He writes,

An idea as such [as something which an agent passively observes] is something I may be passive with respect to. It may be implanted in me by God or the external world or by an Evil Demon. But for that very reason it cannot be identical with my belief about some matter, for my judgments are *my* affair, something I am responsible for, through the exercise of my infinite liberty to affirm or deny.²²

Moran thinks that an agent with a purely theoretical stance toward his own intentional states is incoherent.²³ Our conception of a rational agent is of a subject who is engaged with his first order states in a way articulated by Moran's notion of the deliberative stance. The idea of an agent who bears a purely theoretical stance to his intentional contents, who has a purely third-person view of his own intentionality as we might say, is not a coherent idea.

This idea also comes out in Sydney Shoemaker's discussion of self-blindness. This is an idea that Shoemaker uses to defend certain aspects of the Cartesian Theory

²² (Moran 2001), page 149.

²³ (Moran 2001), page 149.

of Mind, namely that each person has a sort of access to the contents of his intentional states that no one else has—the thesis of first-person access to intentional content. Self-blindness is a condition where a subject, to interpolate Moran's terminology, bears a theoretical stance to the content of his intentional states. The basic idea comes out of Shoemaker's objection to the inner sense theory of self-knowledge. Shoemaker's argument is, in short, that if the inner sense theory of self-knowledge (the idea that we introspect our thoughts in much the same manner that we perceive the world) were true, then our thoughts would enjoy a certain independence from our awareness of them—just as the external world enjoys an independence from our perception of it. Given the independence of the thoughts from the self-blind individual's awareness of these thoughts, the individual's thoughts could change without him being aware of it. The conclusion, then, is that if the view leaves this possibility open, it must be false, for Shoemaker holds that such an individual is not conceptually possible. A being of this sort is not one we would describe as a rational human being. Here is Shoemaker himself on self-blindness.

What I wish to maintain is the impossibility of something I shall call "self-blindness." A self-blind creature would be one which has the conception of the various mental states, and can entertain the thought that it has this or that belief, desire, intention, etc., but which is unable to become aware of the truth of such a thought except in a third-person way. In other words, a self-blind creature could frame and understand ascriptions to itself of various mental states, but would be incapable of knowing by self-acquaintance whether such self-ascriptions were true. Only if self-blindness were a conceptual possibility would it be appropriate to think of the capacity for self-acquaintance as a quasi-perceptual capacity which is something over and above the capacity to have and conceive of the mental states in question. And it is the appropriateness of so thinking of it that I am anxious to deny.²⁴

²⁴ (Shoemaker 1996), pages 30-31.

It is helpful to see that Moran and Shoemaker both critique views of self-knowledge on which knowledge of intentional contents is had by way of a method of introspecting thoughts modeled on our perception of external objects. This is what Shoemaker intends to track when he talks of views that understand "self-acquaintance as a quasi-perceptual capacity" and what Moran has in mind when he critiques those views of self-knowledge modeled on the Cartesian theatre of ideas, wherein the mind's eye inspects its thought contents in a way similar to the way our eyes perceive external objects. The problem with such models, for Moran and Shoemaker, is the introduction of data and distance, we might say, between an agent and his first-order states. On such models, access becomes of a third-person sort: data is collected and inferences are made. Both Moran and Shoemaker think that it is incoherent to see an agent as having merely this sort of third-person access to his own states, of having, in Moran's words, a purely theoretical stance to his own intentional contents.

The overall lesson from Moran and Shoemaker's work is that an agent must be properly engaged with his first-order intentional states if he is to count as rational. To be properly engaged with his first-order intentional states, an agent must have first-person—and not third-person access—to them. This general idea is one I will exploit in both of the following arguments. Let us turn back to the transcendental argument about justification.

Why rational agents must be viewed as having justified second-order beliefs about first-order intentional states

In our present inquiry we focus on the viewpoint of the interpreter in a rational community, though we must also keep in mind that the argument must be made from the point of view of the theorist. To talk about our current practice of rationality is to talk

about a community of rational agents. In delving into the necessary conditions of such a community, one might think that it is enough that all the members of this community be sufficiently rational: to be able to perform simple acts of reasoning based on a shared environment, to be able to engage in dialogical reasoning with one another, to measure each other's reasoning by a normative standard, and so on. Why must it also be the case that those who are rational in such a community be seen as having justified self-reflexive second order beliefs?

A rather simple observation here is that without justification for his second-order beliefs, a subject would lack knowledge of his intentional contents. Must an interpreter view a subject as having knowledge of his intentional contents if that subject is to count as rational? In "First Person Authority" Davidson argues that it must be assumed that speakers know what they mean if they are to be interpretable. That argument, though, concerns knowledge of meaning and the present inquiry concerns knowledge of intentional contents.

Without justification for self-reflexive second-order beliefs, the agent would lack knowledge of a crucial class of beliefs: beliefs about the intentional contents of his own mind. It would certainly be odd to view someone as lacking knowledge of their intentional contents, but we need a reason for thinking that viewing them as lacking such knowledge is viewing them as somehow less than rational. The reason that we must view rational agents as having such knowledge is that to view them as lacking it is to view them as continuously bearing a theoretical stance to their own intentional states, because the lack of justification becomes possible only when we assume that there is some data on which they come to know the content of their intentional states. When one

is seen as bearing a deliberative stance to one's intentional states, the possibility of lack of justification is foreclosed: there is no data or evidence being used such that there can be a failure to use it in the proper way. As Moran and Shoemaker have argued, viewing an agent as merely bearing a theoretical stance to his own intentional states is an incoherent idea. A subject must have ownership of his intentional states; he must, as Moran puts it, stake himself to his intentional states. To view a subject as lacking this is not to view a subject as lacking intentional states, but to view him as having an aberrant relationship to his intentional states when viewed in the light of rational agency. His intentionality is not fully his. In short, he is passive with regard to it, and not its agent.

Why speakers must be viewed as rational agents by interpreters if there is successful communication.

The previous section leaves us with the question of why we must assume that subjects are rational. The above argument only works provided that those in a community of agents must assume that their fellow agents are rational. Davidson gives a compelling argument that such an assumption must be made by interpreters. In “Three Varieties of Knowledge” he writes:

The process of separating meaning and opinion invokes two key principles which must be applicable if a speaker is interpretable: the Principle of Coherence and the Principle of Correspondence. The Principle of Coherence prompts the interpreter to discover a degree of logical consistency in the thought of the speaker; the Principle of Correspondence prompts the interpreter to take the speaker to be responding to the same feature so the world that he (the interpreter) would be responding to under similar circumstances. Both principles can be (and have been) called principles of charity: one principle endows the speaker with a modicum of logic, the other endows him with a degree of what the interpreter takes to be true belief about the world. Successful interpretation necessarily invests the person interpreted with basic rationality. It follows from the nature of correct interpretation that an interpersonal standard of consistency and

correspondence to the facts applies to both the speaker and the speaker's interpreter, to their utterances and to their beliefs.²⁵

If this argument is sound, it appears that we have good reason to believe that speakers must be viewed as rational agents by interpreters if there is successful communication.

Why believe, though, that we must apply these principles in interpretation?

Davidson, like Quine before him, employs a thought experiment with a speaker and an interpreter who do not share the same language, and wherein the interpreter has no knowledge of the speaker's language. Stripping away knowledge of a shared language is supposed to reveal the fundamental features of interpretation. To make sense of the speech of the foreign speaker, the interpreter must assume that the speaker is rational and that his beliefs are largely about the portion of the world they are sharing. Without the latter assumption, the interpreter would be at a loss to assign content to the speaker's utterances. He can do no better than to assume that the speaker is talking about the objects, properties and relations in the environment they share. As for the former assumption, citing Quine, Davidson talks of the necessity of interpreters reading their own logic into the minds of speakers.²⁶ It is hard to imagine how we would go about interpreting someone if we did not assume they were rational to some degree. Points at which we fail to be capable of interpreting one another are either points at which it is clear that both parties are not talking about a shared environment (as when someone is delusional) or at points when the interpreter cannot attribute rational thought to the speaker (as when someone is schizophrenic). The thought experiment then requires us to imagine things from the standpoint of the interpreter; this view point, it

²⁵ (Davidson 2001), page 211.

²⁶ (Davidson 2001), page 149.

seem to me, reveals at least one conceptual truth: without assuming that the speaker is rational, there is no way for successful interpretation to proceed. It is hard to see how we would begin to interpret the words of another without assuming that their method of reasoning about the world is not sufficiently similar to ours.

So, in sum, if an agent is viewed as rational, he must be viewed as having self-reflexive justified second-order beliefs. Further, if an agent is interpretable, he must be assumed to be rational. Thus, if there is true communication and agents are interpretable, then they are rational, and in virtue of the assumption of rationality, they must be viewed as having self-reflexive justified second-order beliefs, for to assume that they could lack such justification is to assume that they continuously bear a theoretical stance to themselves.

A Transcendental Argument about First-Person Access

In this section, I shall offer another argument related to the idea of engagement and rationality. This argument concerns the type of access that a subject must have to his intentional states to count as rational. Sydney Shoemaker has offered various arguments for the conclusion that a subject must have first-person access to his intentional content if he is to count as rational. Charles Siewert discusses four of these arguments in "Self-Knowledge and Rationality: Shoemaker on Self-Blindness." We shall focus on one that concerns the need for a rational agent to revise his beliefs, and also focus on Siewert's commentary on the argument.

In "Self-Intimation and Second Order Beliefs" Shoemaker offers a succinct summary of the need for a rational agent to revise his beliefs for coherency.

It is a condition of being a rational subject that one's belief system will regularly be revised with the aim of achieving and preserving consistency

and internal coherence, and that such revision requires awareness on the part of the subject of what the contents of the system are.²⁷

It is important to bring out the role that self-reflexive beliefs play in this account. They are what provide awareness to the subject about what his first-order contents are. It is these self-reflexive beliefs that must be accessed in a distinctly first-person way, a way that does not depend on evidence or inference. If these self-reflexive beliefs were accessed by inference and evidence, then they would not be sufficient for revising the intentional contents of a rational agent. To this, Siewert adds that it may also be required of a rational agent to not only be able to revise his beliefs for coherence, but also to be able to justify such revisions. This demand seems reasonable: it would be odd if a rational agent revised certain beliefs and was unable to justify the revision for reasons other than inability to remember.

That point of course should be relatively uncontroversial. The significant claim is that one must have first-person access to one's intentional contents in order to perform the rational acts of revision and justification with regard to one's intentional contents. In a sense, both the demand that a rational agent be capable of revising intentional states for coherency and the demand that he be able to justify such revisions is a demand for the same ability: the ability to access first-order and second-order intentional states in a first-person way—a way that does not depend on evidence or inference. Here is a quote from Charles Siewert that helps to bring out why this is so.

If the practice of justifying, as we engage in it, requires the ability to cite accurately what beliefs, desires, and experiences we have, and if there is often no available third-person source for these citations, one must be able

²⁷ (Shoemaker 2009) page 39.

to represent one's own attitudes and experiences to oneself accurately by a distinctively first-person means.²⁸

There are many instances when no third-person sources for citations are available. When I engage in practical reasoning, I do not wait to hear myself speak, for instance, before I access the intentional content that my speech makes public. Let us say I am trying to decide whether to vote for Barack Obama in the 2012 presidential election. I draw on the contents of many intentional states in trying to come to a conclusion. For instance, my belief that Barack Obama is a pragmatist and not an ideologue factors into my reasoning. It is not as if I must assert "Barack Obama is a pragmatist and not an ideologue" to know the content of my thought. I access the content of the belief directly. The point is not merely one about how I *in fact* access the intentional content, though. The point to bring out is that if I did have to access my intentional content by waiting to hear what I said, I would not be properly engaged with my intentional content. The possibility of self-blindness would be open, for my intentional contents would be logically independent of my self-reflexive states that track them. I would, as Moran would say, seem to lose my stake in them if it turned out that I had to access them in a third-person way. They would not—in a very important sense—be mine. If we lacked first-person access to our intentional content, we would lack the proper engagement with those contents and we would have to give up on the picture sketched between the tight connection between self-reflexive states and the intentional states they track, a connection that does not depend on any evidence. If we give up such a picture, then we give up on rational agency as we know it. Let us end with a summary of the argument just given.

²⁸ (Siewert 2003) page 136.

(P1) In order to be rational, one must revise one's beliefs in light of inconsistency and be capable of justifying those revisions.

(P2) Revising one's intentional contents in light of inconsistency and justifying such revision requires first-person access to one's intentional contents.

(C1) If we are rational, we must have first-person access to our intentional contents.

Explaining Self-Knowledge

As we saw, the most basic notion of this overall picture is the transparency of intentional content. Because it concerns the direct and unmediated access each person has to his own intentional states, we can understand this by thinking of individuals in isolation. Unlike the notion of the transparency of intentional content, we cannot screen off the relationship between the individual and others to understand the rest of the notions of self-knowledge. Once we accept it is just the subject who can access his intentional states in a direct way, we can clearly see the truth of the thesis of asymmetry of access to intentional content. We have seen that this notion helps explain first-person access to intentional content and the notions of first-person authority; the transparency of intentional content is also central to an understanding of the connection between self-knowledge and rationality. Before I said that first-person access to intentional content helps explain first-person authority of self-reports and it helps explain the assumption of first-person authority that is embedded in our communicative practices. Without first-person access to intentional content, it would be puzzling why we assume that others are authorities about their intentional contents and why we grant their self-reports of those contents authority. Understanding the story about access helps us to see why these two notions first-person authority arise in the context of communication.

Merely seeing the connection between notions of access and authority does not fully explain self-knowledge. We also had to see how our second-order beliefs could be

justified in the absence of evidence. Locating the source of his justification required looking deeper into the necessary features of the structure of the justificatory and communicative practices in a community of rational agents. We saw that one's justification for one's self-knowledge lies in facts about how others must view a rational agent if he is to be interpretable. We also saw that we must have first-person access to our intentional content if we are to count as rational agents. If we lacked such access, we would be in the position of Shoemaker's self-blind individual, who is related to his own intentional contents in a way insufficient for rationality. Though it was not discussed in the two main transcendental arguments, we can also say that if a subject lacked first-person authority, meaning that he was such that interpreters did not assume that he was an authority about his thoughts, then it would be hard to see how such a subject could be rational. Like first-person access to intentional content, the two notions of first-person authority seem crucial for rationality. I cannot give a lengthy defense at this point, but let me offer some remarks that can make an intuitive case for the notions of first-person authority being crucial for rationality. To see that the two notions of first-person authority are crucial for rationality, one can attempt to imagine an agent who has first-person access to his intentional contents, but is not presumed by his neighbors to be an authority over what he thinks and makes self-reports on his thoughts that do not have any more authority than his neighbors' reports. Assuming that the communicative environment is normal, there does not seem to be a possible situation in which a rational agent lacks the two aspects of first-person authority under discussion. Our rationality depends first and foremost on first-person access to intentional content,

although the two aspects of first-person authority appear to be important to rationality as well.

To end the chapter, I should note that it is a consequence of this view of self-knowledge that there is a sense that we are infallible about what we think. At first glance, this may seem damning. Surely there are some times when we are mistaken about our thoughts. My view allows for mistakes, but only mistakes of a certain sort. We must admit that there is sense in which we cannot be mistaken about the content of our intentional states. If we access these without evidence or inferences, there it is hard to see how we could be mistaken about our intentional contents. However, we can make mistakes about what we think when we improperly describe our thoughts. Everyone is familiar with this. If someone tries to disparage comparisons by saying, "Comparisons are odorous," we could jocularly ask him why he thinks that comparisons give off a bad smell. Realizing that he misspoke, he will likely go about correcting himself and soon enough get to a proper articulation of the thought. All the while, though, he is not mistaken about the content of the thought: that comparisons are odious. What he gets wrong is not the content but the proper linguistic expression of the content. It turns out that we cannot be mistaken about the content of our intentional states, though we are sometimes mistaken about what counts as an adequate expression of them.

CHAPTER 3 EXTERNALISMS

I said earlier that one does not need to be steeped in philosophy to see that we know our minds in a special way. The same cannot be said of the truth of content externalism. Perhaps this is because self-knowledge plays a more fundamental role in communication, a role that helps us to more readily see that it is true. The idea that our thoughts have content, that they carry information about the world, is not foreign to most. However, questions about what this content depends on are less accessible to someone unfamiliar with the literature on content externalism and internalism. Self-knowledge is surely a more familiar concept than the concept of content externalism, the idea that intentional content depends on features of the environment. I hope to make content externalism clear by the end of the chapter, to articulate the arguments that support it, and to distinguish it and another form of externalism.

Thinking about Twin Earth

In “The Meaning of ‘Meaning’” Hilary Putnam asks philosophers to consider Twin Earth, a world which is like ours in every detail, except that instead of having lakes, rivers, oceans and rain composed of H₂O, the lakes, rivers, oceans and rain on Twin Earth are made up of another chemical element with a long and complicated chemical formula that for ease we can abbreviate as XYZ. This familiar scenario has been used to advance a version of content externalism. Putnam asks us to imagine two individuals who are the same on the inside, who share the same phenomenological properties (the world seems the same to each of them) and other intrinsic properties (their brain states are the same, for example). One individual is on Earth, and the other is on Twin Earth. The relevant question for content externalism is this. When these individuals assert,

“Water is vital to human life,” do they share the same belief or not? Here is a short argument that says they do not. Due to the difference in their external environments, one believes that XYZ is vital to human life, and the other believes that H₂O is vital to human life. Remember that Twin Earth has also taught us that water is necessarily H₂O (more on this later). Thus, we can conclude that the individual on Earth believes that water is vital to human life and the individual on Twin Earth believes that a water-like substance (sometimes called "watery stuff") is vital to human life. Remember that the individuals are the same on the inside; they have the same intrinsic properties. If we are fine with each step so far, then we must conclude that what determines the content of the belief is not solely a matter of what is inside of a subject. It is partly a matter of a subject's environment.¹

Thinking about Arthritis

Another influential argument for the same conclusion is found in the work of Tyler Burge, in particular “Individualism and the Mental.” Burge's argument differs from the previous one in that he asks us to focus not on the chemical properties of substances in the environment (being H₂O and being XYZ, for example), but on the linguistic meaning of words. In his famous example, like Putnam's Twin Earth, we have two individuals with the same intrinsic properties in different environments. One individual is in the actual world; the other is in a counterfactual world. Both individuals believe that they have arthritis in their thighs. We know this because it is stipulated that both would assent to the sentence, “I have arthritis in my thigh.” However, “arthritis” means different things in each community. In the actual world, “arthritis” means a rheumatoid ailment

¹ There is more to say about the Twin Earth argument. I give a brief characterization here, so as to introduce both it and Burge's version of externalism before going into more detail about each argument.

exclusively of the joints. In the counterfactual world, “arthritis” means any rheumatoid ailment, not merely a rheumatoid ailment of the joints. In each community, the individuals use the word “arthritis” in a way that suggests that each individual at least partly grasps the intersubjective concept expressed by the word. The individual who lives in the *actual* world has a less than full grasp on the intersubjective concept because he thinks that arthritis is any rheumatoid ailment and not an ailment exclusively of the joints. Since it is proper to describe each of these beliefs using the term “arthritis,” and because the linguistic meaning of “arthritis” differs at each world, Burge concludes that the intentional content of the beliefs differ. The subject in the actual world believes that he has arthritis (that is, a rheumatoid ailment exclusively of the joints) in his thigh. The subject in the counterfactual world believes that he has “tharthritis” (that is, a rheumatoid ailment of the joints *or* the muscles) in his thigh. The beliefs have different contents. Yet, the individuals are the same on the inside. Therefore, what determines the content of a belief is not simply a matter of what is inside of a subject, but partially determined by what is outside.

It seems that the following thesis is general enough to capture what is important to both conclusions.

The thesis of content externalism: The contents of intentional states are not merely a function of the intrinsic properties of their owners.

In order to properly understand this thesis, let us review some important terminology.

Let us discuss intrinsic properties first. A wedding band, for example, has intrinsic and extrinsic properties. Some of the features of my own wedding band depend on the environment independent of my ring being a certain way. For example, being an indicator of my marital status is an extrinsic property of the band, because it requires the

institution of marriage, not something that is a function of my intrinsic properties.

Weighing one ounce is also among its extrinsic properties, because its weight depends on the gravitational field of the environment. My ring's mass, however, is a feature of my ring that is not dependent on the nature of the environment independent of the ring. My ring has the same mass wherever it is. This is not true of its weight and its ability to serve as an indicator of my marital status. On Mars my ring would have a different weight. In a possible world with no institution of marriage, my ring would not indicate anything about my marital status. We can see that there is a test for whether or not something is an intrinsic property. If the environment independent of the individual who has the property cannot possibly be changed in a way such that it changes the property in question, then the property is intrinsic. We can use a modified version of Descartes's evil demon thought experiment to provide a ready example of a situation in which the content of intentional states depends on just the intrinsic properties of an individual. Let us say that my thoughts are as they are now, and that the external world is by and large as I take it to be. The people I see are really there, as are the other middle-sized objects, and so on. As I sleep tonight, suppose that an evil demon obliterates everything in the world except my mind (and himself). Suppose that he is able to manipulate my mind in such a way that I am not aware that anything has changed. My thoughts about the world appear to me as they did before the evil demon's destructive work. Such a situation is coherent only if intentional content depends on intrinsic properties alone: in focusing on the properties of mine on which the content depends, the content of my intentional states must surely depend on my intrinsic properties alone for the situation to

be capable of coherence, because if my intentional states depended on my extrinsic properties, they would be radically altered with the alteration of my extrinsic properties.

I should add, in case there is a question, that the notion of dependence in question is a logical one. For example, to say that free will is required for moral responsibility is to express a logical requirement or logical dependence. Free will is a necessary condition of moral responsibility. To say, as the content externalist does, that the content of intentional states depends on the environment independent of a subject being a certain way is to express a similar logical requirement, a similar necessary condition. Without the environment being such and such a way, one would not have such and such thought. We can put this talk of logical dependence in terms of something being a function of something else. So, to say that x is not merely a function of y is to express a relation of logical dependence. One's culinary skill for example is not merely a function of one's genetic make-up: many other factors are involved in whether or not someone can cook well. To be a content externalist is to say that one's thoughts are not merely a function of an individual's intrinsic properties, but a function of something more than merely one's intrinsic properties, namely one's extrinsic properties as well.

Understanding why the Arguments have been so Popular

We have seen two quick arguments for content externalism. We need to have a more careful look at them now, and also discuss why they have become so popular. To do this, I will discuss each argument and surrounding issues separately. Let me begin with a discussion of Putnam's Twin Earth.

Natural Kind Externalism

The externalism that is supported by reflection on Twin Earth has been called “natural kind externalism” by Jessica Brown.² I shall adopt her terminology in this essay. The Twin Earth thought experiment was introduced by Hilary Putnam in his well-known paper, “The Meaning of ‘Meaning’.” It is worth noting that the question of whether intentional content depends on a thinker’s extrinsic properties is not even asked in “The Meaning of ‘Meaning’.” Because of this, and because I think it is important to sort out the various philosophical theses supported by Putnam’s famous paper, I shall begin my discussion here not by discussing Twin Earth and content externalism, but by focusing on his primary aim of the paper: to show that the traditional theory of meaning is incoherent because it rests of two assumptions that cannot be, in Putnam’s words, “jointly satisfied.”³ The assumptions are:

Assumption I: That knowing the meaning of a term is just a matter of being in a certain psychological state.

Assumption II: That the meaning of a term (in the sense of “intension”) determines its extension (in the sense that sameness of intension entails sameness of extension).⁴

The Twin Earth thought experiment was originally designed to show two things: that one of the traditional assumptions must be rejected in order to maintain a coherent theory of meaning, and that psychological states alone do not determine the extension of terms.

Let us see how thinking about Twin Earth supports these goals.

Let us begin by reviewing how Putnam’s argument is supposed to work. We are asked to imagine a world, Twin Earth, that is nearly identical to our own, except for one

² See (Brown 2004).

³ (Pessin and Goldberg 1996), page 6.

⁴ (Pessin and Goldberg 1996), page 6.

important difference: in this world, there is no H₂O. There is a substance that appears to be water; it looks like water and it does what water does. However, its chemical structure is different. It is described by a long and complicated chemical formula that we can abbreviate as XYZ. On Twin Earth, XYZ is called “water,” and this substance plays the same role in the lives of Twin Earthians as it plays in the lives of Earthians. Now, take an Earthian and a Twin Earthian. Suppose they each have the same type-identical psychological states. And let us stipulate that both are ignorant of the chemical structure of the substances they refer to using the word “water.” Putnam urges us to think that the extension of the term “water” is different at each world. On Earth “water” refers to H₂O, and on Twin Earth it refers to XYZ. This may seem harmless enough. If we grant that the extension of “water” is different at each world, then we will see that an individual’s psychological state alone does not determine extension of the term “water,” for the two individuals are in the same type-identical psychological state and the extension of “water” is different at each world. Of course, there is nothing special about the term “water” here. The argument is supposed to generalize to other terms, the lesson being that something more is relevant than an individual’s psychological states when it comes to determining extension.

If we are with Putnam so far, we can see how the Twin Earth thought experiment shows that no theory of meaning can be grounded by both Assumption I and Assumption II. We cannot accept both that the extension of the term “water” is different at each world and that the two assumptions are true. If the extension of “water” is different, then our speakers cannot be said to have knowledge of the meaning of “water,” for they are in the same psychological state and the meaning of “water” is

different per Assumption II. Putnam says that we must reject one of the assumptions.⁵ He favors rejecting Assumption I.

This conclusion has perhaps been overshadowed by the idea that Twin Earth shows us something important about *a posteriori* necessities. *A posteriori* necessities are necessary truths that cannot be known *a priori*.⁶ The discovery of such truths by Putnam, and also Saul Kripke, was seen as a groundbreaking philosophical insight, because it was traditionally thought that all necessary truths were knowable only *a priori*. Let us review the argument for these truths, drawing on both the work of Putnam and Kripke.

Perhaps the best place to start is with an important notion employed by the argument, the notion of a rigid designator.⁷ A rigid designator is a term that refers to the same object in all possible worlds. Here is Kripke.

Let's call something a *rigid designator* if in every possible world it designates the same object; a *nonrigid* or *accidental designator* if that is not the case.⁸

⁵ Although Putnam does not explore the option that the extension of "water" is the same at each world, it is possible that the extension of "water" at both worlds is H₂O or XYZ or whatever plays the water role. Eddy Zemach (1976) makes a strong case for just this, basing his argument on our actual practice of referring to substances.

⁶ By knowing something "*a priori*" I mean knowing something without the benefit of empirical information.

⁷ Putnam seems happy to talk of rigid terms and rigid designation.

Kripke calls a designator "rigid" (in a given sentence) if (in that sentence) it refers to the same individual in every possible world in which the designator designates. If we extend the notion of rigidity to substance names, then we may express Kripke's theory and mine by saying that the term "water" is rigid. (Putnam 1975), page 16.

⁸ (Kripke 1980), page 48.

For Putnam and Kripke, natural kind terms such as “water” are rigid designators that have their actual-world referents in every possible world, or if an actual-world referent of a rigid designator does not exist in a given possible world, then that rigid designator does not refer. “Gold,” for example, refers to Au in all possible worlds and no other substance, no matter how superficially similar that substance may be to gold. And if there is no Au in a world, “gold” has no referent in that world.

Based on this view of reference fixing, Putnam and Kripke think that it is logically impossible for water to be anything other H₂O. Once we fix the reference of “water” based on the story of rigid designation, it is an *a posteriori* necessity that nothing that is not H₂O counts as water.

Nothing I have so far recounted directly supports content externalism.⁹ In the original paper, as I noted, Putnam does not say anything about content externalism, although he did offer some commentary later. Consider the following quote from his “Introduction” to *The Twin Earth Chronicles*, a book published 20 years after the original publication of “The Meaning of ‘Meaning’”.

Of course, denying that meanings are in the head must have consequences for the philosophy of mind, but at the time I wrote those words I was unsure as to just what those consequences were. After all, such accomplishments as knowing the meaning of words and using words meaningfully are paradigmatic “mental abilities”; yet, I was not sure, when I wrote “The Meaning of ‘Meaning’,” whether the moral of that essay should be that we shouldn’t think of the meanings of words as lying in the mind at all, or

⁹ Putnam’s rejection of Assumption I may seem to support some type of externalism. However, it does not support content externalism, but a rather uninteresting form of externalism about what is required for knowledge. Nearly everyone would agree that a proposition is known only if it is true. The truth of a proposition is in most cases an extrinsic property of that proposition. One may be able to argue that if a proposition is true in virtue of its meaning alone, then its truth is an intrinsic property of that proposition. There are a great many propositions, of course, whose truth is an external matter. Even so, the truth of a proposition does not depend on just the intrinsic properties of one who believes that proposition. Therefore, an individual’s knowledge is not dependent on an individual’s intrinsic properties.

whether (like John Dewey and William James) we should stop thinking of the mind as something “in the head” and think of it rather as a system of environment-involving capacities and interactions. In the end, I equivocated between these views. I said, on the one hand, that “meanings just ain’t in the head,” and, on the other hand, that the notion of the mind is ambiguous, and that, in one sense of “mental state” (I called mental states, in this supposed sense, ‘narrow mental states’), our mental states are entirely in our heads, and in another sense (I called mental states in this supposed second sense “broad mental states”) states are individuated by our relations to our environment and to other speakers and not simply by what goes on in our brains. Subsequently, under the influence of Tyler Burge and more recently of John McDowell as well, I have come to think that this conceded too much to the idea that the mind can be thought of as a private theater (situated inside the head).¹⁰

This quote makes it very clear that Putnam made no direct argument for content externalism in the original paper. At this point, there are some important questions we need to ask: What exactly is the argument for content externalism inspired by Twin Earth? What connection does this argument bear, if any, to the morals that Putnam originally draws from Twin Earth? Why did the argument for content externalism inspired by Twin Earth become so popular among philosophers?

Let us begin with the first question. Of Twin Earth, it is common to read philosophers writing things like the following: “Although such a thought experiment was designed to establish semantic externalism, it can be extended to mental contents as well.”¹¹ Of course, it is not so clear that the argument is being extended, as much as Twin Earth is being deployed as a backdrop or as a handy thought experiment for a different argument. Colin McGinn was perhaps the first to use Twin Earth to argue for content externalism.¹² His proposal is as follows. When an individual on Earth assents

¹⁰ (Pessin and Goldberg 1996), page xvii.

¹¹ (Lau and Deutsch 2008).

¹² See (McGinn 1977).

to the sentence, "Water is vital to human life," the content of his belief is true if and only if H₂O is vital to human life. Contrarily, when an individual on Twin Earth assents to the sentence, "Water is vital to human life," the content of his belief is true if and only if XYZ is vital to human life. As we know, the individuals are the same in terms of their intrinsic properties. The beliefs are different, and the difference must lie in the respective environments. So, their intentional content is not merely a function of their intrinsic properties. Therefore, content externalism is true.

Here is the argument generalized.

(P1) Beliefs and other intentional states are to be individuated by their truth conditions.

(P2) It is possible for individuals to have the same intrinsic properties and yet assent to a sentence that has different truth conditions in different worlds.

(C1) Therefore, it is possible to have two individuals who share all the same intrinsic properties and have different beliefs.

(C2) Therefore, some intentional content is not solely a function of an individual's intrinsic properties.

(P1) seems to be a premise that nearly everyone would accept. The truth of (P2) is supposed to be seen by reflecting on Twin Earth. (C1) follows from (P1) and (P2), and (C2) is merely a different articulation of (C1). Setting things up like this helps us to see that the question turns on the truth conditions of the sentences assented to. Those who think that water is necessarily H₂O will think that when one assents to the sentence "water is vital to life" on Twin Earth one does not believe that water is vital to life, because the previous sentence is true on Twin Earth if and only if XYZ is vital to life, and nothing is water if it is not H₂O.

It is important to see that adopting the story of *a posteriori* necessities answers the question of what the proper truth conditions are. But we might pause to ask

ourselves whether this is the only story one can adopt. Some may think that water is not necessarily H₂O. In this case “water is vital to life” is true if and only if H₂O or XYZ or whatever plays the water role is vital to life.¹³ I will give a more in-depth argument for how we should think of the truth conditions for intentional states in Chapter 5. For now, let me note that there is an important difference with regard to the truth conditions of sentences full stop and truth conditions of sentences in an intentional context. By “intentional context,” I mean a context having to do with the content of intentional states. When discussing the truth conditions of a sentence assented to by a subject, for example, we are discussing truth conditions in an intentional context. Consider our sentence “Water is vital to human life.” If a philosopher thinks that water is necessarily H₂O, he will judge that the sentence is true if and only if H₂O is vital to human life. Even if someone believes that, one may still think that the truth conditions of the sentence in an intentional context may differ. For example, let us suppose that I know nothing of water's chemistry. I say, “Water is vital to human life.” In the context of determining intentional content, we may have good reason for thinking that the truth conditions of this sentence change when seen in an intentional context. After all, I may know nothing about H₂O, in which case we should ask ourselves whether it makes sense to attribute to me beliefs about H₂O.

Some may think that how we decide on the proper truth conditions of “water is vital to life” relative to a subject's intentional state comes down to how we think about

¹³ In “The Transparency of Mental Content,” Boghossian states that the Twin Earth thought experiment shows that “either mental contents are not individuated individualistically or they are not individuated in terms of their truth conditions” (Boghossian 2008), page 160. I think that this is mistaken. One could accept (P1) and still reject (P2). One could, that is, hold (P1) and defend content internalism by arguing that the truth conditions of the beliefs about water are the same (see (Crane 1991) for such a move).

one's personal concept of water. For the natural kind externalist, our concepts are shaped, in a rather strong way, by the nature of the environment. The following example brings this out. Take an individual who has been raised on Earth. One night he is taken from his bed while asleep and transported unawares to Twin Earth. When he wakes the next day, the world around him appears normal to him. He has no idea of what happened. He proceeds to go about his daily affairs. Remember that there would be nothing to trip him up: the world seems the same to him.¹⁴ One question we might ask here is this: Are his thoughts different in virtue of his new environment? Most externalist say no. They think that initially when he thinks about water, he thinks about H₂O. After a certain period of time, though, his thoughts about water would switch from being thoughts about H₂O to being thoughts about XYZ in virtue of the new environment shaping his concepts. Remember that nothing has changed for the subject. According to the content externalist, one day he will have been on Twin Earth long enough and his thoughts will thereby change their contents without him realizing it. From his point of view, there will have been no change to his concept of water. He will have learned no new information about water (or watery stuff if you like); he will certainly not have learned that the watery stuff is XYZ. Yet, his concept of water, the concept that is part of the structure of his thoughts about water will have changed. His belief that water is vital to human life is true if and only if XYZ is vital to human life.

One gets the sense that many are moved by this thought experiment to think that the concept would change without the subject's awareness. We do not need to decide what to say about this now. I am merely trying to sketch the details of the content

¹⁴ Let us suppose that we have kidnapped and successfully hidden his doppelganger on Twin Earth.

externalist argument and to lay out some points that will be important later. Two points bear repeating. First, the conclusion that the truth conditions of the respective Earthian and Twin Earthian beliefs differ can be supported by the content externalist who adopts the story of *a posteriori* necessities. Second, it looks like even if one adopts the story of *a posteriori* necessities, one can still resist the conclusion that the truth conditions differ by noting that the truth conditions may be the same when viewed in an intentional context. When looking at the subject's concept of water, we may find that the truth conditions of the intentional states of the Earthian and Twin Earthian are actually the same. These points will be discussed in more depth in Chapter 5.

This Twin Earth inspired argument, as I see it, is not supported by Putnam's argument against psychologism or semantic internalism, but by the conclusions he draws about the essence of substances.¹⁵ Someone might think that Putnam's argument against semantic internalism—the view that meaning is a function of an individual's intrinsic properties—might be the driving force behind the argument for content externalism. However, it turns out that his essentialist conclusions of natural kinds, for which Kripke is typically given equal credit, are what are most important to Twin Earth argument for content externalism.

Why did the argument for content externalism inspired by Twin Earth become so popular among philosophers? There is no doubt that the popularity of Putnam's Twin Earth thought experiment has lent some credibility and authority to the Twin Earth inspired argument for content externalism. There should also be no doubt that a general distrust of the Cartesian Theory of Mind, one aspect of which is the idea that the mind is

¹⁵ In Chapter 5, I urge that we make a distinction between semantic externalism and content externalism.

independent of the environment, has made the argument for content externalism more appealing. It seems natural that these factors would work together to incline some to see natural kind externalism as part of a package that represents philosophical progress. It is not hard to see why one would think that semantic externalism, natural kind externalism, anti-Cartesianism about the mind, and conclusions about *a posteriori* necessities are an attractive collective of views for a contemporary philosopher to hold. We will see that it is not necessary to adopt this package wholesale. We have good reason for rejecting some of these theses while accepting others.

Social Externalism

Let us move on to Burge's version of content externalism, which is sometimes called "social externalism" due to Burge's focus on the role that sentence meaning plays in shaping our thoughts. To review, here is a condensed version of Burge's argument. Burge asks us to imagine two different situations, one actual and one counterfactual. He asks us first to imagine an individual in the actual world who has many true beliefs about arthritis and the false belief that he has arthritis in his thigh. He cannot, of course, have arthritis in his thigh, by the very definition of the term. Such a situation seems coherent. Next we are asked to imagine a counterfactual situation in which the same individual lives the same life as he did in the first situation, at least from his perspective.

Burge writes,

He has the same physiological history, the same diseases, the same internal physical occurrences. He goes through the same motions, engages in the same behavior, has the same sensory intake (physiologically described). His dispositions to respond to stimuli are explained in physical theory as the effects of the same proximate causes. All of this extends to his interactions with linguistic expressions. He says and hears the same words (word forms) at the same time he actually does. He develops the disposition to assent to "Arthritis can occur in the thigh" and "I have arthritis

in the thigh” as a result of the same physically described proximate causes.¹⁶

The key difference is that in the counterfactual situation “arthritis” means any rheumatoid ailment. If people in this situation were to describe the subject’s beliefs as being about arthritis, they would use content clauses (that-clauses for beliefs, for example) that involve the term “arthritis” (just as in the first situation). “Arthritis” means something different in the counterfactual situation. Therefore, the respective beliefs would differ while the intrinsic properties of the individuals would remain the same.

Here is another way to make the point. In both scenarios the subjects’ beliefs about their respective thigh pain would be characterized by those around them in the following way: “He believes that he has arthritis in his thigh.” In the first scenario, because of the linguistic meaning of “arthritis,” the subject believes that he has a rheumatoid ailment *exclusively of the joints* in his thigh. In the second scenario, because of the linguistic meaning of “arthritis,” the subject believes that he has a general rheumatoid ailment in his thigh. In short, in the actual case, the subject’s beliefs are about arthritis, and in the counterfactual case they are not about arthritis. What makes the beliefs differ is not anything intrinsic. It is stipulated that all the subjects’ intrinsic properties are the same. The properties of linguistic meaning are what have changed. These are extrinsic properties.

Burge’s argument for social externalism has received more skepticism than the arguments for natural kind externalism. A number of authors have noted that we do not attribute beliefs in such a rigid manner, by always taking a speaker’s at his word, as it

¹⁶ (Burge 1979), page 130.

were. Many have pointed out that our actual practice of thought attribution is much more resourceful and complex. Donald Davidson puts this point nicely:

First, it seems to me false that our intuitions speak strongly in favor of understanding and interpreting an agent's speech and thoughts in terms of what others would mean by the same words. For one thing, there is the problem of deciding what group is to determine the norms. But more important, we understand a speaker best when interpret him as he intended to be interpreted; this will explain his actions far better than if we suppose he means and thinks what someone else might mean and think who used the same words 'correctly'.¹⁷

Not a few philosophers have pointed out that in using Burge's method we sometimes end up attributing contradictory beliefs to rational beings, a point that we will focus on later. Burge's arthritis example nicely illustrates this. In the first scenario, because the agent states, "I believe that I have arthritis in his thigh," according to Burge, we should attribute to him the belief that he has a rheumatoid ailment exclusively of the joints in his thigh, the belief that he has an ailment in his thigh that cannot occur in his thigh. Surely, the subject also believes that he cannot have a disease in his thigh that cannot—by definition—occur in this thigh. The objection is that the form of belief attribution that Burge endorses conflicts with a principle of interpretation that says we should not attribute contradictory beliefs to a rational agent.

The details of these objections can wait. For now, I want to note why Burge's argument has been influential. I think that the major reason is that the conclusion can appear to be an inevitable consequence of adopting semantic externalism. In this essay, as I noted, I will understand this as the thesis that meaning is not a function of any one individual's psychological states. If we adopt such a view along with the view that the content of an individual's psychological states is only expressible by means of a

¹⁷ (Davidson 2001), pages 198-199.

public language, then will not content externalism follow? Though I think that the answer to this is no, part of the force of Burge's work lies in how content externalism seems to be an inevitable consequence of adopting semantic externalism along with the view that intentional contents are expressible only via a public language. Were such a consequence truly inevitable, we would have very good reasons to adopt content externalism, because it is very hard to see how linguistic meaning could be a function of any one individual's psychological states and because nothing other than a public language appears capable of express intentional content.

How many Externalisms?

In this section, I want to discuss two quite distinct views about intentional content that sometimes fall under the label of “externalism,” but do not always get as much press as natural kind externalism and social externalism. The first thesis has been called “singular anti-individualism,” and is supported by reflection on the nature of singular thoughts, thoughts that are of a particular individual. I shall call this view “singular thought externalism.” The second has been called “transcendental externalism,” and is supported by some key theses in the work of Donald Davidson.

Singular Thought Externalism

In *Anti-Individualism and Self-Knowledge* Jessica Brown gives the following definition of singular thought externalism: “According to singular, a subject’s thought contents are individuated partly by the particular objects that are in her environment (see, e.g., Perry 1979; Kripke 1980; Evans 1982; Peacocke 1983; McDowell 1986; Salmon 1986; Soames 1987; Kaplan 1989).”¹⁸ This thesis counts as a version of

¹⁸ (Brown 2004), page 13.

content externalism by our previous definition. It differs from natural kind externalism and social externalism, though, because it concerns thoughts about particulars, particulars which some have argued we do not think of by description, but by some other method. In the previously discussed situations, about water and arthritis respectively, the thinker in question could think of the relevant substances and phenomena in his environment by way of a description, for instance, "the watery stuff in my environment." Much of the literature on singular thought externalism is concerned with showing that there are objects that we cannot think of merely by description and exploring the consequences of these types of thoughts.

A good deal has been written about singular thoughts and singular propositions. My aim here is rather narrow: to understand how singular thoughts figure into an argument for content externalism. Here is the argument that Brown outlines. She asks us to imagine a thinker who has what she calls "a perceptually demonstrative thought" about an apple. For the subject, the apple is identified by demonstration. Imagine a counterfactual situation in which, familiarly, everything is the same from the inside, though there is a difference in the external environment: the apple is numerically different. The intuition is supposed to be that the thoughts are different in virtue of being about different apples. The thoughts, then, would be externalistic, since the intrinsic properties of the subjects are the same and the intuition is that the thoughts are different.¹⁹ The thoughts are object-dependent in that they are essentially identified by the objects they are about.

¹⁹ In what follows, I shall refer to thoughts as "externalistic" or "externalistically individuated" as a way to express their dependence on extrinsic properties of their thinkers. I shall likewise refer to thoughts as "internalistic" or "internalistically individuated" as a way to express these thoughts' dependence on only the intrinsic properties of their thinkers.

Transcendental Externalism

The other sort of externalism I want to discuss has been called "transcendental externalism" by Jason Bridges.²⁰ The thesis describes a type of externalism that Bridges identifies in the work of Donald Davidson. This form of externalism falls out of one of Davidson's most famous philosophical theses, the thesis of triangulation, which is a thesis about the necessary conditions for thought and language and about how thoughts can have the content they do. For Davidson, triangulation is necessary for the objectivity of thought because, in short, triangulation is necessary for creatures to have the concept of error—to have the idea that one's perspective on the world is one among others and that one's beliefs about the world may be false. The concept of error is required for a creature to have the concept of belief. The concept of belief is required for thought; therefore, without the concept of error, a creature would not be capable of thought. It is important to note that for Davidson it is not enough for thought that there be various *intrapersonal* perspectives (different ways of seeing things from the same individual's point of view) or that there be two nonlinguistic beings sharing an environment. There must be two linguistic beings (beings with a language) for individuals to have the concept of an objective world outside of their own individual perspectives on it.²¹ If this is true, then from the fact that there is thought at all we can know certain external facts about the world. This is, in a nutshell, Davidson's transcendental externalism. It counts as a form of externalism because it concerns thought's dependence on external properties, on the world outside of the thinker's mind being certain ways.

²⁰ See (Bridges 2006).

²¹ These individuals do not need to share the same language.

Davidson's argument for transcendental externalism that I have sketched above is a fascinating and interesting argument. My aim is not to criticize it here, but to merely lay it out to see how it relates to the previously mentioned forms of externalism. Let us begin with a discussion of triangulation, and show how this view anchors Davidson's transcendental externalism. Let us begin with one of Davidson's own summaries of the thesis.

There is a prelinguistic, precognitive situation which seems to me to constitute a necessary condition that can exist independent of thought and can therefore precede it . . . The basic situation is one that involves two or more creatures simultaneously in interaction with each other and with the world they share; it is what I call triangulation.²²

Like much of Davidson's work, this view is articulated against the backdrop of a systematic philosophical program that involves a number of significant and insightful philosophical theses. Davidson indicates that the thesis of triangulation is essential for explaining two features of thought. He writes,

The triangle that I have indicated is essential to the existence, and hence to the emergence, of thought. For without the triangle, there are two aspects of thought for which we cannot account. These two aspects are the objectivity of thought and the empirical content of the thoughts about the external world.²³

Triangulation must be true if our thoughts are about a shared, intersubjective world, and if our thoughts are about the features of the world we take them to be about. So, in addition to arguing that triangulation is a necessary condition for thought, Davidson also offers another transcendental argument for the conclusion that triangulation is required if our thoughts are about the normal objects we take them to be about. The idea here is

²² (Davidson 2001), page 128.

²³ (Davidson 2001), page 129.

that for thoughts to be about external, middle-sized objects the triangle of two linguistic beings and a shared environment must be in place. If this triangle were not in place, then the content of thoughts would be underdetermined. In short, if there were not two linguistic beings on the scene, then there would be no way to determine what thoughts were about. Of the necessity of triangulation for the determination of content Davidson writes,

For until the triangle is completed connecting two creatures, and each creature with common features of the world, there can be no answer to the question whether a creature, in discriminating between stimuli, is discriminating between stimuli at the sensory surface or somewhere further out, or further in. Without this sharing of reactions to common stimuli, thought and speech would have no particular content—that is, no content at all. It takes two points of view to give location to the cause of a thought, and thus to define its content.²⁴

The idea appears to be that without two creatures who speak a language and who share and interact with the same objective world, there would be no way to decide what their thoughts were of. If there were only one creature on the scene, there would be no fact of the matter about what his thoughts were of. His thoughts could be about anything in the causal chain of events that caused his thought, from the external objects in his visual field to the sensory impingements on his retina, for example. It takes two linguistic beings and a share environment to fix such content.

As we saw, Davidson's transcendental externalism guarantees that if there is thought, then certain external properties must be instantiated. We have also seen that two sorts of externalism fall out of Davidson's thesis of triangulation. The first is transcendental externalism: the idea that triangulation is necessary for thought. The second is Davidson's view that triangulation is necessary if content is not

²⁴ (Davidson 2001), pages 212-213.

underdetermined. How do these views relate to our general definition of content externalism?

The most important thing to see for our purposes is that Davidson's view is not concerned with what determines the nature of content, but the conditions that make it possible for intentional states to have content. One way to put the point is this. The content externalist is interested in what is responsible for determining the content of each individual intentional state, whereas Davidson is concerned with the general necessary conditions on creatures having intentional states with content. As we have seen, for Davidson, the content of intentional states cannot occur in isolation, the conditions that must obtain require the instantiation of extrinsic properties, whereas for Putnam and Burge the focus is on the individuation of content. Thus, we must add a new thesis to articulate a Davidsonian view concerning the necessary conditions of thought.

The thesis of transcendental externalism: The existence of the contents of intentional states require the instantiation of background conditions, conditions which instantiate extrinsic properties.

Davidson's focus is on the background conditions of intentional content, while the other externalisms focus on the individuation of content—whether or not a certain type of thought (and not thought in general) requires the environment to be a certain way. In Chapter 4 we will examine the extent to which these both content externalism and transcendental externalism are in conflict with first-person access to intentional content, and hence in conflict with rationality.

CHAPTER 4 QUESTIONS OF COMPATIBILITY

In Chapter 2, I discussed the thesis of first-person access to intentional content and its importance to rationality. I will now discuss which forms of externalism are compatible with first-person access to intentional content. Because this thesis is so central to the picture of self-knowledge sketched in Chapter 2, forms of externalism that are incompatible with it will be incompatible with the picture of self-knowledge sketched in Chapter 2. Much has been written about the compatibility of content externalism and self-knowledge. Instead of beginning with a history of the debate, I will give my own argument, and then address how my argument relates to the larger body of literature.

The Incompatibility of Social Externalism and First-Person Access to Intentional Content

Let us begin by discussing the compatibility of social externalism and first-person access to intentional content. Brief reflection on Burge's patient shows that there are problems for content externalism and rationality. Recall Burge's patient who believes that he has arthritis in his thigh. According to the externalist theory of content determination, and due to the linguistic meaning of "arthritis" in his community, we should attribute to the patient the belief that he has arthritis in his thigh, or, in other words, the belief that he has an ailment that cannot occur in his thigh in his thigh. The theory, then, has us attributing a belief to him that is not one that a rational agent should hold. A rational agent may surely have false beliefs, but what he cannot have is blatantly contradictory beliefs. What has gone wrong?

It seems that the most obvious way to state the problem is that the externalist method of content attribution is not sensitive enough to how the patient thinks of arthritis, not sensitive enough to his personal concept of arthritis. Both parties, content

internalists and content externalists alike, must look to certain features to determine what someone thinks. The social externalist focuses on the linguistic meaning of the terms in the belief ascription. The content internalist, on the other hand, should look to the concepts a subject associates with linguistic terms (concepts which he must of course say depend on only intrinsic properties). Reflecting on the above case, it seems that content internalism, primarily because of its sensitivity to how the patient thinks of arthritis, is the theory of content determination that is consistent with rationality, and should for that reason be preferred.

What does this have to do with first-person access to intentional content? So far it may appear that the incompatibility is directly between rationality and content externalism. We know from Chapter 2 that there is a connection between rationality and first-person access to intentional content. One of my own arguments concluded that first-person access to intentional content is required for rationality. So, if content externalism runs afoul of rationality, it is liable to do so by first running afoul of first-person access to intentional content.

Focusing on Burge's thought experiment and the type of access we must have to our concepts helps us to sharpen the argument. That we have first-person access to our concepts is entailed by the thesis of first-person access to intentional content and the view that content is structured by concepts. Burge's patient has first-person access to how he thinks about arthritis. He does not, however, have first-person access to the linguistic meaning of "arthritis" in the community that he is in. If one sees this rather obvious point, then one is in a position to see how content externalism is incompatible with first-person access to intentional content. If content externalism is true, then the

individual's access to his intentional content is the same as his access to the linguistic meaning of arthritis. His access to linguistic meaning is the same as his neighbors'. Therefore, his access to the content of his own thoughts is the same as his neighbors' because his access runs by way of an extrinsic property to which they both have equal access.

That, in short, is the argument for incompatibilism. I need to ward off a few objections before I explain how this argument, set against social externalism, generalizes to all forms of content externalism. First, we need to see that the argument above does not require the assumption that one must know all the background conditions of content in order to know the content of one's thought. Any such argument for incompatibilism that required this assumption would have a much too demanding view of what is required for knowledge. For example, as Burge pointed out, it would be too much of a demand to place on knowledge of X to say that to know X one must also know the background conditions on which X depends.¹ We cannot therefore argue that content externalism is incompatible with first-person access to intentional content on the grounds that a subject must know the background conditions that are required for a particular sort of thought. Like Burge, Boghossian also notes that such an argument makes problematic assumptions about knowledge.

As it stands it appears to be making problematic assumptions about the conditions required for knowledge. Consider perceptual knowledge. Someone may know, by looking, that he has a dime in his hand. But it is controversial, to put it mildly, whether he needs to know all the conditions that make such knowledge possible. He need not have checked, for example, that there is no counterfeit money in the vicinity, nor does he need to be able to tell the difference between a genuine dime and every imaginable counterfeit that could have been substituted. The ordinary

¹ See (Burge 1988)

concept of knowledge appears to call for no more than the exclusion of “relevant” alternative hypotheses (however exactly that is to be understood); and mere logical possibility does not confer such relevance.²

Given these warnings, we should be careful that the present argument does not make these unwarranted assumptions. I did not assume that to know that one, for example, believes that p, one must know the background conditions on which one's belief that p depends. Understood in an unrestricted way, this gives us comical results; to know that one believes that p, one must gain all sort of information and settle all sorts of question in order to know one's beliefs. For example, one would have to know whether or not the belief in question supervened on a brain state, and if so, what the nature of the brain state was in order to know the content of the belief.

Though my argument does not employ this assumption, it can seem to employ a similarly suspect assumption: the assumption that to know the content of a thought one must know the content determining properties of that thought. After all, it seems that the individual must know the meaning of "arthritis" to truly know his thought. The linguistic meaning is partly responsible for determining the thought, so to know the thought content one must know the content determining property. This assumption can seem suspiciously similar to the one above; it merely substitutes the requirement to know the background conditions with the requirement to know the content determining properties. We need to get clear on what those properties are, of course. If we just understand "content determining properties" to mean the properties responsible for the content in an unrestricted sense we have some familiar comic results. For instance, if one accepts physicalism about mental states, then it appears that (at a minimum) intentional states

² (Boghossian 2008), pages 146-147.

supervene on physical states. Are the physical properties of the intentional states content determining properties? If so, it appears that one must know the physical properties of one's intentional states to know one's thoughts. If we understand the assumption that to know the content of a thought one must know the content determining properties of that thought in an unrestricted way, then it appears that both the content internalist and content externalist have a problem when it comes to first-person access to intentional content. Both views must hold that there are properties that determine thought content. If those properties are known via other properties, then it appears that first-person access to intentional content is lost.

An incompatibilist might try to save things by arguing that we should understand the assumption in a more restricted way. We should hold that one must know only the *relevant* content determining properties to know the content of his thoughts. This of course raises the question of what the content determining properties are for both views. We have seen some of the candidates for content externalism (e.g., linguistic meaning, microstructural properties of substances), but we have not yet examined any candidates for content internalism. We will hear more on the sorts of intrinsic properties that may be responsible for intentional content in Chapter 6. As for the present maneuver, restricting the principle to just the *relevant* content determining properties seems *ad hoc*.

We do not actually have to make such a maneuver, thankfully, as the argument I made above does not depend on this assumption. The first step in seeing this requires seeing that my argument is one about access and not knowledge. These notions, as we saw from Chapter 2, are closely related, but the basic incompatibility is between content

externalism and first-person access to intentional content. Burge's patient lacks first-person access to the content of his belief, because he must first access the linguistic meaning of "arthritis." He has first-person access to the concept he associates with the meaning of "arthritis," his personal concept of arthritis.³ Content internalism is compatible with first-person access to intentional content because it looks at how the individual thinks of arthritis to determine the content of the individual's thought. In other words, in the example above, it looks to the concept the person associates with "arthritis" and not the intersubjective concept associated with "arthritis." The point is not that one theorist's content determining properties cause problems for first-person access to intentional content while the other theorist's do not, for being aware of one's personal concepts is not achieved via grasping some content-determining property first and grasping the concept second in virtue of the property grasped. Accessing one's personal concepts is immediate. The key point to see is that if content externalism is true, then access to intentional content is not merely a matter of a subject accessing his personal concepts. On content externalism, the determiners of content are not a subject's personal concepts, but other things, things which do not supervene solely on a thinker's intrinsic properties.⁴ For the social externalist, the determiners of content are the linguistic meanings of content clauses in belief ascriptions. These meanings, as Burge's case clearly brings out, are not something that one can have first-person access to.

³ I sometimes refer to subjective concepts as personal concepts.

⁴ I assume of course that personal concepts supervene on a thinker's intrinsic properties.

Burge's case allows us to easily see the connection with rationality. The content externalist has it that Burge's patient believes that he has arthritis in his thigh, which is the belief that he has an ailment in his thigh which cannot occur in his thigh. Surely he also believes that he cannot have a pain in his thigh which cannot—by definition—occur in his thigh. Asked to give an explanation for why he thought that he could have an ailment that cannot occur in his thigh in his thigh, our individual would surely say that he did not realize that arthritis is an ailment exclusively of the joints. He would likely say that he knew that he had a pain in his thigh and mistakenly called it "arthritis" because it felt similar to other arthritic pains he experienced previously. The belief he reveals in the process of justifying his own beliefs is the belief that he has an arthritis-like ailment in his thigh. This is *not* the belief that he had arthritis in his thigh. We can see, then, that the intentional content of the belief that becomes his reason in rational discourse is a content that is shaped by the personal concept he associates with the term "arthritis" and not one that is shaped by the intersubjective concept associated with the term "arthritis."

The Incompatibility of all forms of Content Externalism and First-Person Access to Intentional Content

All other forms of content externalism (natural kind externalism and singular thought externalism) are incompatible with first-person access to intentional content for the reason discussed above. The reason here is that what determines a subject's intentional content is not his personal concepts (that the content internalists says supervene on intrinsic properties) but something else, whether it be the linguistic meanings of terms in intentional state ascriptions, the microstructural properties of substances, or particular objects in the external world. We can see this in the cases of

natural kind externalism and singular thought externalism by using switching cases borrowed from the rather large body of incompatibilist literature. The idea of switching was first introduced by Burge.⁵ Basically, the switching scenarios are scenarios where an individual is switched between Earth and Twin Earth, and does not realize that he is being switched. His life seems the same to him at each place. There is an externalist intuition that the intentional contents of the switcher's thoughts would change after enough time has passed in a given environment. Burge writes, "The thoughts would not switch as one is switched from one actual situation to another twin actual situation. The thoughts would switch only if one remained long enough in the other situation to establish environmental relations necessary for new thoughts" (1988: 114).⁶ So, the content externalist thinks that it is clearly possible that even though a subject is unaware of the switch, some of his intentional contents would still change.

What is the problem with this? Much has been made of switching cases, which will be addressed later in this chapter. For now, I should stress that I use switching because it is familiar and expresses the content externalist's commitments rather clearly. Let us suppose that the subject originally from Earth has been on Twin Earth for the requisite period of time. Intentional states that seem to him to be about water are actually about twater, because there is no longer any H₂O in his environment, and assuming we accept the story about *a posteriori* necessities, only H₂O is water.⁷ He now

⁵ See (Burge 1988).

⁶ I must admit that I fail to see how a certain amount of time would have to pass for the new environment to take hold, as it were, of the intentional contents. One would think that the shift would be immediate, since it is not dependent on the subject's awareness of the switch.

⁷ "Twater" is a term that is sometime uses to refer to the watery substance on Twin Earth. It use suggests that the watery stuff on Twin Earth is not actually water. I do not mean to endorse such a view by using the term. I employ it for ease of exposition.

desires to drink twater and believes that the lakes, rivers and oceans are filled with twater. In my discussion of the switching case, I assume the truth of the story of a *posteriori* necessities. (Recall that I argued earlier that this story was integral to supporting natural kind externalism.) Suppose the switcher is in a discussion with a Twin Earthian about the importance of improving access to clean and potable water in the developing world. He says things like, "Water is a common resource that all people should be entitled to access," "Governments have a responsibility to provide their citizens with access to clean and potable water," and so on. He intends to be speaking of the liquid that surrounds him. His Twin Earthian interlocutor, based on his assertions, will attribute to him beliefs about twater and he will thereby attribute to him the concept of twater. When the switcher says, "water is a common resource that all people should be entitled to access," the content of his belief on Twin Earth can be expressed variously: that twater is a common resource that all people should be entitled to access; that XYZ is a common resource that all people should be entitled to access; that XYZ and not H₂O is a common resource that all people should be entitled to access; that twater and not water is a common resource that all people should be entitled to access. These all express the content of his belief, for the content is cashed out relative to the story of a *posteriori* necessities about what determines the referent of "water" at Twin Earth, and the content externalist's claim that the subject's concepts will be about twater after a certain amount of time.

Let us slightly alter the thought experiment. Suppose that the Twin Earthian interlocutor is aware of the switching. He knows that the individual does not realize that the substance around him is XYZ. Let us suppose that he tells the switcher of this fact.

Let us suppose that he also asks the switcher "Do you really believe that people are not entitled to access water as they are entitled to access twater?" It is easy to imagine the switcher responding incredulously. For throughout the discussion, he had been taken himself to be talking about water and expressing beliefs about water. Even if he were not talking about water, surely he would have taken himself to be expressing beliefs about water. According to the content externalist, though, he was not. And he is mistaken to think that he was thinking about water. His mistake, though, is different from Burge's patient's mistake in at least one sense. We can hold Burge's patient culpable for his mistake in thinking he had arthritis in his thigh, whereas the switcher is ignorant of the meaning of "water" due to the set-up of the thought experiment, and cannot really be held at fault for believing that he believed that twater and not water is a common resource that all are entitled to access. However, this difference is a minor one, and not central to our discussion.

The key similarity is that the content of both the patient and the switcher's intentional states is hostage to the public meaning of the terms of ascriptions used in the respective intentional state attributions. Thus, they have first-person access to the content of their intentional states in so far as they have first-person access to the public meaning of terms—which is to say not at all. If we adopt the externalist method of intentional content attribution, we will attribute intentional content to subjects that is not sensitive to their personal concepts, in a way, in other words, that is in conflict with their rational agency. As we saw in Chapter 2, for a subject to be rational, he must access his intentional content in first-person way. According to content externalism, the patient and the switcher's access to the content of their intentional states goes by way of their

access to the public meanings of the terms of ascriptions of intentional state attributions. Their access is not immediate and direct, and unless the access is immediate and direct, the access is in conflict with rationality. For this reason, content externalism is incompatible with first-person access to intentional content and thereby in conflict with rational agency.

An Argument for the Compatibility of Transcendental Externalism and First-Person Access to Intentional Content

Let us turn to the compatibility first-person access to intentional content and transcendental externalism. Here are the two theses in question.

- **The thesis of transcendental externalism:** The existence of the contents of intentional states require the instantiation of background conditions, conditions which instantiate extrinsic properties.
- **The thesis of first-person access to intentional content:** Each subject has first-person access to his own thoughts because each subject accesses his contents non-inferentially and without the benefit of evidence, whereas others access those contents by way of inference.

In assessing whether or not these claims are compatible, we need to recall that we already rejected the requirement that to know *X*, one must know the background conditions on which *X* depends. So, it should be easy to see that we are not in a position to require that to know what one was thinking, one would have to inquire into the background conditions necessary for a given thought. One does not have first-person access to the background conditions necessary for thought. When we examine the two claims above, it does not appear that they are in conflict. The reason for this should be clear: because transcendental externalism is not a thesis about the individuation of the content of intentional states, but about background conditions

necessary for certain intentional states. It is well worth keeping in mind that one can hold transcendental externalism while rejecting content externalism.

Addressing the Larger Debate

It would be insufficient to end here, as it would neglect what many others have written about the incompatibility of content externalism and self-knowledge. Many of the issues in this debate can be traced to three main sources. First, there are issues that can be traced back to Boghossian's incompatibilist argument in "Content and Self-Knowledge," namely issues that have surrounded switching cases, with which we are already somewhat familiar. Second, there is the incompatibilist argument due to McKinsey in "Anti-Individualism and Privileged Access." Third, the internalism and externalism debate about justification has seemed relevant to many. Let me begin by discussing Boghossian's argument in "Content and Self-Knowledge," and the issues it created, and then move to a discussion of internalism and externalism about justification and also McKinsey's argument.

Boghossian's Incompatibilist Argument

Boghossian uses the switching scenario introduced by Tyler Burge to illustrate how content externalism is incompatible with self-knowledge. Boghossian asks us to imagine that Twin Earth exists and that there is an individual switched unawares between Earth and Twin Earth. The content externalist thinks that after the appropriate amount of time, the individual's concepts will change in virtue of the environmental change even though there will be no introspectible difference in his intentional states. Boghossian notes that if the individual is switched back between Earth and Twin Earth, it appears to be an open question whether the switcher will have two sets of concepts, an Earthian and a Twin Earthian set, or just one set of concepts which are appropriate

to whatever environment is currently shaping those concepts. Boghossian notes that as the story is typically told the individual has only one set at a time, the set appropriate to the given environment. Boghossian employs the popular retelling, the one-concept version, in his own argument.

Boghossian presses the content externalist by focusing on how the switcher would answer questions about the content of his intentional states.

If someone were to ask him, just after one set of twin-earthian concepts has been displaced by a set of earthian ones, whether he has recently thought thoughts involving an arthritis-like concept distinct from *arthritis*, S would presumably say "no." And yet, of course, according to the anti-individualist story, he has. His knowledge of his own past thoughts seems very poor, but not presumably because he simply can't *remember* them. Could it be because he never knew them?⁸

Boghossian then argues that to know what he thinks the switcher must be able to rule out relevant alternatives. A thought with the concept of tharthritis is a relevant alternative, Boghossian argues, to the switcher's present thought that involves the concept of arthritis. Boghossian's point is that content externalism raises the possibility of relevant alternatives to thoughts that is not raised by content internalism. Boghossian argues, then, that to know his thoughts the switcher must not only consider—but also rule out—relevant alternatives to know what he thinks. This means, of course, that he must reason to conclusions about the contents of his thoughts. This is surely incompatible with the immediacy and directness of self-knowledge. If we know our thoughts by reasoning, we do not have self-knowledge of them.

⁸ (Boghossian 2008), page 147.

Challenges to Boghossian's argument

Ted Warfield argues that Boghossian's argument is invalid.⁹ Here is his characterization of Boghossian's argument ("S" stands for the subject of the slow switching, and "P" stands for the proposition in question). Let us call this "Version of 1 of the Slow Switching Argument."

Version 1 of the Slow Switching Argument

- (P1) To know that P by introspection, S must be able to introspectively discriminate P from all relevant alternatives of P.
- (P2) S cannot introspectively discriminate water thought from twater thoughts.
- (P3) If the Switching Case is actual, then twater thoughts are relevant alternatives of water thoughts.
- (C1) S doesn't know that P by introspection.

Warfield argues that all that follows is a much weaker conclusion.

- (C1) If the Switching Case is actual, then S doesn't know that P by introspection.

If this is the conclusion, Warfield argues, then Boghossian has not demonstrated that self-knowledge and content externalism are incompatible. At most he has shown that it is not necessarily true that we have self-knowledge of our thoughts if content externalism is true.

Peter Ludlow argues that Warfield mischaracterizes Boghossian's argument.¹⁰ He argues that (P3) should be cast not in terms of whether the switching case is actual but in terms of whether such cases are prevalent. Here is Ludlow's characterization of Boghossian's argument. Let us call this "Version 2."

Version 2 of the Slow Switching Argument

⁹ See (Warfield 1992).

¹⁰ See (Ludlow 1995).

- (P1) To know that P by introspection, S must be able to introspectively discriminate P from all relevant alternatives of P.
- (P2) S cannot introspectively discriminate water thought from twater thoughts.
- (P3') If switching cases in general are prevalent, then there are relevant alternatives to water thoughts.¹¹
- (P4) Switching cases, in general, are prevalent.
- (C1) S doesn't know that P by introspection.

Ludlow argues that multiple meaning words are common, and that speakers shifting from one community to another will often find themselves using a word to express a thought that has a different meaning in different communities. Here is an example that Ludlow gives of this phenomenon.

Suppose for example that Biff, who knows little of American philosophy, understands that certain of his colleagues are called 'pragmatists'. He may speak of these colleagues at home as pragmatists as well, not knowing that the expression has different meanings in the two worlds that he moves between. During a conversation with one group or another, Biff may even wonder whether a particular member of the group is a pragmatist. Of course the content of the mental episode will depend on Biff's fellow travelers of the moment (assuming he has spent enough time with them). Once again, Biff is an unwitting victim of slow switching.¹²

The prevalence of these sorts of cases is supposed to motivate P3' in Version 2. Ludlow seems to me correct on this score: situations wherein the content of a thought changes due to an environmental shift and without any awareness on the subject's part are quite common if content externalism is true.

¹¹ In this essay, certain premises in arguments may have an inverted comma attached to them, as in (P3') above. In this case, this indicates that (P3') is an iteration of (P3). Attached to premises, the inverted comma should be taken to have the same connotation through this essay.

¹² (Ludlow 1995), page 228.

Warfield has an interesting response to Ludlow's defense of Boghossian's argument. He argues as follows.

I fail to see how Ludlow thinks he has improved upon Boghossian's fallacious argument. The most that follows even granting all of Ludlow's claims is that there is a possible world, the actual world, in which externalism is true and individuals lack privileged self-knowledge of some of their thought contents. This, like Boghossian's argument, shows at most that externalism is *consistent with* a lack of self-knowledge. Ludlow's claim do nothing towards showing that externalism *implies* a lack of self-knowledge.¹³

I take it that Boghossian thinks that it is not just in the actual world that self-knowledge and content externalism are incompatible; his argument aims to show that in any world in which content externalism is true, individuals do not have direct self-knowledge of the content of their thoughts, due to the very nature of content externalism. Content externalism itself does not logically imply or entail the denial of self-knowledge. Rather, it is in conjunction with other plausible claims or premises (P1, P2, P3' and P4 in Version 2) that the truth of content externalism leads to an absurd consequence. We can understand this as a *reductio ad absurdum* argument in favor of content internalism. This is more or less the response that Ludlow has for Warfield's rejoinder.¹⁴ There is no reason to think that self-knowledge and content externalism are compatible in some worlds but not compatible in others. Boghossian's point is that the nature of content externalism makes it such that individuals lack self-knowledge of some of their thought contents.

¹³ (Ludlow 1995), page 233.

¹⁴ See (Ludlow 1997).

How to understand the importance of Boghossian's argument

There may be some lingering reservations about the soundness of Boghossian's argument. For instance, on the face of it, (P1) seems to make problematic assumptions about what is required for self-knowledge.

(P1) To know that P by introspection, S must be able to introspectively discriminate P from all relevant alternatives of P.

In Chapter 2 we discussed how the directness of self-knowledge requires that we know our own thoughts without the benefit of evidence. As it is written, (P1) may seem to require that subjects use introspective evidence to arrive at self-knowledge.

At this point, it is important to note that Boghossian himself does not use the term "introspection" when he characterizes his argument. Here is a revealing passage.

What does S know? By assumption, he is not aware that the switches have taken place and nothing about his qualitative mental life or his perceived environment tips him off. Indeed, S may not even be aware of the existence of twin-earth or of the dependence of content on environment. As far as S is concerned, he has always lived on earth.¹⁵

We should be a bit cautious about using the term "introspection," then, in a proper characterization of Boghossian's argument. The key to understanding Boghossian's argument is seeing that if content externalism is true, then there will be relevant alternatives to certain thoughts that a subject must rule out to know what he thinks. Let us recast the argument as follows, without the use of the term "introspection." Let us call this "Version 3."

Version 3 of the Slow Switching Argument

(P1) If there are relevant alternatives to p, then S must rule out those alternatives to know p.

¹⁵ (Boghossian 2008), page 148..

- (P2) If content externalism is true, then there are relevant alternatives to certain thought contents (such as the belief that water is vital to life and the belief that twater is vital to life).
- (P3) S must rule out relevant alternatives to have knowledge of certain thoughts.
- (P4) Ruling out relevant alternatives involves reasoning.
- (C1) Knowing our thoughts by reasoning to them is not sufficient for the directness required for self-knowledge.

The support for (P3) and (P4) comes from the following passage.

S has to be able to exclude the possibility that his thought involved the concept *arthrititis* rather than the concept *tharthrititis*, before he can be said to know what his thought is. But this means that he has to *reason* his way to a conclusion about his thought; and reason to it, moreover, from evidence about his external environment which, by assumption, he does not possess. How, then, can he know his thought at all?— much less know it directly?¹⁶

I want to draw attention to these two questions that Boghossian asks here. They seem to suggest that there are two possible conclusions we can draw from this argument.

(C1) from Version 3 of the Slow Switching Argument is related to the question of how we can know our thoughts directly if we must reason to their contents. This is one way to understand the argument that Boghossian gives. It does not seem to me, though, to be the best way to understand the argument. Let me explain.

The primary question seems to be the question of how the subject can know his thoughts *at all* if content externalism is true. The basic idea here is not that the subject must reason to his thought content by ruling out relevant alternatives but that he is *in no position to rule out relevant alternatives* based on how the situation is set up. Content externalism creates the possibility of relevant alternatives to certain thought contents, and such alternatives cannot even be considered, let alone rule out, in the slow

¹⁶ (Boghossian 2008), page 148.

switching case. The switcher, remember, does not have any information about the switching or about the difference between water and twater, so he does not have enough information to conceive of relevant alternatives to thoughts he takes to be about water. We can characterize this argument as follows. Let us call this "Version 4."

Version 4 of the Slow Switching Argument

- (P1) If there are relevant alternatives to p, then S must rule out those alternatives to know p.
- (P2) If content externalism is true, then there are relevant alternatives to certain thought contents (such as the belief that water is vital to life and the belief that twater is vital to life).
- (P3) S is not capable of ruling out the relevant alternatives to certain thoughts.
- (P4) S cannot know the content of certain thoughts because S is not in a position to rule out relevant alternatives.
- (C1) Therefore, if content externalism is true, S cannot know the content of certain thoughts.

This argument should be seen in light of the primary question (How can the switcher know his thought at all?), because if the subject is not in a position to conceive of relevant alternatives, then the point that he must reason to the content of his thoughts by ruling out relevant alternatives does not get a foothold. It is fine, of course, to make the point that *even if* the subject could conceive of relevant alternatives to certain thought contents, then to know those thought contents he must reason to those by ruling out those relevant alternatives. This point would be relevant if it turned out that the subject could in fact conceive of relevant alternatives. However, it is clear from the way the switching case is described that the subject cannot conceive of the relevant alternatives, as he has no information about the switch or of the underlying chemical structure of the water-like substance in the environment.

Of course, all the versions of the slow switching argument discussed above depend on some relatively controversial premises, mainly the connection between discriminating between relevant alternatives and knowledge. Aside from being somewhat controversial, this connection between discrimination, relevant alternatives and knowledge, though interesting in its own right, is orthogonal to question of the compatibility between self-knowledge and content externalism. The argument should be understood without appeal to discrimination and relevant alternatives, so it turns out that the focus on discrimination and relevant alternatives in the incompatibilist literature is a red herring.

Because it is controversial that S's knowing p requires S to discriminate p from relevant alternatives to p, we should be wary of using that premise in our argument. I think that we can use the switching case to support the following sort of argument, retaining the spirit of Boghossian's argument, while freeing ourselves of the questionable commitment to the connection between discrimination and knowledge. What is required is translating the talk of relevant alternatives into talk about personal and intersubjective concepts. Let us call this "Version 5 of the Slow Switching Argument."

Version 5 of the Slow Switching Argument

- (P1) There are cases in which S's personal concept of X can stay the same even if the intersubjective concept of X changes.
- (P2) If S's thought contents about X are shaped by the intersubjective concept of X, in other words, if content externalism is true, then there are cases where S's thought will change without a change in personal concept.
- (P3) The change in thought content will be due to a factor to which the subject does not have first-person access.

(C1) If content externalism is true, subjects lack first-person access to certain thought contents.

The truth of (P1) and (P2) seen by reflecting on the switching cases. I think this talk of personal and intersubjective concepts is sufficiently faithful to the spirit of Boghossian's argument. Consider a previously quoted passage from his paper about the nature of the subject's concepts.

If someone were to ask him, just after one set of twin-earthian concepts has been displaced by a set of earthian ones, whether he has recently thought thoughts involving an arthritis-like concept distinct from *arthritis*, S would presumably say "no." And yet, of course, according to the anti-individualist story, he has. His knowledge of his own past thoughts seems very poor, but not presumably because he simply can't *remember* them. Could it be because he never knew them?

Asking the subject this sort of question, a question about his concepts, is the right one to ask here. It brings out rather clearly that there has been no change in how he thinks about water. We can grant that the intersubjective concept of water has changed such that this concept picks out XYZ and nothing else. However, the subject would not describe his own concept of water as exclusively picking out XYZ. If we grant that there is a distinction between personal and intersubjective concepts, then there is no reason to resist this sort of argument.¹⁷ Properly articulated, slow switching can be used to show the fatal flaw of content externalism: that the content of thoughts is determined via the intersubjective concepts associated with terms. These are not concepts a subject has direct awareness of or first-person access to. Therefore, there are cases in which a subject's access to a content of a thought will not be of a first-person sort, but rather the

¹⁷ I offer some considerations in Chapter 5 that lend support to the distinction between subjective and intersubjective concepts.

same as everyone else's. I should also add here that intersubjective concepts instantiate extrinsic properties while personal concepts, properly understood, do not.

To resist such an argument it will not do to argue, as Warfield does, that the switching case is not a relevant alternative. Ludlow's point is that there are many times when a subject switches between linguistic communities shows that there are many actual cases where the concept a subject associates with a term and the intersubjective concept associated with that term will differ. "Football" is an English word that provides a ready example. Suppose there is an American who does not know that the British call soccer "football." As far as he knows, "football" refers only to American football. When vacationing in England, he visits a pub and starts discussing sports with the pub's British clientele. Suppose he says, "Football is the greatest sport on Earth." According to the content externalist, his thought is that soccer is the greatest sport on Earth, even though he has no idea that soccer is called "football" in Britain. He intends to assert, of course, that American football is the greatest sport on Earth, as he believes that American football is the greatest sport on Earth. This is a much more homey example than the switching case that involves Earth and Twin Earth. And homey examples are plentiful; there are many cases where the concept a subject associates with a term differs from the intersubjective concept associated with that term. Thus, if content externalism is true, there are many cases where subjects do not have first-person access to their thoughts.

Other Incompatibilist Arguments

Phenomenal Properties and Privileged, First-Person Access to Intentional Content

As for other incompatibilist arguments, I want to begin by discussing an argument that can seem similar to my own argument for incompatibilism. Distinguishing this argument from my own will help to point out certain virtues of my own argument.

In the switching case, the switcher's qualitative mental life stays the same while his thought contents switch. There are many who think that one's phenomenal properties are the properties to which one has first-person access. There is a temptation, then, to offer the following sort of argument. If a subject has privileged access to his intentional contents it is by way of privileged access to phenomenal properties. The phenomenal properties of the switcher stay the same in all environments even though there is a change in intentional content in different environments. The environmental properties that are responsible for these changes are not properties a subject has first-person access to. Therefore, thought contents are not always contents a subject has privileged access to. This sort of argument uses the switching case as a backdrop to make the point.

The problem with this argument is that it makes heady assumptions about the dependence of intentional content of phenomenal properties.¹⁸ We would need to have an independent argument for that dependence to properly support the incompatibilist argument sketched in this section. I discuss these sorts of arguments in Chapter 6, but nothing in my own argument requires such dependence of intentional content on

¹⁸ This argument also assumes that phenomenal properties supervene on intrinsic properties (phenomenal internalism) rather than extrinsic properties (phenomenal externalism).

phenomenal properties. What is required is that intentional content is to be cashed out in terms of personal concepts and not intersubjective concepts if subjects are to have first-person access to their intentional contents. No specific relationship between personal concepts and phenomenal properties is required (although the content internalist should hold that personal concepts supervene on just intrinsic properties). Of course, there may be some connection between personal concepts and phenomenal properties yet to be spelled out. However, my own argument does not require an articulation of that connection, if there is one. The argument I make is independent of phenomenal properties, and more directly gets to the heart of the incompatibility of content externalism and first-person access to intentional content. It may turn out that the above argument is also an argument for incompatibilism, but if it is, it will be so only in a derivative way, based on some grounding of personal concepts in phenomenal properties.

Content Externalism and *A Priori* Access to the External World

In addition to the slow switching style incompatibilist arguments I have been discussing, there is another prominent incompatibilist form of argument, first introduced by McKinsey in "Anti-Individualism and Privileged Access." McKinsey argued that if content externalism is true and privileged access is true, then we can gain *a priori* knowledge of the empirical world, which is absurd. To avoid this absurdity, we must reject either content externalism or privileged access. The argument is roughly as follows. If content externalism is true, then certain thoughts have their contents only if the environment is a certain way. If privileged access is true, then we know the contents of our thoughts in an *a priori* fashion, without, that is, the benefit of empirical investigation. However, this appears to entail that in virtue of the *a priori* access we

have to our thoughts and the environmental dependence of their content, we can know empirical facts based purely on *a priori* investigation.

Much has been written about this argument. I will not address this debate here. An argument for incompatibilism has already been advanced which I think gets to the heart of the incompatibility more straightforwardly than the McKinsey style arguments. And we have already discussed some of the difficulty of assimilating talk of privileged access or self-knowledge to talk of *a priori* knowledge, namely that reducing or assimilating the concept of privileged access to *a priori* knowledge seems to reverse the proper order of the reduction. We appear to have a much clearer grasp on the concept of first-person access than the concept of *a priori* knowledge. So, that point alone should give us some reason to suspect that the above argument is not as clean and straightforward as it first appears. The focus of this project has been on first-person access to intentional content, both its necessity for rational agency and its incompatibility with content externalism. Though some version of McKinsey's argument may be sound, we need not seek such a version here, as we have already seen an argument that shows that first-person access to intentional content is incompatible with content externalism.

Internalism and Externalism about Justification

So far we have discussed externalism as it relates to language (semantic externalism), to intentionality (content externalism), and to experience (phenomenal externalism). Externalism is often invoked in epistemology as well. We shall now examine the extent to which externalism in epistemology bears on incompatibilism.

A good deal has been written about externalism and internalism in epistemology. Traditionally, philosophers have tended to be epistemic internalists, understanding

knowledge as true, justified belief of a proposition, with justification understood as somehow relative to the believer's perspective. Justification has traditionally turned on a subject's personal evidence or reasons for the proposition in question. More recently, though, due to the work of reliabilists, philosophers began to question epistemic internalism. Perhaps justification is a more objective matter, the thought runs. A subject may have knowledge of a proposition if such knowledge is the result of a reliable method, even if that method is not something of which the subject is aware. It is best to see the internalism and externalism debate in epistemology as one about justification. So understood, we can define the general positions as follows.

The thesis of epistemic externalism: The justification for one's belief that p is not solely a function of one's intrinsic properties

The thesis of epistemic internalism: The justification for a one's belief that p is solely a function of one's intrinsic properties.

Before moving further, it is worth remembering that my view of how each person knows his own mind is one that endorses epistemic externalism. The justification for one's second-order beliefs about first-order intentional states does not supervene solely on one's intrinsic properties. Rather, it depends on external factors.

In "Externalism and Mind and Epistemology" Jessica Brown raises some worries about the compatibility of any plausible form of epistemic internalism with content externalism. She quite usefully outlines three versions of epistemic internalism and then measures their compatibility with content externalism.

Supervenience of the mental (SM): whether a thinker is justified in believing p supervenes on that thinker's occurrent and dispositional mental states.

Supervenience of the accessible (SA): whether a thinker is justified in believing p supervenes on those states to which she has special access.

Access internalism (AI): One has special access to one's justificatory states.¹⁹

Brown argues that SM is the only version that is potentially compatible with content externalism, and that there are problems with SM. If her arguments are successful, it looks like epistemic internalism is incompatible with content externalism. The issue of the compatibility of these two theses is no doubt an interesting one. However, it does not track the issue we confronted in the argument for incompatibility above. The issue there, in short, was about whether content externalism is compatible with first-person access to intentional content, not with certain views about justification. If we have independently plausible reasons for adopting epistemic internalism, then we would have another argument against content externalism. I cannot help myself to such an argument, given what I have said about the justification for self-knowledge.²⁰ Whether or not that will convince epistemic internalists about self-knowledge to reconsider their views is an issue for another time. It is enough at this point to note that there is room for another incompatibilist argument against content externalism based on epistemological concerns. Some may think that the incompatibilist argument against content externalism works only if we assume epistemic internalism. However, this is not true. Even though Brown convincingly argues that all plausible versions of epistemic internalism may be incompatible with content externalism, the argument I advance above is independent of these concerns.

¹⁹ (Goldberg 2007), page 32-33.

²⁰ For what it is worth, it seems to me that the question of the correct view of justification is actually relative to the sort of knowledge in question. It seems to me to be likely that justification is externalistic for certain domains of knowledge and internalistic for others.

To sum up, we have seen that content externalism is incompatible with first-person access to intentional content. Some philosophers may view this as a difficulty, given the widespread acceptance of content externalism. We may seem to be faced with a dilemma: reject content externalism or first-person access to intentional content. Either option seems troubling. If we reject content externalism, we are turning our backs on powerful philosophical arguments. If we accept these powerful arguments, we will have to turn our backs not only on first-person access to intentional content but also on rationality. In Chapter 5, I argue that the arguments supporting content externalism are not as strong as some have thought. Seeing this will relieve the pressure that some may feel about giving up on content externalism.

CHAPTER 5 RETHINKING CONTENT EXTERNALISM

In the debate of whether content externalism is compatible with self-knowledge, it is clear that many philosophers do not want to give up content externalism. Some have thought that if self-knowledge conflicts with content externalism, then we need to revise our ordinary conception of self-knowledge. As I said earlier, I think that we should reject content externalism if there is truly a conflict. Having shown that there is a conflict between content externalism and first-person access to intentional content, it is now time to come to terms with rejecting content externalism. In this chapter I shall respond to the arguments for the types of content externalism we discussed in Chapter 3. By reexamining the arguments for content externalism afresh, I hope that they will seem less powerful than they have seemed to many philosophers. The purpose of this chapter is to show that we do not have compelling reasons to adopt content externalism based on the arguments for social externalism, natural kind externalism or singular thought externalism. Not only do we have reason for rejecting these arguments due to the incompatibility of their conclusions with first-person access to intentional content, but it also turns out that there are not good independent reasons for adopting them in the first place. Thus, the goal of this chapter is to relieve any anxiety some may have for rejecting content externalism.

Rethinking Social Externalism

Here is a generalized form of the argument that is shared by both natural kind externalism and social externalism.

- (P1) Beliefs and other intentional states are to be individuated by their truth conditions.

- (P2) It is possible for individuals to have the same intrinsic properties and yet assent to a sentence that has different truth conditions in different worlds.
- (C1) Therefore, it is possible to have two individuals who share all the same intrinsic properties and have different beliefs.
- (C2) Therefore, some intentional content is not solely a function of an individual's intrinsic properties.

The important question in assessing arguments for social externalism is this. At each respective world, what are the truth conditions of the sentence assented to by the subjects in the thought experiments? Let us recall Burge's thought experiment. Both subjects assent to "There is arthritis in my thigh." Part of the story of the thought experiments is that in the actual world "arthritis" has its normal meaning and in the counterfactual world "arthritis" means any rheumatoid ailment in the body. In the thought experiment, the individuals share all the same intrinsic properties. "Arthritis" means different things at each world, so the truth conditions of the beliefs will differ as a direct result of the difference in meaning of "arthritis" in each world. The intrinsic properties of the subjects are the same. The content of their beliefs differ. Therefore, intentional content is not a function of an individual's intrinsic properties.

The Limits of the Disquotation Principle

One way to see the problem with social externalism is by examining its commitment to what has been called "the disquotation principle." The disquotation principle concerns the relationship between the language of intentional state reports (ascriptions) and the content of intentional states. Saul Kripke names this principle in "A Puzzle about Belief." He defines the principle of disquotation as: "where 'p' is to be replaced, inside and outside of all quotation marks, by any appropriate standard English sentence: 'If a normal English speaker, on reflection, sincerely assents to 'p', then he

believes that 'p'."¹ The idea is that if a speaker sincerely assents to a sentence, then he believes the content which that sentence expresses. In other words, the sentence he assents to is the vehicle appropriate for expressing the content of his belief. This principle may sound obviously true. Surely if one sincerely assents to "p", then one believes that p. When I sincerely assert, "Antananarivo is the capital of Madagascar," it seems obvious that I believe that Antananarivo is the capital of Madagascar. This principle clearly guides our interpretative practices, but there are instances when more is needed than a simple application of it; we shall see that sometimes the sentence to which a subject assents is not the sentence that best captures the content of his belief.

The disquotation principle is fundamental to the argument for social externalism. A good deal of criticism of social externalism focuses on problems arising from a strict application of the disquotation principle. One main concern has been Burge's idea that the public meaning of the sentence a believer assents to is what, *in all instances*, gives the content of a subject's beliefs. Here, again, is a quotation from Davidson that captures this concern.

It seems false that our intuitions speak strongly in favor of understanding and interpreting an agent's speech and thoughts in terms of what others would mean by the same words. For one thing, there is the problem of deciding what group is to determine the norms. But more important, we understand a speaker best when we interpret him as he intended to be interpreted; this will explain his actions far better than if we suppose he means and thinks what someone else might mean and think who used the same words 'correctly'.²

We can see the limits of the disquotation principle in Kripke's case of Pierre. In "A Puzzle about Belief" Kripke discusses the case of the Frenchman Pierre. While living in

¹ (Kripke 1979), page 895.

² (Davidson 2001), page 198-199.

France, monolingual Pierre comes to assent to the sentence "Londres est jolie" from hearing stories of London and its charm. Later, he moves to England, and begins to learn English. Bilingual Pierre lives in London, which he does not find pretty. He fails to make the connection that the Londres from the stories he has heard and his current city are the selfsame place. Kripke thinks that there is a problem. Pierre has not relinquished his belief that London is pretty; he would still assent to the sentence "Londres est jolie." He also clearly believes that London is not pretty, because he would assent to "London is not pretty." His beliefs are clearly contradictory. Kripke thinks that there is no plausible way out of this puzzle. We must accept that he believes both that London is and is not pretty.

John Biro argues that Kripke's puzzle dissolves once we see that his argument assumes that we must *always* apply the principle of disquotation in determining the content of intentional states.³ Once we realize, Biro points out, that we need not always rely merely on the disquotation principle to determine what someone believes, the puzzle disappears. It is not hard to see that Pierre does not realize that the names "London" and "Londres" are co-referring. Based on this, we should conclude that his concepts of London and Londres differ. Since concepts clearly play a role in constituting beliefs, the use of *just* the disquotation principle should give us pause, as it is not always sensitive enough to the concepts a subject associates with the terms used in the intentional state ascriptions. Biro notes that we should draw the following morals from the Pierre case.

The general principle about belief attribution which these cases serve to underline is, roughly, this: to the extent to which someone is not fully aware

³ See (Biro 1984).

of the actual or conventional references of the names he uses, such reference should not be used by third parties to land him with beliefs which only someone who is aware of them would rationally hold. The belief we must attribute to someone we are trying to interpret is his belief, not ours, and there is no guarantee that we do this if in specifying the content of the belief we use expressions whose semantic content may be obscured from him by idiosyncratic semantic beliefs he happens to hold concerning those expressions.⁴

In the case of Pierre, it is easy to imagine an interlocutor asking him how he can believe that the same thing both is pretty and is not pretty. At some point in his discussion with the interlocutor, Pierre would no doubt realize that “London” and “Londres” are co-referring. At which point he would no doubt cease to assent to both “London is pretty” and “Londres est jolie.” The disquotation principle is our default principle is assessing the content of someone’s intentional states. After all, in assessing the content of a speaker’s intentional states, we start with the words a speaker himself uses to describe his intentional states. However, there are clearly cases, as in the case of Pierre, where using this principles leads us to attribute content to a speaker’s intentional states that is not correct.

Like Kripke’s example of Pierre, Burge’s arthritis example also adheres too rigidly to the disquotation principle. If a patient tells his doctor, “I think I have arthritis in my thigh. I have had pain there for a week,” the disquotation principle tells us to attribute to him the belief that he has arthritis in his thigh. To be sure, it seems very natural to apply the disquotation principle. However, in cases where we attribute to a subject inconsistent beliefs, we should do more work to assess the content of his belief. Rather than attribute to the patient the belief that he has an ailment exclusively of joints to an area of his body other than his joints, we should attempt to get a better sense of his

⁴ (Biro 1984), page 273.

concept of arthritis. It is clear that the patient thinks arthritis is an ailment, but he does not understand the limits of its locality. When we are sensitive to his concept of arthritis, it seems proper to attribute to him the belief that he has an ailment in his thigh that feels to him like arthritis (but of course is not). Usually the disquotation principle is a reliable guide to content determination, as speakers often use terms competently. However, as the examples above show, there are clearly cases where they do not.

If an agent is confused about the meaning of a certain term, then the disquotation principle will not be a reliable guide to determine the content of intentional states that ascribed to him (by himself or others) using the term about which he is confused. We need to use a more complicated method to determine what the subject believes, a method which involves getting a better idea of the concept the speaker intends to express in his use of the term about which he is confused. Tim Crane in "All the Difference in the World" analyzes Burge's arthritis case as follows. Given what we know of the situation, the subject associates the concept of a rheumatoid ailment that can occur both in the joints and in the muscles with the term "arthritis." Based on this, we should attribute to him the concept of tharthritis (Crane's terminology), a concept that picks out any rheumatoid ailment. He thus has the true belief that he has tharthritis in his thigh, and the false meta-linguistic belief that "arthritis" means any rheumatoid ailment. Here is Crane.

On this diagnosis, then, Alf has a true belief, I have tharthritis in my thigh, a false belief to the effect that "I have arthritis in my thigh" is the right sentence to express this belief, and thus makes a false statement, "I have arthritis in my thigh." This means that though his belief is true, he says something false when he attempts to express it. This sounds paradoxical, but it becomes clear when we distinguish, as we did when discussing

Putnam, between the meanings of sentence in public languages and the contents of beliefs.⁵

The method of determining thought content expressed above, a method that involves more than applying the disquotation principle, allows us capture the subject's concept when we attribute intentional states to him. The idea is that the personal concept that an agent associates with a term is what constitutes an agent's thought. There are times when the personal and intersubjective concepts diverge. Frequently, an agent's personal concept lines up with the respective intersubjective concept. It is hard to see how we could communicate effectively if this were not so. However, malapropisms, spoonerisms, and the like show that there are instances (often humorous ones) where the personal concept an agent associates with a term does not correspond to the intersubjective concept associated with that term. In order to properly interpret someone, and thereby attribute the correct intentional states to someone required for interpretation, we often need to distinguish between both the personal meaning and personal concept someone associates with a term and the intersubjective meaning and intersubjective concept associated with that term. There are other elements we must draw on in interpretation as well, such as speaker intentions. We must draw on these elements, for example, when someone says, "I believe that comparisons are odorous." Applying the disquotation principle alone will have us attributing to the speaker the belief that comparisons have a smell. There are other elements we must distinguish between in order to properly interpret the speaker, such as the personal concept being associated with "odorous" and the intersubjective concept associated with that same

⁵ (Crane 1991), page 298.

term. In sum, our normal method for attributing the content of intentional states to others is much more flexible and complex than a mere application of the disquotation principle.

Many others have noted that the argument for social externalism relies on a questionable assumption. This assumption expresses an explicit commitment to the disquotation principle. Kent Bach writes,

Underlying Burge's interpretation of this thought experiment is the assumption that, if one literally and correctly uses a term in the 'that'-clause of an attitude attribution, one is imputing to the subject the notion expressed by the term and is, further, including it in the content of the attitude being attributed.⁶

This is a questionable assumption, for even if one is correctly using a term in the that-clause of a belief attribution, there is no guarantee that the subject and attributor's understanding of the term is similar enough to warrant the attribution of the attributor's concept of a term to the subject.

Here is Brian Loar's description of the assumption in question from "Social Content and Psychological Content."

Sameness of the de dicto or oblique occurrence of a general term in two belief ascriptions implies, if everything else is the same, sameness of the psychological content of the two beliefs thus ascribed.⁷

We have good reason to reject this assumption. Kripke's Pierre shows that often times we cannot use the disquotation principle to correctly determine what someone believes, as we end up attributing obviously contradictory beliefs to a supposed rational agent.⁸

⁶ (Bach 1988), page 92.

⁷ (Loar 1985), page 183.

⁸ This is not to suggest that to count as a rational agent one must have a completely consistent set of beliefs. This would be too high a standard. However, in the case where the beliefs are blatantly contradictory, such as in the belief that London is pretty and the belief that London is not pretty, it would violate principles of interpretation to attribute such beliefs to a rational agent. Furthermore, it is clear from the example, that the subject associates two different concepts with

There is no principled reason that we must always attribute to subjects the intersubjective concepts associated with terms of intentional state ascriptions.

Communication and the attribution of intentional content is much more complex than is allowed for by merely applying the disquotation principle. For this reason, we have a principled reason for rejecting social externalism. It is not merely that it is in conflict with first-person access to intentional content. It also forces on us an implausible and false principle of belief attribution.

Rethinking Natural Kind Externalism

As I noted earlier, the arguments for social externalism and natural kind externalism share a similar form.

- (P1) Beliefs and other intentional states are to be individuated by their truth conditions.
- (P2) It is possible for individuals to have the same intrinsic properties and yet assent to a sentence that has different truth conditions in different worlds.
- (C1) Therefore, it is possible to have two individuals who share all the same intrinsic properties and have different beliefs.
- (C2) Therefore, some intentional content is not solely a function of an individual's intrinsic properties.

While the well-known arguments for natural kind externalism and social externalism share this same form, they differ in how they motivate (P2). As for natural kind externalism, Twin Earth is typically invoked to support the idea that two individuals who share all the same intrinsic properties can assent to sentences with different truth conditions in different worlds. They would both, for example, assent to the sentence, "Water is vital to life." All can agree that the sentence is true if and only if water is vital to

the terms "London" and "Londres." We should attribute different beliefs based on the differences in personal concepts.

life. However, not everyone agrees on what exactly water is. The intuition that seems to have won out is that something is water if and only if it is H₂O. So, on Twin Earth “water is vital to life” is true if and only if XYZ is vital to life, and, on Earth, “water is vital to life” is true if and only if H₂O is vital to life.

I want to discuss two responses to this argument. The first response is from Eddy Zemach, who argues against the idea that water is necessarily H₂O. The second response is from Tim Crane, who argues that the thinkers on Earth and Twin Earth who assent to the sentence “Water is vital to life” share the same thought content in virtue of sharing the same concept of water. I believe that both of these responses to various aspects of the argument for natural kind externalism are successful, and once properly appreciated, show us that the Twin Earth argument for content externalism is not as strong as has been thought.

Zemach’s Response

In “Putnam’s Theory on the Reference of Substance Terms” Zemach challenges the idea that the extension of the term “water” differs on Earth and Twin Earth using Putnam’s own method of reference determination. For Putnam, to determine the extension of a substance term, we first look to what most people refer to using that term. Next, we examine the microstructural properties of that substance. Only substances with those microstructural properties are counted in the extension of the term. Thus, on Earth, we use “water” to refer to the substance in the lakes and streams, and so on. This substance is H₂O. Therefore, “water” refers to H₂O and nothing else.

Zemach begins his paper by noting that it far from clear that Earthians and Twin Earthians are in different languages communities. If they are in the same language community, then by Putnam’s own method of reference fixing the extension of “water” is

H₂O and XYZ. Even if there is a way to show that English speakers on Earth and Twin Earth are in different language communities, Zemach notes that it is far from clear that those on Earth would not call the liquid that functions and looks like water “water.”

Earthmen, too, may call the substance in the seas, lakes, etc., of Twin Earth by the name “water.” They will certainly do so before they become chemically sophisticated and discover that Earth water has a different molecular structure from that of Twin Earth water, and they might continue to do so even after having made this discovery. That is, they may say that what has been discovered is that some water is made out of H₂O molecules and some water is made out of XYZ molecules, but both are equally water. After all, this is exactly what we say of so many other materials: paper may have widely different chemical structures, and so may sand, and cloth, and stone, and hair, and glue, and . . . Chemical constitution is not always decisive in determining our usage of substance names.⁹

It seems to me that this response has not been fully appreciated by most. Putnam appears to have too much confidence in the response that XYZ would not be called “water.” In “The Meaning of ‘Meaning’,” he writes,

If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that “water” has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that “water” on Twin Earth is XYZ, and the Earthian spaceship will report somewhat as follows: “On Twin Earth the word ‘water’ means XYZ.”¹⁰

Putnam does note that by “means” here he has something in mind like “has the extension.” Even with this distinction, it is still an open question whether or not the Earthians in the spaceship would think that XYZ *is* water or some other substance. Even if we grant that the extension of “water” differs at each world, it is still another question whether XYZ is water.

⁹ (Zemach 1976), page 62.

¹⁰ (Putnam 1975), page 10.

To strengthen the point that substance terms may have disjunctive extensions, Zemach notes that on Earth we actually use the term “water” in such a way. He writes,

Consider the case of heavy water. Nothing that is composed of D_2O molecules is composed of H_2O molecules, yet (Saul Kripke to the contrary) heavy water is commonly regarded as a kind of water. The same holds for aggregates of T_2O , HDO , HTO , and DTO molecules (the number of varieties is eighteen, since in each case the oxygen can be either O^{16} , or O^{17} , or O^{18}). All these, we say, are different kinds of water. Moreover, there is no chemical constitution common to all bodies of ordinary water; different samples differ chemically. Ordinary water contains some D_2O , some H_2O_2 , some H_2O_3 , a large amount of sodium chloride, and various minerals . . . if English speakers have the right to call all these liquids “water,” they may also add aggregates of XYZ molecules to this list, and Putnam will have to accept this classification as an uncontestable datum.¹¹

The actual practice of English speakers applying the term “water” to aggregates of molecules other than H_2O suggests that according to Putnam’s own story of reference the extension of “water” is not in fact H_2O and therefore not necessarily H_2O . The lesson in all this is that the essence of water is not found in its underlying microstructural properties. Rather, the essence of water is found in its functional and appearance properties. If a substance functions as water does and looks like water, then it counts as water.

Crane’s Response

While Zemach’s focus is on the nature of substances, Tim Crane’s response in “All the Difference in the World” is directed at the nature of intentional content. Crane begins by noting that the argument for natural kind externalism turns on how we understand the truth conditions of the sentence the Earthian and Twin Earthian assent to. Crane asks,

Why should we believe it [that the truth conditions of the sentences differ at each world]? Why should a difference in the chemical structure of water affect the truth conditions of “water is wet”? Why should we not say, for

¹¹ (Zemach 1976), pages 62-63.

instance, that if there were such a substance as XYZ, all this would show is that “not all water has the same microstructure” (Mellor (1977), p. 303). And indeed, we knew this already, since it would surely be stipulative to deny that heavy water (D_2O) is really water. So why does the fact that most of our water is H_2O entail that the truth conditions of “water is wet” differ across Earth and Twin Earth?¹²

The idea here is that if we consider the point of Zemach (and Mellor) about the extension of “water” being disjunctive, then it is far from clear that the truth conditions of the sentence would differ at each world. If the truth conditions do not differ, then, the respective content of the intentional states will not differ, and then a Twin Earth style thought experiment could not be used to show that natural kind externalism is true. This would be enough for someone who rejects the idea that water is necessarily H_2O . Of course, there are many who accept this idea, and also in general accept that other substances have essential natures that are based in their microstructural properties. For these individuals, Crane offers some additional considerations.

Crane offers another response, which we can call the common concept response. The basic idea here is that the truth conditions of the sentence in question (for example, “water is vital to life”) should be cashed out relative to each individual’s personal concept of water.

Take aluminum and molybdenum, two practically indistinguishable metals whose names (and relative scarcity) are switched on Earth and Twin Earth. And suppose my Twin and I are atom-for-atom identical. Putnam says that the meaning of “aluminum” spoken by my Twin and me is underdetermined by our narrow states—the states we share. Our narrow states determine molybdenum and aluminum as the extension of our uses of the word “aluminum.” But why not therefore say that neither of us has a full understanding of the meanings of our words? Why not say that we have the same concept (call it “molyminum”) that applies to aluminum and

¹² (Crane 1991), page 290.

molybdenum alike? The concept molyminum will distinguish less finely between substances than the concepts aluminum and molybdenum.¹³

One way to understand this suggestion is in terms of personal and intersubjective concepts. There is the personal concept of molyminum (that corresponds to the personal sense of “aluminum” and “molybdenum” shared by both Crane and his Twin in the example above), and there are the intersubjective concept of aluminum and the intersubjective concept of molybdenum, which neither Crane nor his Twin fully grasp. Their personal concepts do not distinguish between the substances aluminum and molybdenum, while the respective intersubjective concepts of aluminum and of molybdenum of course do. In characterizing a subject’s intentional states we should cash out the truth conditions in terms of personal concepts as opposed to intersubjective ones. In characterizing the public meanings of sentences, we should cash out the truth conditions in terms of the intersubjective concept (the concept shared by those in the know).¹⁴ Thus, the previously discussed sentence “water is vital to life” will have the same truth conditions when it comes to the Earthian and Twin Earthian beliefs about the water-like substances in their respective environments, because they share the same concept. Their beliefs are true if and only if H₂O or XYZ (or whatever plays the watery role) is vital to human life, because the Earthian and Twin Earthian have the same personal concept of water, a concept which applies equally to H₂O or XYZ or anything that functions and looks like water. The public meaning of “water” will not be determined by anyone’s personal concept, but will be determined by those in the

¹³ (Crane 1991), page 291.

¹⁴ These two sentences essentially say that we should be content internalists and semantic externalists.

know. I have suggested that the reference for “water” (what the intersubjective concept picks out) should be considered to be whatever plays a water role and looks like water, though someone can reject this and still adopt Crane’s approach, provided that one allows for a distinction between personal and intersubjective concepts. For example, there are those that think that water is necessarily H₂O, and therefore that the public meaning of “water” will be constrained by this fact. As Putnam has argued, if “water” has a different extension at Earth and Twin Earth, then by the assumption that meaning determines reference, the total meaning of “water” differs at Earth and Twin Earth. However the public meaning of “water” should be determined, we can leave that question open and still respond to the argument for natural kind externalism using Crane’s strategy: the truth conditions of intentional states should be cashed out in terms of personal and not intersubjective concepts.

Aside from providing a way to avoid the content externalist’s conclusion, we have independent reasons to prefer this sort of strategy. Distinguishing between personal and intersubjective concepts, cashing out a subject’s intentional states relative to personal concepts, and cashing out public meaning relative to intersubjective concepts are all required to make sense of certain forms of communication. As Crane points out, a distinction between conventional (or linguistic) meaning and speaker meaning (what the speaker intends to express) is required to make sense of many non-literal uses of language, such as punning. The attendant intentional states involved in instances of non-literal uses of language also require us to make a distinction between personal and intersubjective concepts. Sometimes speech contains puns where the speaker intentionally uses language in an aberrant way for effect. Sometimes speech contains

unintentional misuses of language, such as malapropisms or spoonerism. To make sense of the content of such utterances, we need to distinguish between conventional meaning and speaker meaning. Take the example we have been using from Shakespeare's Constable Dogberry, who says, "Comparisons are odorous" (*Much Ado About Nothing*, Act 3, Scene V). We do not interpret the speech to mean that comparisons give off a smell, even though that is the literal meaning. We take the speaker to be intending to speak truthfully and mistaken in his use of words. We know that he meant to say that comparisons are odious. Likewise, we do not attribute to him the belief that comparisons give off a smell, but the belief that they are very unpleasant. To do this we need to distinguish between the concept the subject associates with "odious" and the intersubjective concept associated with "odious." Crane's strategy is one we employ all the time, and not one merely crafted as a response to the arguments for natural kind externalism.

Taking stock

The responses of Crane and Zemach have different targets. Zemach sets out to argue against Putnam's theory of how substance terms refer. Crane set out to argue against the idea that Twin Earth can be used to show that natural kind externalism is true. Zemach's response can be used to argue against natural kind externalism in the following way. If we accept Zemach's argument, then we can see that water is not necessarily H₂O. Substances with other microstructures can be water, as long as they have all the functional and appearance properties of water. Thus, if we have an Earthian and Twin Earthian who share the same intrinsic properties, their beliefs about the water-like substance around them will not differ. When they assent to the sentence, "Water is vital to human life," they assent to a sentence that has the same truth conditions at each

world. On Earth the sentence is true if and only if water (H₂O or XYZ or another aggregate that functions likewise) is vital to human life. On Twin Earth, the truth conditions are the same. Thus, the respective beliefs are the same.

This line of argument is further supported by Crane's response, for we can say that the Twin Earthian and the Earthian share the same concept of water, a concept that tracks the functional and appearance properties of water, or "superficial properties" as John Locke called them. I think that Zemach's response is correct and complimentary to Crane's response. I encourage philosophers to reevaluate the success of Putnam's argument for the reference of substance terms, although that point is not one on which the main points of this chapter turn. As Putnam's conclusion about the reference of substance terms is well accepted by philosophers, I should stress that one can reject Zemach's argument and still reject natural kind externalism using Crane's response. That is, one can hold that "water" necessarily refers to H₂O *and* that the aforementioned beliefs of the Twin Earthian and Earthian do not differ. Granted the *sentence* "Water is vital to human life" has different truth conditions at each world for someone who believes that "water" necessarily refers to H₂O, but if we accept that the truth conditions for intentional states should be cashed out relative to a subject's concepts and not the intersubjective concept of his community, then the sentence "water is vital to human life" is not the best sentence to capture their shared personal concept of water. Perhaps, "watery stuff is vital to human life" is the sentence that best captures their belief, as the concepts that they associate with "water" do not include information about its chemical properties, at least in the version of Twin Earth which we have been discussing. Those who do not wish to give up on Putnam's (and Kripke's) famous doctrine of the reference

of substance terms, and feel pressure to reevaluate commitments to natural kind externalism, should be relieved that they can reject content externalism and still hold the doctrine of *a posteriori* necessities. The distinction between speaker meaning and sentence meaning and the attendant distinction between personal and intersubjective concepts that are required for the responses to social externalism and natural kind externalism should not be seen merely as the artificial machinery of a response to a philosophical argument, but as machinery that is required to make sense of many aspects of communication. As mentioned previously, there many instances in communication when we must make a distinction between not only conventional meaning and speaker meaning but also the concept a subject associates with a word and the intersubjective concept associated with a word.

Rethinking Singular Thought Externalism

In Chapter 3, we discussed Jessica Brown's definition of singular thought externalism: "According to singular anti-individualism, a subject's thought contents are individuated partly by the particular objects that are in her environment (see, e.g., Perry 1979; Kripke 1980; Evans 1982; Peacocke 1983; McDowell 1986; Salmon 1986; Soames 1987; Kaplan 1989)."¹⁵ Let us review the argument for the conclusion that singular thoughts are individuated externalistically.¹⁶ Brown asks us to imagine a thinker who has a perceptually demonstrative thought about an apple. To have such a thought, the subject must have first identified the apple by demonstration. Imagine a counterfactual situation in which, familiarly, everything is the same from the inside, as it

¹⁵ (Brown 2004), page 13.

¹⁶ Again, the adverb "externalistically" is a shorthand way of referring to individuation that depends on a thinker's extrinsic properties.

were, and the difference lies outside: there is a different apple. The intuition is supposed to be that the thoughts are different in virtue of being about different apples. The thoughts, then, would be individuated externalistically, since the intrinsic properties of the subjects are the same and the intuition is that the thoughts are different. The thoughts are object-dependent in that they are essentially identified by the objects they are about.

The above argument is too quick. To accept the externalist conclusion, we need to hear more about the role that perceptions of objects play in the recipe of singular thoughts. Once we see the steps of the argument laid out, we shall see that there are options for resistance.

Arguments for Singular Thoughts being Individuated Externalistically

First, we should note that there are two possible conceptions of singular thought. One conception derives from the classical view of Bertrand Russell's according to which singular thoughts involve unmediated access to their objects. Russell thought that we must be acquainted with objects in order to have singular thoughts of them. Russellian acquaintance is famously demanding: we can be acquainted only with objects that we cannot possibly misidentify; this rules out singular thought about material, extra-mental objects and any objects outside of sense-data, universals and possibly the self. Another conception allows that singular thoughts may be mediated by some conceptual element. Thoughts are singular in so far as the concepts that constitute the thought are object-dependent. The nature of the concept of the object is determined relationally; it would not be the concept it is unless it were related in the proper way to the object that it is of. This second conception, the singular concept or *de re* concept view, is more common among philosophers today. I discuss a form of it first, then I discuss how a more

classically Russellian conception of singular thought can be used in an argument for the conclusion that some of our singular thoughts are individuated externalistically.

From *de re* concepts to content externalism

In “Getting a Thing into a Thought,” Kent Bach defends a conception of singular thought that departs from the classical Russellian view. He writes,

Russellian acquaintance is an unmediated cognitive relation. It suggests the idea of pure *de re* thought, even though, of course, for Russell one cannot bear this relation to material objects other than oneself. Aside from its inspiration from Russell, whatever appeal there is to the idea of pure *de re* thought depend on a false dichotomy: the only alternative to thinking of an object under a description (i.e. individual concept), which does not really count as thinking of it in a *de re* way, is to think of it directly, in an unmediated way. But this dichotomy misrepresents the relevant contrast. The contrast is not between mediated and unmediated thinking.¹⁷

Bach clearly thinks that we can have singular thoughts that do not involve the unmediated access to objects Russell envisioned. To articulate his conception of singular thought, Bach deploys the notion of a *de re* representation, a representation of an object that instantiates extrinsic properties. This is what I shall mean by the phrase “*de re* concept of an object.” In order to have such a singular thought of an object we must have the right sort of cognitive connection to the object. Bach, in “Getting a Thing into a Thought,” discusses this first step in the recipe for singular thought. He uses the term “representational connection” to capture the sort of contact a subject must have with an object initially to later have a singular thought of it. Bach notes that to have such a connection, we need not actually visually perceive the object. I can have a singular thought about your mother, for example, if you tell me enough about her. Bach readily

¹⁷ (Jeshion 2010), page 54-55.

admits that it is not always easy to say what counts as a representation connection sufficient for singular thought.

So, for example, does seeing a photograph or film of someone put one in a representational connection with them? Hearing someone's voice? Does reading someone's name do the trick, even outside the context of communication? For example, can you have singular thoughts about someone whose name you read on a luggage tag, in a phone book, or on a tombstone? I'm inclined to think not.¹⁸

We need not settle questions like this here. It is enough to note that singular thought begins with a representational connection between a subject and an object. A clear case of such a connection is visual perception. Less clear cases are ones such as seeing a name in a phone book or on a luggage tag.

Next, the concept that factors into the singular thought that the subject has of the object as a result of the representational connection must meet some sort of strong uniqueness condition.¹⁹ The concept must be uniquely attached—in some way—to the object it is of in order to establish the uniqueness that is required for singular thought. We can understand the uniqueness condition here in the following way. Two numerically distinct objects that cause qualitatively identical perceptions in a subject will be the source of two different concepts in virtue of their difference in the numerical identity of their objects. In a case where a subject perceives numerically distinct objects as identical, the concepts he uses to think of those objects are individuated by the distinct objects they were caused by. Thus, we have a clear case wherein the subject

¹⁸ (Jeshion 2010)

¹⁹ I should note that a subject could of course have more than one concept of the same object and not realize that the concepts co-refer. I may think of a person, for example, as my neighbor Mr. Green, and not realize that he is also the priest to whom I confess every Sunday. It seems possible that I can have distinct singular thoughts about the same object, assuming that singular thoughts can be constituted by concepts of objects and not the objects themselves.

instantiates the same intrinsic properties when thinking two different thoughts. The difference is in the external world. Thus, the thoughts instantiate extrinsic properties of their thinker in virtue of their respective concepts being uniquely tied to distinct particular objects.

Here is summary of the argument.

- (P1) Numerically distinct objects that are qualitatively identical to subjects when initially perceived cause distinct concepts despite being perceived as identical. [Uniqueness condition]
- (P2) Thoughts about objects perceived will be constituted by a subject's concept of the object perceived.
- (P3) In a case where a subject perceives numerically distinct objects as identical, the concepts he uses to think of those objects are individuated by the objects they were caused by.
- (P4) The thoughts in these cases will be constituted in part by the concepts of the objects.
- (C1) Therefore, the thoughts will have different individuation conditions due to the difference in external objects.
- (C2) The difference in individuation conditions is due to a difference in extrinsic properties of the thinkers.

The uniqueness condition is driving the externalist argument here, and I think that one can resist the argument by rejecting it. Consider the following case as a counterexample to the uniqueness condition. Imagine that I am shopping for tables. I see Table 1 in Store A. I note how it looks and its location next to the bookcases. I like Table 1, but I want to look at another store before I buy a table, so I decide to go to Store B. As I leave Store A to go to Store B, I have formed a concept of Table 1 based on my visual perception of it. I do not see anything I like at Store B and therefore return to Store A. Unbeknownst to me, someone has bought Table 1. (Let us say that someone liked Table 1 so much that they had to have the floor model and did not want one from the

storeroom.) By the time I am back in the store, a new table, Table 2, has been assembled by the store's staff in the very same spot. It looks to me exactly as Table 1 did. The question is whether I do in fact form a new concept based on perceiving this new table or not. Let us say that I just cannot decide whether I want to buy a table and leave Store A. As I leave I think, *that table is of good quality and is being sold at a reasonable price*. The question to ask here is not: What does my phrase 'that table' pick out? The question is: How many concepts do I have? Do I have a concept of both Table 1 and Table 2? One who holds the uniqueness condition will say yes. It is far from clear to me that this is the correct response. It seems to me that there is only one concept, the concept of the table that I saw in the store that looked thus and so and was located near the bookcases. Based on the situation, I do not have the ability to discriminate between Table 1 and Table 2. Had the staff explained to me what happened, I would surely be able to discriminate between the two tables. However, as it stands, it seems to me that I have one concept, a concept that is not fine-grained enough to distinguish between the two tables. Notice that I am not demanding anything as strict as Russellian acquaintance here. The point is that if the subject cannot discriminate between the two concepts (the concept of Table 1 and the concept of Table 2), then we have to ask whether there are really two concepts in play. It seems to me that there are not in virtue of the subjective indistinguishability of the two putative concepts. It seems that there is but one concept. I should note that I can still surely refer to both tables. Let us say that I go back into the store, point to Table 2, and say, "I want that table." I surely refer to Table 2. Such is the nature of demonstration. However, I am not able to form two thoughts that pick out both Table 1 and Table 2. If I have more information, say, that

Table 1 was sold and Table 2 is the one I am considering whether to buy, then I can discriminate between the two tables in virtue of the new information, thus providing me with distinct concepts of Table 1 and Table 2.

Someone may press me here on the question of what my thoughts are about. Surely, the line runs, we want our thoughts to be about the right objects. When I leave Store 1 the first time after seeing Table A, and say, "That table is of good quality and is being sold at a reasonable price," then my thought is of Table A. And when I return to Store 1 later and in seeing Table 2 I say, "I must really admit that that table is of good quality and is being sold at a reasonable price," then my thought is of Table B. Thus, the uniqueness condition must be preserved so as to capture the right intuitive result that these two thoughts are distinct in virtue of being about different objects. To this response, it should be noted that there is nothing in my response that says that my thoughts cannot be about both Table 1 and Table 2. On my account, my concept can pick out either table. I do not, after all, have enough information to tell one table from the other. Some may worry that this is not an account of singular thought. I address this in a later section called "Are there Singular Thoughts?"

From neo-Russellian singular thoughts to content externalism

Another argument from singular thought to content externalism proceeds by way of modifying Russell's conception of acquaintance. Such a view would not limit the objects of acquaintance to sense-data, universals and possibly the self, but also include mind-independent objects. On such a view, when we have a singular thought of an object, the object itself is in the thought; there is no mediating conceptual element. Such a view would block the above response that I gave to the argument from *de re* concepts to content externalism, as there is now no conceptually mediating element. The object

itself is what constitutes the thought. Although the idea that we can be directly acquainted with external objects is perhaps not yet clear, it is clear how content externalism follows from such a conception of singular thought. If an external object is a constituent of a thought, then that thought will be extrinsic in virtue of the thought instantiating extrinsic properties.

In addressing such an argument, it will be helpful to lay out some distinctions. We need to talk about the distinction between Russellian and Fregean propositions and between what have been called *de re* and *de dicto* attitude reports. Getting clear on these distinctions will help us to address the argument at hand.

To oversimplify, Russell and Frege both believed that propositions were structured entities, but they famously disagreed about the types of entities that structure propositions. Russell thought that it was objects themselves and properties and Frege thought that it was senses—objective concepts of entities. Their foundational work on propositions has given rise to talk of Russellian and Fregean propositions. When it comes to intentional contexts, the Fregean claim is that the proposition someone is intentionally related to is constituted by senses themselves and not the referents of those senses. The Russellian claim is that the referents of the terms in the sentence that expresses the proposition are part of what constitute the proposition in question. Take the sentence “Sam believes that Nashville is a boring city.” For the sake of example, let us say that the noun clause “that Nashville is a boring city” expresses a proposition. In this case, the Russellian thinks that this proposition is structured in part by Nashville itself, and the Fregean thinks that it is structured in part by the objective

concept of Nashville.²⁰ It may seem that if we use these propositions to individuate intentional states, then these intentional states will be externalistic, both Nashville itself and the objective concept of Nashville do not depend solely on the intrinsic properties of any individual thinker. We shall discuss this worry in a moment. However, the point to focus on is that one can have a view of singular propositions such that they are individuated by objects themselves or by concepts of objects. The argument from Neo-Russellian singular thought to content externalism focuses on singular thoughts as thoughts that can be characterized by singular propositions. Thoughts, that is, that are directly tied to objects in the world in a way that makes them externalistic in a straightforward way.

I bring up the distinction between Russellian and Fregean propositions because understanding the proper relationship between propositions and intentional content is relevant to the argument at hand. Philosophers often talk of propositional attitudes, attitudes individuals can take toward a proposition (though not everyone accepts that propositions are the objects of thought). It can seem quite natural to say that a proposition that has Nashville and the property of being a boring city as constituents

²⁰ In "Sense and Reference" Frege offers the following analogy to help us understand the idea of an objective concept.

The reference of a proper name is the object itself which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object itself. The following analogy will perhaps clarify these relationships. Somebody observes the Moon through a telescope. I compare the Moon itself to the reference; it is the object of the observation, mediated by the real image projected by the object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or experience. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers. (Frege 1892), page 566.

describes Sam's thought about Nashville. When we look at the Russellian proposition that describes the belief, we can simply read off of it that Nashville is the object. Clearly, on such a reading, belief is externalistic in virtue of have an external object as a constituent. However, we should exercise some care here when we talk about propositions being the objects of intentional attitudes. There is a difference between saying that the Russellian proposition *is* the object of the belief and saying that it *characterizes* the belief. There are at least a few worries one could raise here about construing the objects of intentional states as being propositions. In the case of singular thought, to say that the object of a singular thought is a singular proposition is to make it seem that singular thought actually does not involve an unmediated connection with an object. If what one is directed at in having a singular thought is a proposition structured by an object, then it seems that we do not in fact have unmediated access to that object if our thought is about the object only in virtue of the object being a constituent in a Russellian proposition. This suggests that it is best to think of Russellian propositions as *characterizing* the content of singular thoughts.

Suppose that the Neo-Russellian is happy with such talk of Russellian propositions characterizing the content of singular thought as opposed to serving as the objects of such thoughts. Does not the very nature of singular thought still ensure that not all thoughts are internalistic? Here we need to ask ourselves about the relationship between thinker and the object he has unmediated access to. What type of relationship is this? As I noted earlier, Bach thinks that a subject must first have some representational connection with an object to later have a singular thought of it. This much seems obviously true. However, how can we understand a thought of an object

that is not mediated by any conceptual element? Some work here has been done by Neo-Fregeans like Gareth Evans and John McDowell in articulating Fregean modes of presentation as both descriptive and non-descriptive. Francois Recanati sums up this distinction as follows.

Modes of presentation are now [on the Neo-Fregean account] ways the object is given to the subject, and an object may be given directly, in experience, or indirectly, via descriptions. Non-descriptive modes of presentation are ways the object is (directly) given to the subject in experience, while descriptive modes of presentation are ways the object is (indirectly) given via properties which it uniquely instantiates.²¹

One way to think of this distinction is in terms of the types of linguistic resources the subject has available to describe thoughts with these two types of modes of presentation. For a descriptive mode of presentation, the subject can use descriptions to describe the object. Suppose I sincerely assert, “The 44th President of the United States lives in Washington, D.C.” The description “the 44th President of the United States” gives information that does not involve demonstratives. Contrast this with the following belief ascription. Suppose I assert, “The man who cut in front of me while we were both jogging around the White House is rude.” Suppose that both beliefs are actually about President Obama and that I was not aware that the man who cut me off was actually the president. The non-description mode of presentation is marked by an ascription that employs demonstration; it refers to how Obama was presented to me. It references an experience of mine in an important way such that someone who is not familiar enough with that experience will not understand who my thought is about. If someone were jogging alongside me, then he would surely be also able to think of the man using the same (or sufficiently similar) non-descriptive mode of presentation

²¹ (Jeshion 2010), page 148.

because we shared a sufficiently similar experience. The point we should stress is that accepting this Neo-Fregean account of modes of presentation accepts that singular thoughts are mediated by some sort of conceptual element. Once one admits this, then the straightforward, simple argument from neo-Russellian singular thought to content externalism is abandoned, and we are back to the argument from *de re* concepts to content externalism.

***De re* and *de dicto* ascriptions**

Let us consider a related point about *de re* and *de dicto* thought ascriptions. It may seem that various ways of ascribing an intentional states can lend plausibility to the view that there are singular thoughts that are directly about objects. Philosophers often make a distinction between *de re* and *de dicto* thoughts. Using an example from Quine, Burge notes that that there is a grammatical distinction between two different ways to ascribe belief.

In epistemic contexts, the grammatical distinction is between belief in a proposition and belief of something that it is such and such. Many examples, here as in necessity contexts, are ambiguous. But some are not. Thus

- (a) Ortcutt believes the proposition that someone is a spy.
- (b) Someone in particular is believed by Ortcutt to be a spy.²²

Certainly we have two distinct ways to ascribe a belief to someone. (A) seems, on the face of things, to relate the believer to a proposition (that someone is a spy). (B) seems to relate the believer to an individual (Ortcutt) and to a property (being a spy). We may have good reason to prefer one type of ascription in favor of another in order to disambiguate between the *de re* and *de dicto* reading of “Ortcutt believes that someone

²² (Burge 1977), page 340.

is a spy.” If we know that Ortcutt has a particular individual in mind, then we can reach more precision in our belief report by using (b). If Ortcutt believes the general proposition that there is at least one spy, then we should use (a) in reporting his belief. It is not hard to see how someone could think that *de re* ascriptions describe singular thoughts.

- (c) Ortcutt believes that Jones is a spy.
- (d) Ortcutt believes of Jones that he is a spy.

In (d), “Jones” is outside the scope of the intentional verb. Thus, the truth of (d) as an ascription seems to require the existence of the referent of “Jones” in a way that (c) does not. We may think that (d) expresses a singular thought. If the belief attributed by (d) has Jones as an actual constituent, then it is clearly externalistic.

To reach this conclusion, though, is to read too much into the syntax of these belief ascriptions. *De re* and *de dicto* ascriptions are ways of describing intentional states and should not be expected to always exactly express intentional content. Bach makes a similar point.

A “believes-that” report can report a singular belief and a “believes-of” report can report a general belief. In short, the form of a belief report does not determine the type of belief being reported. (Jeshion 2010: 45). It is easy to imagine examples of *de dicto* ascriptions that can ascribe beliefs about a particular object.

- (e) Smith believes that Jones is a spy.

Bach is surely right. We can also easily imagine *de re* ascriptions that describe beliefs that are not singular, in the sense that they do not uniquely pick out an object. Let us say that Ortcutt believes there is spy in his midst but does not know who it is. Smith who knows that Jones is the spy in Ortcutt's midst could use the following ascription to attribute a belief to Ortcutt.

(d) Ortcutt believes of Jones that he is a spy.

From what we know of the situation, this ascription should not attribute a singular thought to Ortcutt, because Ortcutt does not have a particular individual in mind. Smith uses the above attribution because he knows Jones is the spy in question. But from the context of the situation, it seems that this ascription can be used to attribute a general and not singular belief to Ortcutt.

Are there Singular Thoughts?

If there is no good argument from singular thought externalism to content externalism, do we need to revise our conception of singular thoughts? That is, if it is not the case that the objects themselves do not structure the thought in the way we supposed, what is left of singular thought? It can seem that we must choose between the classical Russellian picture where we can have singular thoughts only about entities we cannot possibly misidentify or the classical Fregean picture on which we think of the world always via some description and never directly. We have reasons for resisting the starkness of both pictures. It is true that we must always think of objects via some conceptual element. However, some of our concepts are *de re* concepts, concepts that are of objects. But, we rejected the uniqueness condition. This means that we are left with the idea of *de re* concepts that seem incapable of picking out objects uniquely. We want direct, unmediated access to external objects, but such access is not possible.

In lieu of such access, what we can have is *de re* concepts that *approach* the uniqueness condition. We can have concepts of particulars that are based on many representational connections. This is what our concepts of the particulars most familiar to us are like. With each representational connection our *de re* concepts become more fine-grained, enabling us to be in a position to discriminate between particulars. Thus,

when I gain more information about Table 1 and Table 2, I become more equipped to have a particular thought about each respective particular. There is less chance of more than one particular satisfying the concept. To rule out completely the possibility of more than one particular satisfying a concept, we must hold that concepts have their identity based on the particulars that caused them. This has implausible consequences though, for it means that a subject could have two distinct concepts, Table 1 and Table 2, and not be aware of the distinction between the concepts. The uniqueness condition secures the unique satisfaction of a subject's concepts at the potential cost of making the identity of those concepts opaque to a subject. On these grounds, it should be rejected. So, we can have singular thoughts, although we must be clear that the *de re* concepts that structure these thoughts cannot be tied to their objects in a way that is not sensitive to a subject's perspective. If a subject cannot discriminate between two putatively distinct concepts that are of different objects, then there are not really two concepts in play. Rejecting that strong uniqueness condition is not to reject that we can have *de re* concepts, but to reject that concepts can be individuated solely by the objects that in some manner are responsible for giving rise to those concepts.

Diagnosing the Arguments

In one way or another, each of the above arguments for content externalism assumes a false connection between the language of intentional state reports and the content of the intentional states. Quite a number of philosophers have noted this assumption. Here are some examples.

Kent Bach in "Do Belief Reports Report Beliefs?" calls it "the specification assumption." It is the assumption, in Bach's words, "that the 'that'- clause in a belief

report specifies the thing that the believer must believe if the belief report is to be true.”²³ Bach argues that this assumption is false.

The Specification Assumption is false: even though their ‘that’-clauses express propositions, belief reports do not in general specify things that people believe (or disbelieve) – they merely describe or characterize them. A ‘that’-clause is not a specifier (much less a proper name, as is sometimes casually suggested) of the thing believed but is merely a descriptor of it. A belief report can be true even if what the believer believes is more specific than the proposition expressed by the ‘that’-clause used to characterize what he believes.²⁴

David Lewis offers a similar moral in discussing the Causal Theory of Reference that one finds in Kripke and Putnam.

The New Theory of Reference teaches that meanings ain't in the head. That may be right—it depends on which of the many sorts of semantic values that new theorists of reference must distinguish best deserve the name "meanings." If it is right, it applies inter alia to the sentences whereby we express our beliefs to others and to ourselves. But the proper moral is not that beliefs ain't in the head. The proper moral is that beliefs are ill-characterized by the meanings of the sentences that express them. Hilary may express one of his beliefs by the sentences "Elms are threatened by disease," although the meaning of this sentence, in some sense of "meaning," depends on more than is in his head. But if so, then it seems that what Hilary believes and what his sentence means cannot be quite the same.²⁵

Lewis here endorses a distinction between the public meaning of belief ascriptions and the content of the intentional states they are used to describe. Here is Kirk Ludwig on the distinction between the public meanings of ascriptions and the content of intentional states.

It will be natural to press into service to describe our thoughts about the world the language we have introduced for talking about it. As in the case of talk about objects around us, in talking about our thoughts or the thoughts

²³ (Bach 1997), page 221.

²⁴ (Bach 1997), 225.

²⁵ (Lewis 1979) page 526.

of others about those objects, what may be of most importance is what we think about an object, rather than how we refer to it. Thus, it is natural and convenient to press into service in describing each other's thoughts sentences using directly referring terms. When we use such devices, we only partially specify the content of each other's Cartesian Thoughts. But this serves most of our purposes well enough. In addition, often we don't have any very good idea of how an individual picks out an object, though we know what object it is he is thinking of and what he wants to say about it. This is what happens when someone uses a proper name or other directly referring term to pick out an object to say something about it. We know which object, but because the semantics of the referring term underdetermines how the user picks it out, we will often not know how he is picking it out on a particular occasion. In such cases, our best evidence for what the speaker thinks is what he says, and we do no better than to use his sentence (or a sentence synonymous with it) to characterize his mental state, even if in doing so we undercharacterize his Cartesian Thoughts.²⁶

There are other philosophers who argue that public meaning of intentional state ascriptions does not always determine the content of intentional states.²⁷

Ludwig's idea of a Cartesian thought is an idea of an internalistic thought. In Chapter 4 we saw that this sort of thought is required for first-person access to intentional content, which is itself a prerequisite for rational agency. This seemed to force a rejection of content externalism, which many believed had strong support. We have seen that the arguments for content externalism can all be met, so we have eliminated one reason for unease about adopting content internalism. Other concerns about adopting this view will be addressed in Chapter 6.

Content Externalism, Semantic Externalism

One final reason that the arguments for content externalism can seem compelling is that it is sometimes assumed that there is no distinction between semantic externalism and content externalism. Here is a brief description of this line of thought. If

²⁶ (Ludwig 1996b), page 445.

²⁷ See (Biro 1984) and (Jackson 2003a, 2003b, 2004) in addition to the Bach and Loar quotes on pages 120 and 121 of this essay.

it is true that the public meaning of terms differ based on Twin Earth style thought experiments, then intentional state ascriptions that involve those terms will differ in meaning and, therefore, the content of the respective intentional states will differ. I have given responses previously in this chapter that can answer this line of thought. We saw that we should allow some slack, as it were, between the language of an intentional state report and the content that report is supposed to express. If there is no distinction, though, between semantic externalism and content externalism, we can allow for no such slack. At times, philosophers do not make the distinction between these two sorts of externalism. For example, in her introduction to *New Essays on Semantic Externalism and Self-Knowledge*, Susan Nuccetelli writes, "Semantic externalism or anti-individualism is often cast as the rejection of semantic internalism, a view favored by philosophers at least since Descartes."²⁸ What reason do we have for accepting the distinction between these two forms of externalism?

A simple point is that these two types of externalism concern different subject matters. Content externalism is a thesis about intentionality and semantic externalism is a thesis about meaning. This point is not enough to show that the distinction between the content of an intentional state and the content of an intentional state ascription holds up, of course. It could be that we have good reason for thinking that the content of intentional state ascriptions always determine intentional content. If such a connection exists, then it does not appear to be possible to accept semantic externalism and reject content externalism. I do not think that we have good reasons for thinking that the content of intentional state ascriptions determine intentional content. In fact, reasons

²⁸ (Nuccetelli 2003), page 1.

were given above for thinking the opposite. Making sense of many types of speech (such as intentional and unintentional misuses of words) requires that we make a distinction between not only the meaning of what is said (sentence meaning) and what the speaker intends to convey (speaker meaning), but also the personal and intersubjective concepts that mirror sentence meaning and speaker meaning. Unless there is an argument for the conclusion that there is no distinction between the content of intentional state ascriptions and their respective intentional contents, there is no reason to think that one cannot adopt semantic externalism while rejecting content externalism. I have given reasons for thinking that we have good reason (namely, to explain many types of communication) for seeing these externalist theses as distinct.

CHAPTER 6 LIVING WITH CONTENT INTERNALISM

So far we have seen that first-person access to intentional content is incompatible with content externalism and that the arguments for content externalism were not as strong as perhaps was thought. These considerations support content internalism indirectly. At this point it is natural to turn to a more direct evaluation of content internalism.

The goal of this chapter is twofold. First, I aim to show that content internalism does not face insurmountable objections. Some have thought that if content internalism is true, then intentional content is not expressible in a public language.¹ If this objection were correct, then we would never be in a position to truly communicate our thoughts to others. This consequence of content internalism would perhaps be just as troubling as content externalism's conflict with first-person access to intentional content. If this objection is right, it would seem that going internalist about intentional content secures first-person access for subjects at the cost of making it impossible for subjects to understand each others' minds. We surely want a theory of intentionality to be consistent both with the ability of subjects to access their intentional contents in a first-person way and with their ability to know each other's minds. Another potential objection against content internalism is that it is a view of intentionality that leaves us unable to access the external world directly and, therefore, cut off from reality in a dangerous way. Due to this epistemic consequence, the argument concludes, content internalism should be rejected. I also address potential metaphysical worries by arguing

¹ For examples see: (Fodor 1987), (Horgan, Tienson, and Graham 2004), and (Loar 1988).

that content internalism is not only compatible with physicalism, but also that content externalism appears incompatible with certain versions of physicalism.

The second goal of this chapter is to sketch some accounts that explain how it could be that intrinsic properties are the only properties on which intentional content depends. I discuss the prospects for how intrinsic, phenomenal properties may play the role of the properties that ground intentional content. This account is somewhat lengthy. After discussing it, I suggest a way we may be able to adopt the key features of this account while rejecting the dependence of intentional properties on phenomenal properties. I do not, in the end, endorse a fully developed account of the properties that are responsible for intentional content. I merely sketch some options to give more shape to the direction in which our internalistic theory of intentionality may take. In the end, achieving my twofold goal of answering objections to content internalism and sketching accounts of how to understand the dependence of intentional content on intrinsic properties will hopefully make life with content internalism seem more livable than some may have previously thought.

Objections to Content Internalism

The Need for a Private Language to Describe Thoughts

There is a worry the content internalism entails that at some level the contents of intentional states are ineffable. If nominal clauses (such as that-clauses) are not adequate devices for determining content, then it may seem that non-linguistic, non-public devices must be used to describe intentional content. If all the resources we have for making our thoughts public are private resources, then there appears to be a struggle about how we can possibly make public what we are thinking. Thus, it may seem that if content internalism must be true, there is a real problem of seeing how it

could be true, given that we seem to be able to communicate our thoughts so well with others. Have we secured privileged access to content at the cost of walling this content off from the rest of the world?

Content internalists themselves are partially guilty of giving rise to this putative problem. Here is Brian Loar in “Phenomenal Intentionality as the Basis of Mental Content”:

Mental content is individuated more fine-grainedly than the interpersonally shared "oblique" content of certain that-clauses. I rather think the phenomenon is all-pervasive, that for virtually any that-clause a similar underspecification of content can be shown. A closely related point is this. Consider any perceptually nuanced conception of mine. I can invent a neologism to express that conception, and use it in self-ascribing that-clauses. But the that-clauses are then secondary: what matters is my reflexive grasp of the perceptual concept, *its* psychological content. That-clauses as they are standardly used apparently capture too little information, even on oblique interpretations, and that information is not of the right sort: that-clauses are more about socially shared concepts and their referents than about the various perceptually-based and other ways in which thoughts conceive their referents. They are not especially psychologically informative.²

This passage suggests, with its talk of neologisms, that semantic content, the content of public language vehicles, like words and sentences, is ill suited to describe intentional content. If nominal clauses are not appropriate vehicles for expressing our thoughts and the thoughts of others, one really begins to wonder what other options there are. As

Burge has put it,

There is at present no well-explained, well-understood, much less well-tested, individualistic language—or individualistic reinterpretation of the linguistic forms currently in use in psychology—that could serve as surrogate.³

² (Loar 2003), pages 229-230.

³ (Burge 1986), page 9.

It seems that if a new sort of language is required for individuals to articulate their thoughts if content internalism is true, then we have reason to be suspicious of content internalism. We seem faced with a problem. We accept that a given nominal clauses may not properly capture a subject's perspective. We admit that there is no individualistic, private language. We appear to lack a system for properly expressing intentional content. It appears we have merely inverted our problem. First, it appeared that the truth of content externalism entailed that we do not know our own minds in a special way. Now it appears that the truth of content internalism entails that no one else can ever truly know our minds, because we do not have a language capable of properly expressing the narrow content of our thoughts.

Many philosophers have taken this worry seriously. My response to it should not come as much of a surprise. I noted that there is a distinction between semantic externalism and content externalism. We should keep in mind that these theses concern different subject matters. The above worry arises in part from a failure to distinguish the two theses. If one thinks that semantic internalism is a non-negotiable part of the package of content internalism, then one may run into the troubles outlined above, for we do not allow for the possibility that public meanings can describe narrow content. However, as long as we understand that we can accept semantic externalism and content internalism, then we are part of the way to dissolving the above worry. The other point we need to keep in mind is that nominal clauses that attribute content in intentional state attributions *can* capture individualistic content, although it may require some work. The idea here is that while there are obviously cases in which a given nominal clause does not attribute the right content because it is not sensitive to the

subject's perspective, there is *some* nominal clause that can do the job. Recall Pierre who assents to the sentences, "Londres est jolie" and "London is ugly." If I attribute the beliefs to him that London is pretty and that London is not pretty, I will have clearly attributed beliefs to him that do not capture his perspective, because these beliefs do not capture the concepts Pierre associates with "London" and "Londres." From this, it seems reasonable to conclude that the most obvious nominal clause an interpreter can use to attribute a thought does not always properly express a subject's thought. We need a further argument, though, to get to the conclusion that nominal clauses are in principle not suited for attributing intentional content to thinkers. As John Biro notes, "from the fact that the wrong that-clause does not capture what someone has in mind, it does not follow that some other that-clause could not."⁴

There is no reason to think that nominal clauses, which have their content externalistically, cannot attribute internalistically individuated intentional content. It may be true that the obvious nominal clause will not do at times, but there are other nominal clauses that can properly attribute the content, in the sense that the content attributed is fine-grained enough to capture the subject's perspective.⁵ Let us consider Burge's arthritis case in order to better see the point. Burge suggests that we should use the term "arthritis" in belief ascriptions about the individual who believes he has arthritis in his thigh. Of course, this is, as we have already discussed, what we typically do. We typically use a subject's own words to attribute intentional states to him. It is very natural

⁴ (Biro 1992), page 288.

⁵ Of course, there are cases that involve demonstratives that require a certain amount of contextual information in order to attribute the proper content. Let us say that you are in London, and I say that Pierre believes that this city [as I point to the ground] is pretty. To know what city is referred to by the demonstrative, you must have certain contextual information. In this case, you must know what city you are in.

to imagine us saying that he believes he has arthritis in his thigh. This, however, does not capture the content of the intentional state as well as the following attribution: "He believes that he has a rheumatoid ailment in his thigh." He of course believes that this rheumatoid ailment is called "arthritis," but this metalinguistic belief can be factored out of the belief he has about the pain in this thigh. To capture the subject's perspective we should avoid using the term "arthritis." After all, if we did use the term "arthritis," we would run into a problem of attributing obviously contradictory beliefs to him. After the doctor tells him that the pain in his thigh cannot be arthritis because arthritis occurs only in the joints, we would not want to say of him that he believes both *that he has arthritis in an area outside of his joints* and *that arthritis is a disease exclusively of the joints*. Not recognizing that there are times when the obvious choice of words in an intentional state ascription will fail to capture the intentional content of the state can lead us to having a much too restrictive view of the resources we have for intentional state attributions. On the other hand, denying the role that nominal clauses have in attributing intentional content is implausible and leaves one puzzled over how our thoughts can be truly described by others. Taking a more cautious approach, one that recognizes both the necessary role that nominal clauses (and their dependence on extrinsic properties of thinkers for their meaning) play in intentional content attribution and the fact that there is always some nominal clause or other available to describe intentional content, allows us to avoid the consequence that content internalism entails that the contents of thoughts are ineffable.

Paving the Route to Skepticism about the External World?

In Chapter 3, I quoted Michael Luntly in what seems to be a good representation of the general worry that content internalism severs our minds from external reality in a problematic way.

Although representationalism is a safeguard against idealism, having got a specification of representations independently of what they stand for, the representationalist's problem is to show how, given their constitutive separateness from the world, they nevertheless have the property of being about the world. It is not clear that this makes sense. Despite the ease with which many theorists assume that they have safeguarded realism by endorsing representationalism, they do so at the risk of leaving our thoughts (representations) wholly out of touch with the world.⁶

This seems to be a general worry about content internalism. Ever since Descartes's *Meditations* much philosophical work has been motivated by a desire to avoid skepticism. Part of what allows Descartes skepticism about the external world to get a foothold is his theory of ideas, which is internalistic in that a thinker's having a certain idea does not require the environment being a certain way. So, it is not difficult to see how content internalism can seem to leave us vulnerable to skepticism about knowledge of the external world. After all, this is the view of intentional content that underlines one of the greatest sources of skepticism about our knowledge of the external world, Descartes's *Meditations*. If the contents of our thoughts do not depend on the environment being a certain way, if they are truly independent of how the world is, then our thoughts are logically independent of the environment. Thus, it is possible that the external world is not as it appears to be in our representation of it. Davidson sums up this well in "The Myth of the Subjective."

⁶ (Luntley 1999), page 15.

If the ultimate evidence for our schemes and theories, the raw material on which they are based, is subjective in the way I have described, then so is whatever is directly based on it: our beliefs, desires, intentions, and what we mean by our words. Though these are the progeny of our 'view of the world'—indeed, taken together, they constitute our view of the world—nevertheless they too retain the Cartesian independence from what they purport to be about that the evidence on which they are based had: like sensations they could be just as they are, and the world be very different. Our beliefs purport to represent something objective, but the character of their subjectivity prevents us from taking the first step in determining whether they correspond to what they pretend to represent.⁷

We saw that content externalism has implausible epistemic consequences. It my intention to clear up any suspected negative epistemic consequences of content internalism, so as to not leave us in the paradoxical position of being saddled with implausible epistemic consequences for both views of intentional content. Far from adequately addressing the larger debate of how content internalism and content externalism affect our ability to have knowledge of the external world, my goal in the next section is much more modest. I want to show two things.⁸ First, I want to show that the worries about content internalism can be addressed. Second, I want to remind the reader that content externalism has its own difficulties with regard to our knowledge of the external world.

Content internalism and skepticism about the external world

If it is true that content internalism does not rule out skepticism about the external world, then turning this fact into an argument will require the content externalist to show

⁷ (Davidson 2001), page 43.

⁸ There has been a large discussion of whether content externalism gives us *a priori* knowledge of the empirical world, following Michael McKinsey's "Anti-Individualism and Privileged Access." The idea is that if we can know the empirical content of our thoughts in an *a priori* fashion, then we can gain *a priori* knowledge of the empirical world. However, this seems to be absurd. Thus, we have a *reductio ad absurdum* of content externalism.

that his view can better handle the challenge of external world skepticism. How, then, does content externalism help us avoid skepticism about the external world?

Let us be clear what sort of skepticism we are worried about. The worry appears to one of global skepticism that Descartes invites in with his *Meditations*: the possibility that the external world is radically different from how we take it to be, and therefore seemingly outside of the scope of our knowledge. If my beliefs depend on percepts that are systematically false, then it appears that all of my beliefs will be false as a result. Content internalism fails to foreclose this possibility, of course. In the Cartesian fashion, we wonder what we can know of the world if we leave out knowledge that requires us to make an inference about the external world. In addition to the basic Cartesian knowledge that one is thinking and that one exists, what else can one add to one's body of knowledge? It seems to me that the situation is not as hopeless as Descartes allows. For example, we saw that there are necessary conditions on thought. Recall the thesis of transcendental externalism.

The thesis of transcendental externalism: The existence of the contents of intentional states require the instantiation of background conditions, conditions which instantiate extrinsic properties.

If we can know this by dint of philosophical argument, then it appears that we have a way to break out, as it were, of the intentional content which supervenes solely on intrinsic properties. In short, the contents of particular intentional states guarantee nothing about the portion of the external world they purport to represent; however, the existence of these contentful states guarantees *something* about the external world. But, again, what is this something?

If one accepts Davidson's idea that triangulation is a necessary condition of thought, then we can be guaranteed that there is another individual on the scene, and

that there is a scene—a shared environment. This, though, hardly secures knowledge of the external world in the way we would want. We can know that there is an external world, but we do not know much of its nature, or whether or not our representations of it are accurate. In addition to this idea of triangulation, Davidson has also advanced an argument for the conclusion that we cannot be massively mistaken about the world in the way that global skepticism suggests is possible. He lays this out most fully in “A Coherence Theory of Truth and Knowledge.” The basic idea is that the publicity of language and the nature of belief give us the conclusion that most of our beliefs are true. The starting point for seeing the truth of this conclusion is the viewpoint of an interpreter. In interpreting the speech and thought of another, Davidson thinks that we employ the principle of charity. Here is what Davidson says of the principle of charity.

The principle directs the interpreter to translate or interpret so as to read some of his own standards of truth into the pattern of sentences held true by the speaker. The point of the principle is to make the speaker intelligible, since too great deviations from consistency and correctness leave no common ground on which to judge either conformity or difference. From a formal point of view, the principle of charity helps solve the problem of the interaction of meaning and belief by restraining the degrees of freedom allowed belief while determining how to interpret words.⁹

The interpreter must, to make the speaker intelligible, interpret in a way that maximizes the truth of the speaker’s beliefs. If it is true that thought requires a share environment between thinkers, then it appears that the nature of interpretation entails that most of the beliefs of the thinkers will be true. If this is so, then it appears that we cannot be massively mistaken about the nature of the external world. Of course, knowledge is typically thought of as something more than merely true belief. Davidson says that beliefs are justified in the sense that “they are supported by numerous other beliefs

⁹ (Davidson 2001), pages 148-149.

(otherwise they wouldn't be the beliefs they are), and have a presumption in favor of their truth."¹⁰ Justification for beliefs then seems to fall out of the holism of intentional content; intentional states have their content in virtue of their relation to other intentional states. On Davidson's view, beliefs are the only sorts of entities that can justify other beliefs. He writes, "What distinguishes a coherence theory is simply the claim that nothing can count as a reason for holding a belief but another belief" (ibid: 141). From his holism about intentional content, then, the beliefs required to count as reasons are already standing at the ready. We seem to get the justification for a certain belief in virtue of the other beliefs that are by their nature true and required to give the certain belief the content that it has.

Perhaps this rather complicated story is a possible response to skepticism about the external world that a content internalist who also accepts transcendental externalism can adopt. It requires other theoretical commitments that we cannot get into now. And, of course, to be fair there are other responses to skepticism to which the content internalist can appeal. There are contextualist responses. There is the Moorean response. It is far from settled that adopting content internalism leads to radical skepticism about our knowledge of the external world. It is certainly an assumption that drives the traditional argument for skepticism about the external world; however, there are well-known ways for the content internalist to resist skepticism of the external world.

Content externalism and our knowledge of the external world

Sometimes it is seen as a virtue of content externalism that it rules out the possibility of skepticism about the external world (see Putnam "Brains in a vat" in his

¹⁰ (Davidson 2001), page 153.

1981, for an example). On the other hand, some thinkers have noted that content externalism has implausible consequences for our knowledge of the external world. In short, it has been argued that the conjunction of content externalism and *a priori* access to intentional content entails that we can have *a priori* knowledge of the external world, which is an absurd consequence. In “What the Externalist Can Know A Priori,” Paul Boghossian argues that because content externalism individuates concepts in terms of the extensions of the words which express those concepts, the thinker who has *a priori* access to his thought contents can—based on *a priori* investigation alone—gain an empirical fact about the world, namely that whatever is in the extension of the word that expresses the concept that is a constituent of his thought exists. This argument has been much discussed, and a proper treatment of it would require more space than I can allot it here. However, bringing it up is enough to show that content externalism has its own difficulties when it comes to the dialectic about skepticism about the external world. I do not wish to go into these in too much depth, for we have already seen that content externalism is incompatible with first-person access to intentional content. Once we acknowledge that the truth of first-person access to intentional content is required for rationality, we should seek to adopt a theory of the independence of intentional content from the environment that preserves first-person access to intentional content. If the content externalist has problems squaring his view with both rationality and our knowledge of the external world, then it seems to be in trouble indeed. We saw that the content internalist can respond to the skeptic. At this point in our investigation, then, the charge that content internalism is more susceptible to skepticism about our knowledge of the external world seems to lack much force, if content externalism has its own

problems with regard to our knowledge of the external world and with regard to first-person access to intentional content.

What about Physicalism?

The popularity of physicalism warrants a discussion of its compatibility with the two views of intentional content we have been discussing. In particular, I want to discuss arguments intended to show that content externalism is incompatible with various versions of physicalism. Exploring these arguments can perhaps further mollify concerns about adopting content internalism. Some philosophers have point out that content externalism is incompatible with the token-identity theory of mind, sometimes understood as the view that every particular mental state is identical to a physical state of the brain.¹¹ The argument, roughly, runs as follows. All mental states are brain states. Brain states depend on properties that are intrinsic to their owners. However, if certain intentional states instantiate extrinsic properties, then they are not merely a function of intrinsic properties. Therefore, certain intentional states are not brain states or they are not mental. This conclusion forms a dilemma for the token-identity theorist who accepts content externalism. Neither horn is acceptable.¹² The token-identity theorist does not want to save things by saying that intentional states are not in fact brain states, because

¹¹ See (McGinn 1989) and (Burge 1986, 1993) for examples of arguments for incompatibility.

¹² Some have noted that this argument assumes content essentialism, the view that intentional states have their contents essentially (see (Frances 2007) for an example). This thesis should strike us as obviously true. We have been assuming its truth for the entire essay. One way to articulate this thesis is: intentional states are to be individuated by their contents, although there may be other factors used to individuate intentional states (such as intentional mood, that is, whether it is a belief, desire, hope, etc.). The idea is that my belief that Antananarivo is the capital of Madagascar has its content essentially. If the content were different, then the belief would be different. It is hard to see how such a thesis could be rejected. If it is to be rejected to preserve the token-identity theory, then I suggest that we can take this a *reductio ad absurdum* of the token-identity theory.

that seems to contradict his thesis that all mental states are brain states. Of course, saying that intentional states are not mental states is one way to save the view, but this comes at a great cost. Intentional states seem to be paradigmatic mental states.

Donald Davidson responds to this argument in “Knowing One’s Own Mind.” He thinks that a token-identity theorist like himself can accept externalism and the token-identity theory. The argument for the incompatibility, Davidson argues, involves the false assumption that “if a thought is identified by a relation to something outside the head, it isn’t wholly in the head.”¹³ Davidson rejects this assumption.

It should be clear that it doesn’t follow, simply from the fact that meanings are identified in part by relations to objects outside the head, that meanings aren’t in the head. To suppose this would be as bad as to argue that because my being sunburned presupposes the existence of the sun, my sunburn isn’t a condition of my skin. My sunburned skin may be indistinguishable from someone else’s skin that achieved its burn by other means (our skins may be identical in the ‘necktie sense’); yet one of us is really sunburned and the other not.¹⁴

We should proceed with some care here. First, let us note that the argument for the incompatibility of content externalism and the token-identity theory of mind does not require the assumption that “if a thought is identified by a relation to something outside the head, it isn’t wholly in the head.” If this assumption were true, then virtually all thought would be externalistically individuated. The theses of content externalism and content internalism are not about how thoughts are identified but about the sorts of properties on which a subject’s thoughts supervene. All thoughts are identified—at some level—by way of public language, so if one holds the assumption above, then all

¹³ (Davidson 2001), page 31.

¹⁴ (Davidson 2001), page 31.

thought is outside of the head. Even if this is right, though, perhaps there is still something to Davidson's response.

In a sense Davidson's point is well taken. Simply because intentional content is externalistic does not mean that content is "outside of the head." Content externalism is a thesis about the dependence of intentional content on the environment. A content externalist does not need to appeal to talk of thoughts being inside or outside of the head to state the thesis. The thesis is one of dependence, typically articulated using the notion of supervenience, a thesis that does not by itself have any implications for the locations of thoughts.

However, one who accepts content externalism and holds the view that token mental states are identical to token physical states is committed to a thesis about the spatial location of thoughts. The reason for this is fairly straightforward. If one holds that mental states are token-identical to physical states, and intentional states are mental states, then intentional states are identical to physical states. That much follows from the token-identity theory of mind. When combined with content externalism, we get the conclusion that intentional states supervene on extrinsic properties of thinkers, and therefore supervene on physical states that are not instantiated by merely intrinsic properties of thinkers. Not only do the intentional states supervene on these physical states, but also according to the token-identity theorist, they are identical to them. These physical states with which intentional states are identical will be spatial located outside of the head of the thinker. Thus, thoughts in a very real and non-metaphorical sense will be outside of the head. Davidson is surely right that content externalism by itself does not entail that thoughts are outside of the head. However, conjoining content

externalism with the token-identity theory of mind does have that consequence.

Conjoining such views, then, is a mistake.

Of course, there are other views of physicalism that differ from the token-identity theory of mind. The problem described above will be a problem for physicalists who hold the following claims.

- **The thesis of content externalism:** Intentional content supervenes on extrinsic properties of thinkers.
- **The thesis of content essentialism:** Intentional states and events are individuated in terms of their content.
- The claim that intentional events are mental events.
- **The thesis of reductive physicalism:** Mental events are reducible to physical events.

Accepting all these claims is a problem because we can derive the conclusion that intentional events supervene on and are reducible to physical events which instantiate extrinsic properties of thinkers, and this leads to the dilemma just discussed. There seems to be real problems for those who wish to hold content externalism, content essentialism, and some form of reductive physicalism on which all mental events are reducible to physical events.

Content internalism does not create these problems for reductive physicalism, and is friendly to the project of reducing the mental to the physical. If the physicalist adopts content internalism, he has a tidy story about how intentional states are causally efficacious mental states. The story goes like this. Intentional states have their contents essentially.¹⁵ These contents supervene of the intrinsic properties of their owners.

¹⁵ My desire to finish the marathon would be a different desire if it had a different content, for example.

These intrinsic properties on which content supervenes are or are reducible to physical properties of thinkers (most stories will mostly have it that these are properties of the brain). The dilemma that content externalism forces on some varieties physicalism — that intentional events are actually physical tokens outside of a thinker's head or they are not mental—is not a problem for content internalism.

Approaches to Internalistic (Narrow) Content

So far we have seen that there are responses to the major objections to content internalism, and that content internalism is compatible with reductive physicalism, whereas content externalism is incompatible with certain forms of reductive physicalism. Adopting content internalism means one is committed to the view that intentional content depends only on a thinker's intrinsic properties. There is a natural curiosity about which of a thinker's intrinsic properties are the ones on which intentional content depends. In this section, I shall sketch two directions such an account can take, one that grounds intentional content in phenomenal properties, and one that does not. I argue that we should take the latter approach because it is more conservative. First, I shall sketch a theory we can call "Phenomenal Intentionality, " which grounds the nature of intentionality in phenomenal consciousness.

Phenomenal Intentionality

In recent years, it has not been uncommon for philosophers to see intentionality as a phenomenon that could be studied separately from phenomenal consciousness. The philosophical problems of intentionality, it was thought, could be pursued without reference to phenomenal consciousness and vice versa. Colin McGinn sums this attitude up well in "Consciousness and Content."

Recent years have witnessed a curious asymmetry of attitude with respect to these two problems [a naturalistic account of subjectivity and mental representation]. While there has been much optimism about the prospects of success in accounting for intentionality, pessimism about explaining consciousness has deepened progressively. We can, it is felt, explain what makes a mental state have the content it has; at least there is no huge barrier of principle in the way of our doing so. But, it is commonly conceded, we have no remotely plausible account of what makes a mental state have the phenomenological character it has.¹⁶

The view I shall be exploring suggests that there is a much tighter connection between phenomenal consciousness and intentionality than had previously been thought by many. In “Phenomenal Intentionality Without Compromise,” Katalin Farkas gives a good characterization of the view of phenomenal intentionality: “the intrinsic directedness of certain conscious mental events which is inseparable from these events’ phenomenal character.”¹⁷ This harkens back to Brentano’s definition of intentionality as the mind’s directedness to its objects. We are all of course familiar with this phenomenology of occurrent intentional events. Our intentional mental events seem to be directed upon objects. When I think, *President Obama is doing his best to lift us out of the recession*, my thought seems to be directed upon Obama and his efforts. What is the relationship between the phenomenology of my intentional states (how my intentional states seem to me) and their intentionality (the intentional content they have)?

Charles Siewert outlines four positions one can take regarding the relationship.

- (a) Consciousness is explanatorily derived from intentionality.
- (b) Consciousness is underived and separable from intentionality.
- (c) Consciousness is underived but also inseparable from intentionality.
- (d) Consciousness is underived from, inseparable from, and essential to intentionality.^{18,19}

¹⁶ (McGinn 1988), page 295.

¹⁷ (Farkas 2008a), page 273.

¹⁸(Siewert 2006).

By “consciousness” Siewert means phenomenal consciousness. In this section, we shall be interested in how phenomenal properties could ground intentionality, so we shall be interested primarily in (d). I shall assume that phenomenal properties are intrinsic properties in the remainder of this chapter.

John Searle’s defense of the connection principle seems to be one argument that can support (d) and help us to understand how intentional content could depend on phenomenal properties. The connection principle does not directly state the relationship between phenomenology and intentionality; however, the support that Searle provides for the principle provides us with a framework for exploring how phenomenal consciousness might be essential to intentionality. The connection principle is the thesis that “all unconscious intentional states are in principle accessible to consciousness.”^{20,21}

Searle claims that the conclusion depends on the following premises.

- (P1) There is a distinction between intrinsic intentionality and as-if intentionality; only intrinsic intentionality is genuinely mental.
- (P2) Unconscious intentional states are intrinsic.
- (P3) Intrinsic intentional states, whether conscious or unconscious, always have aspectual shape.

¹⁹ Here I use “phenomenology” to the manner in which our intentional states are presented to us. The phenomenology of an intentional state is related to phenomenal consciousness. In so far as an intentional state is presented to us—in so far as it is made occurrent—it will be a phenomenally conscious mental state.

²⁰ (Searle 1992), page 156.

²¹ Searle offers other formulations. Here is one such:

Specifically, we understand the notion of an unconscious mental state only as a possible content of consciousness, only as the sort of thing which, though not conscious, and perhaps impossible to bring to consciousness for various reasons, is nonetheless the sort of thing that could be or could have been conscious (Searle 1991), page 51.

- (P4) The aspectual feature cannot be exhaustively or completely characterized solely in terms of third-person, behavioral, or even neurophysiological predicates. None of these is sufficient to give an exhaustive account of aspectual shape.
- (P5) But the ontology of unconscious mental states, at the time they are unconscious, consists entirely in the existence of purely neurophysiological phenomena.

From these premises he thinks that the connection principle follows.

- (C1) The notion of an unconscious intentional state is the notion of a state that is a possible thought or experience.

From (5) and the connection principle as it is expressed in (C1), Searle thinks that we can conclude the following.

- (C2) The ontology of the unconscious consists of objective features of the brain capable of causing subjective conscious thoughts.²²

There are a number of issues that are running through this argument. I will address the argument directly in a later section. For now, I want to use it as a backdrop for an argument for the conclusion that intentional content depends on phenomenal properties.

Why might intentional contents depend on phenomenal properties?

(P1), (P3), and (P4) seem to be candidate components that can play a role in an answer to this question. (P1) is a thesis about what sorts of entities can have intrinsic intentionality. Searle clearly thinks that only minds can have intrinsic intentionality. A painted portrait of someone has *as-if* intentionality, the story goes, because only the painter has the ability to represent the person. The picture can clearly represent the person, but it does so only in virtue of deriving its intentionality from the intrinsic intentionality of the painter's intentional states. (P3) and (P4) concern aspectual shape. This is Searle's term for capturing how the conditions of satisfaction of an intentional

²² (Searle 1992), pages 156-160.

state are presented to a subject. This idea corresponds to Frege's idea of a mode of presentation. I shall use these terms as more or less interchangeable throughout this essay: as capturing how something is presented to a subject. It is undeniable that we can think of the same object under different modes of presentation or, as Searle says, "under some aspects and not others" (ibid: 157). I think it is true that all phenomenally conscious intentional states have aspectual shape. The truth of (P4), as stated above, is less obvious. Let us consider an example that Searle gives.

A person may indeed exhibit water-seeking behavior, but any water-seeking behavior will also be H₂O-seeking behavior. So there is no way the behavior, construed without reference to a mental component, can constitute wanting water rather than wanting H₂O. Notice that is not enough to suggest that we might get the person to respond affirmatively to the question "Do you want water?" and negatively "Do you want H₂O?", because the affirmative and negative responses are themselves insufficient to fix the aspectual shape under which the person interprets the question and the answer. There is no way just from the behavior to determine whether the person means by "H₂O" what I mean by "H₂O" and whether the person means by "water" what I mean by "water." No amount of behavioral facts constitute the fact that the person represents what he wants under one aspect and not under another.²³

It seems true that a purely behavioral approach to the interpretation of the desire described above is not sufficient for determining the content of the desire. Reference must be made to a mental component in order to properly determine the content of the desire. However, it seems that we can get a grip on the subject's concept of water by asking him the sorts of questions Searle outlines above. If we get a proper answer to the questions, then it seems that we can fix the aspectual shape under which the person desires water. If he says, "Yes, I want water, but I don't care whether or not it is H₂O. I just want some damn water to drink," then it seems that we can grasp the

²³ (Searle 1992), page 158.

aspectual shape of the desire. Sufficient communication with the subject can yield the aspectual shape in question. Interpreters are not in principle blocked from understanding the way in which subjects think of something.

At any rate, we can use (P1), (P3) and a revised version of (P4)—(P4')—in an argument for the fundamental role that phenomenal properties have in grounding intentionality. Here are some key premises in our revised argument.

- (P1) There is a distinction between intrinsic intentionality and as-if intentionality; only intrinsic intentionality is genuinely mental.
- (P3) Intrinsic intentional states, whether conscious or unconscious, always have aspectual shape.
- (P4') The aspectual shape of a thinker's intentional states supervenes of the phenomenal properties of his mental states.

We also need some further claims, claims which rule out other possible groundings of intentionality. We need to establish a connection between intrinsic intentionality and aspectual shape. It may seem to us, as it does to Searle, that human minds, endowed as they are with phenomenal consciousness, are the only entities capable of having intrinsic intentionality. However, we need something more than a mere seeming here.

We can fill in the gap in the argument by focusing on a point from our discussion of content externalism, a point related to Searle's (P4). We saw that often times neglecting the subject's perspective can fail to capture the nature of his intentional content. If we do not recognize that Kripke's Pierre is clearly operating with two concepts, one of the London from his experiences and one of Londres from the stories of his youth, then we are liable to attribute to him beliefs about the same city, when in fact his beliefs are not about the same city. In short, attributing intentional states without an eye to the subject's perspective attributes intentional states that are not fine-grained

enough. Only when we account for aspectual shape—or the subject’s perspective of the intentional content in question—can we attribute intentional states precisely enough. In a useful discussion of the research program of phenomenal intentionality, Terry Horgan and Uriah Kriegel offer the following articulation of this idea.

One recurrent claim is that only phenomenal intentionality has *determinate content* in and of itself (Searle 1991, 1992, Loar 1995, Horgan and Tienson 2002, Strawson 2004, Horgan and Graham 2009). We may state this thesis as follows:

Determinate Content. Necessarily, for any intentional state M with content C, if C is non-derivatively determinate, then M is phenomenally intentional.²⁴

Searle makes this point by stressing that it is only phenomenally conscious intentional states or potentially phenomenally conscious intentional states that are capable of having intrinsic intentionality. Adding the thesis of determinate content to our argument, we have the following.

- (P1) There is a distinction between intrinsic intentionality and as-if intentionality; only intrinsic intentionality is genuinely mental.
- (P3) Intrinsic intentional states, whether conscious or unconscious, always have aspectual shape.
- (P4') The aspectual shape of a thinker’s intentional states supervenes of the phenomenal properties of his mental states.

The thesis of determinate content: All contentful intentional states have determinate content solely in virtue of their phenomenal properties or aspectual shape.

From these premises, we can draw the following conclusion.

- (C3) That intentional content of a thinker’s intentional state is determinate depends on phenomenal properties of a thinker.

²⁴ (Horgan and Kriegel 2011)

The support from the thesis of determinate content comes from two sources: first from reflecting on the how the thinker's perspective—his way of thinking of an object—is essential to determining the content of his intentional state, and second from the idea that his way of thinking of an object supervenes on his phenomenal properties. If we ignore the aspectual shape, we can still determine some intentional content, but such content will not be sufficiently determinate. It will not be sufficiently precise.

How should we understand unconscious intentional states and phenomenal consciousness?

Even if the above argument for (C3) is successful, we still have not shown that all intentionality is grounded in phenomenal consciousness, for there are of course many unconscious intentional states. I believe that Tallahassee is the capital of Florida even when I do not consciously have this belief before my mind. Phenomenal properties may be responsible for the content of our phenomenally conscious intentional states, but what about those states which are not phenomenally conscious?

Here we need to answer the question posed earlier about how we should understand the connection between unconscious intentional states and phenomenally conscious intentional states. If the connection principle is true, then it looks as if we could have full support for Siewert's (d): Consciousness is underived from, inseparable from, and essential to intentionality. Let us turn back to Searle's argument and the conclusions he draws from it to measure his support of the connection principle.

- (P1) There is a distinction between intrinsic intentionality and as-if intentionality; only intrinsic intentionality is genuinely mental.
- (P2) Unconscious intentional states are intrinsic.
- (P3) Intrinsic intentional states, whether conscious or unconscious, always have aspectual shape.

- (P4) The aspectual feature cannot be exhaustively or completely characterized solely in terms of third-person, behavioral, or even neurophysiological predicates. None of these is sufficient to give an exhaustive account of aspectual shape.
- (P5) But the ontology of unconscious mental states, at the time they are unconscious, consists entirely in the existence of purely neurophysiological phenomena.

From these premises we are supposed to be able to derive (C1).

- (C1) The notion of an unconscious intentional state is the notion of a state that is a possible thought or experience.

(C1) expresses the connection principle. It is worth evaluating the success of this argument. From (P1), (P2), and (P3), we can conclude that all intrinsic intentional states—those states that do not have their content derivatively—have aspectual shape. How do we get from this claim—that all intentional states have aspectual shape—to the claim that all unconscious intentional states are potentially conscious intentional states? What else is needed to show us that if a state of a thinker is not potentially conscious, then that state is not mental and not intentional?

(P4) is supposed to show us that aspectual shape cannot be described by third-person or scientific descriptions of mental states. Yet, these appear to be the only descriptions available to us, because, as (P5) states, unconscious mental states are constituted by the phenomena that the types of predicates mentioned in (P4) pick out. Searle notes that this is a puzzle. He posits that accepting the connection principle can solve the puzzle. This form of argument should worry us. We set out to argue for the connection principle. We reached a puzzle: aspectual features are not describable in neurophysiological predicates or other “third-person” predicates, and these seem to be the only predicates applicable to unconscious mental states, as unconscious mental states are a “purely neurophysiological phenomena.” It certainly seems to beg the

question on Searle's part to merely state that a solution to this puzzle can be reached by adopting the connection principle. After all, someone could equally say that based on (P1)-(P5) we must conclude that unconscious states do not have aspectual shape. However, if we conclude this, then we must reject that unconscious intentional states have intrinsic intentionality.²⁵ It appears that Searle has two options here, neither of which is desirable. First, he can posit (C1) as a solution, but that begs the question. Second, in light of the puzzle, he can admit that there is no way that unconscious states have aspectual shape because the predicates of neurophysiology do not capture aspectual shape. However, this forces him to reject other key theses. For instance, if he holds that unconscious intentional states lack aspectual shape, then he must reject that unconscious intentional states are intrinsic.

Are there any sound arguments for the connection principle?

Kirk Ludwig offers another argument for the connection principle.²⁶ I think it is worth exploring that argument before we try to fix Searle's, for we are still attempting to justify the thesis that "consciousness is underived from, inseparable from, and essential to intentionality." Ludwig's argument for his version of the connection principle focuses on what is required for determining ownership of unconscious mental states. Ludwig endorses the following version of the connection principle:

CP: Nothing is a mental state unless it is a conscious state or it is a disposition to produce a conscious mental state.²⁷

²⁵ See (Ludwig 1993) for criticism of Searle.

²⁶ (Ludwig 1996a).

²⁷ (Ludwig 1996a), page 31.

To argue for this conclusion, Ludwig first endorses the thesis that a mental state is identified as X's mental state just in case X has first-person knowledge of that state. He writes,

One's having first person knowledge of a particular mental state is sufficient for it to be one's own mental state and sufficient for it to be no one else's mental state. Thus, in the case of conscious mental states, we can say that a token conscious mental state is X's rather than Y's because X has first person knowledge of it.²⁸

What is meant by "first person knowledge" here? Ludwig writes the following.

One has a kind of knowledge of one's own conscious mental states at the time at which they are conscious which no else could have of those mental states. This difference in the kind of knowledge we have of our own and other people's conscious mental states is well illustrated in the methodology of investigations of perception. Contrast the way we find out how a thing looks to ourselves and how it looks to someone else. In our own case, we do not have to ask ourselves for a report of how a thing looks to know how it looks, or to see this by some observation of our behavior. In the case of other subjects, however, we have no access to how things look to them other than by their reports about it or what differences it makes to their behavior or performance on various tasks we set them. I will call this kind of knowledge we have of our own mental states that no one else does or could have first person knowledge.²⁹

The point here seems to be about access and not about knowledge. Indeed, this point is similar to the thesis of first-person access to intentional content, although it is about not merely intentional states, but all mental states. The point seems to be that the access we have to our mental states enables us to know them in a special way. The most basic point here is about access.

After noting that X's first-person access to mental state Y is sufficient for Y being X's mental state, Ludwig goes on to point out that first-person access will not help

²⁸ (Ludwig 1996a), pages 31-32.

²⁹ (Ludwig 1996), page 31.

determine ownership of unconscious mental states, for there is no direct access to them. If there were, then they would not be unconscious. Assuming that unconscious mental states belong to individuals, then, there are three options for what could assign ownership to such states. The first option is that it is a brute fact that an unconscious state belongs to X, a fact which admits of no explanation. The second option is that the unconscious state is X's because that state is "causally located in his body." The third option, the one Ludwig prefers, is that "a token unconscious mental state is a particular person's mental state because it bears a special relation to that person's conscious mental states." He argues that the third option is the only viable one. In leaving the connection between unconscious states and persons unexplained, the first option does not rule out potential connections between certain unconscious states and owners that we want intuitively to rule out. For example, it does not foreclose the possibility that one of your unconscious states is mine. For the second option we must determine what counts as some person's body. This becomes problematic, as Ludwig notes, because "the only way to do that is to appeal to its relations to his mental states. But then mental states being causally located in his body could not ground those mental states as his, because what makes it his body is that *his* mental states are causally located in it." The support for this claim in Ludwig's paper is not lengthy, although the idea appears to be that there is nothing strong enough to ground one's bodily identity other than mental states. Any attempt that does not reference mental states in specifying what counts as one's body specifies something merely contingent. The idea appears to be a Cartesian one: one is not related to one's body essentially. It is possible for one to be a mind without a body. The only thing that can determine that body Y is X's body is that Y is

causally related to mental states to which X has first-person access. We can be sure that mental states to which X has first-person access are X's. We can only understand other states to be X's in so far as they are causally related to mental states which are essentially X's: those states to which X has first-person access. The third option is to hold that unconscious, dispositional states are X's in so far as they are dispositional states which when made manifest are such that only X would have first-person access to them. This, I think, is a plausible way to show that phenomenal consciousness is what grounds all unconscious intentional states. For a state of an individual to count as a mental state it must either be a conscious state or a state that is potentially conscious. Restricted to intentionality, we can use this articulation of the connection principle to show that it is necessary that for a state of an individual to count as his intentional state it must either be conscious or a disposition to produce a conscious intentional state (or, in other words, a potentially conscious intentional state).

We have discussed Searle's support for the connection principle to set-up two different arguments. The first argument was adopted from some of the premises that Searle endorses in his argument for the connection principle. This argument does not appeal to Searle's questionable (P4). Rather, the argument appeals to the role that phenomenal properties have in determining aspectual shape. The aspectual shape of intentional states is what gives intentional states determinate content. Thus, determinate content depends on aspectual shape. Aspectual shape depends on phenomenal properties. Thus, phenomenal properties ground intrinsic intentionality, in the sense that if a state does not have determinate content, then it does not qualify as intrinsic, and intrinsic intentional states are the most fundamental intentional states.

The second argument appeals to the role that phenomenal consciousness has in grounding unconscious intentional states, which must be potentially conscious or else we have no way to determine whose they are. We are surely able to specify the owners of unconscious intentional states. Thus, there is a real sense in which these states are grounded in phenomenal consciousness. Thus, I endorse CP, and Ludwig's argument for it. We can combine (C3)—the view that the fact that intentional content of a thinker's intentional state is determinate depends on phenomenal properties of a thinker—with CP to conclude that a thinker's intentional content, both of conscious and unconscious intentional states, depends on his phenomenal properties. Call the view that a thinker's intentional content depends on his phenomenal properties "dependence of intentional content on phenomenal properties."

Intentionality and Aspectual Shape

Even though the view above that grounds intentional content in phenomenal properties is gaining popularity, it is being still met with a good deal of skepticism. After all, as we have discussed, it runs contrary to a commonplace understanding of the bifurcation of the study of intentionality and phenomenal consciousness. It is important to remember that it is not the only account of narrow content. "The Phenomenal Intentionality Research Program," as it has been called by Horgan and Kriegel, provides one explanation for how intentional content depends on intrinsic properties. It would be quite controversial, though, if it were the only such account, as it does require a rather drastic revision of the typical way of conceiving of the connection between phenomenal consciousness and intentionality. It would be good to have an account of the sorts of intrinsic properties of thinkers on which intentional contents depend that left the nature of the connection between phenomenal consciousness and intentionality open. Such

account could deploy a notion of aspectual shape, or modes of presentation, as the argument above, but leave the connection between aspectual shape and phenomenal properties open.

Let us recall the argument for the conclusion that the dependence of intentional content on phenomenal properties is necessary if content is determinate.

- (P1) There is a distinction between intrinsic intentionality and as-if intentionality; only intrinsic intentionality is genuinely mental.
- (P3) Intrinsic intentional states, whether conscious or unconscious, always have aspectual shape.
- (P4') The aspectual shape of a thinker's intentional states supervenes of his phenomenal properties of mental states.

The thesis of determinate content: All contentful intentional states have determinate content solely in virtue of their phenomenal properties or aspectual shape.

From these premises, we can draw the following conclusion.

- (C3) That intentional content of a thinker's intentional state is determinate depends on phenomenal properties of a thinker.
- (P4') clearly states that aspectual shape depends on phenomenal properties. We can revise the above argument in a way that is not committed to this dependence, though. For the purposes of content internalism, we still need aspectual shape to depend on intrinsic properties, a dependence we have previously discussed. We have talked of personal concepts, which I take to be modes of presentation, ways a subject can think of an entity. The question we need an answer to is this. It is possible for these modes of presentation to be narrow and not depend on phenomenal properties? If these modes of presentation are narrow and do not depend on phenomenal properties, then this approach to understanding narrow content would be less controversial than the approach that depends on phenomenal properties. It might be the case that modes of

presentation depend on only intrinsic properties of brains, an account that would clearly be congenial to physicalists.

Aspectual shape that does not depend on phenomenal properties

Frege was certainly not committed to modes of presentation being dependent on phenomenal properties. In fact, if we understand modes of presentations as Fregean senses, then we cannot think of them as narrow either, for the content of Fregean senses does not depend on only a thinker's intrinsic properties. Recall Frege's analogy of the telescope.

The reference of a proper name is the object itself which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lies the sense, which is indeed no longer subjective like the idea, but is yet not the object itself. The following analogy will perhaps clarify these relationships. Somebody observes the Moon through a telescope. I compare the Moon itself to the reference; it is the object of the observation, mediated by the real image projected by the object glass in the interior of the telescope, and by the retinal image of the observer. The former I compare to the sense, the latter is like the idea or experience. The optical image in the telescope is indeed one-sided and dependent upon the standpoint of observation; but it is still objective, inasmuch as it can be used by several observers.

If we can find a way to think of modes of presentation in both a distinctively non-Fregean way, as personal or subjective, and a way that does not have them depending on phenomenal properties, then we may have the key to an account of narrow content that is not dependent on rather questionable assumptions about the dependence of intentional content of phenomenal properties. On such an account, there are personal concepts, like the ideas Frege mentions in the quote above and there are intersubjective (or objective) concepts, like Fregean senses. On my way of understanding things, one's way of thinking of an entity can be called a mode of presentation. This mode of presentation of an object can match up with the

intersubjective concept of that entity is so far as the mode of presentation picks out the same entity as the intersubjective concept. In the case of Burge's patient, to use a familiar example, his mode of presentation of arthritis does not match the intersubjective concept of arthritis. Even if someone accepts this, one may have concerns about how we are to understand these personal concepts, or ideas as Frege calls them, as not being dependent on phenomenal properties. Frege's languages above seems to class ideas and experiences together in a way that suggests that ideas are private and dependent on phenomenal properties in the same way that experiences are. How are we to understand modes of presentation as both personal and not dependent on phenomenal properties?

It is hard to deny that there is something it is like to think of an entity under some aspect or other. When I think of Obama and his efforts to lift us out of a recession, there appears to be something it is like for me to have this thought, although it is hard to say very clearly what it is like other than to appeal to similar experiences had by others. Certainly what it is like to have a thought is not as phenomenally rich as what it is like to drink hot coffee, although there may be some phenomenology to occurrent thoughts. Granting that occurrent thoughts have phenomenology is still somewhat controversial. I think it is safe to reckon that there is some sense in which occurrent thoughts have a vague phenomenology. It is another step, though, to the conclusion that the aspectual shape of intentional states *depends* on the phenomenology of thoughts. This step involves the claim that the reason that aspectual shape is responsible for the determinateness of content is because of the phenomenal properties on which aspectual shape supervenes. There is some intuitive plausibility to this claim. We

should be careful, though, of distinguishing intentional content that has an accompanying phenomenology from intentional content that is what it is because of its phenomenology. We can grant that there is something it is like to conceive of an entity under a certain aspect and yet reject that what it is like to have that state plays a role in determining the content of that state. What reason has been given for the conclusion that aspectual shape depends on phenomenal properties?

Searle's work seems to provide an argument. As we saw, he holds that conscious and non-conscious states have aspectual shape. They have their aspectual shape in virtue of being potentially conscious. This is what the connection principle is supposed to show. This argument seems to offer a rationale for why aspectual shape depends on phenomenal properties. It can seem to provide a satisfactory story of the dependence. If these states were not potentially conscious, then they would not have aspectual shape. However, this is not the sort of dependence we are after. According to Searle, unconscious intentional states have aspectual shape when they are not conscious. They have the intentional content they have even when they are not supervening on phenomenal properties (even when, that is, they are not phenomenally conscious states). This is not to say that they are not supervening on intrinsic properties. It is just to say that they are not supervening on phenomenal properties when they are unconscious, even though they are fully contentful intentional states. The important point to see here is that the connection principle cannot be used to show that intentionality depends on phenomenology in a way sufficient to ground all intentional content in phenomenal properties. It can show that there is a sense in which intentional content depends on phenomenal consciousness. It may be that in virtue of being

phenomenally conscious that unconscious intentional states have their intentionality, while it is true that their content does not depend on the phenomenal properties of their owners. If unconscious intentional states have aspectual shape (and therefore determinate content) and this content does not supervene on phenomenal properties, then we must conclude that it supervenes on other intrinsic properties of thinkers if we are content internalists. Properties of thinker's brains seem to be likely candidates to play this role. Could it be that aspectual shape supervenes of the intrinsic properties of thinkers' brains? Such an account of content internalism does not make the controversial commitment to grounding intentional content in phenomenology. It allows that there may be some connection, such as the claim that occurrent intentional states have a phenomenology, but it does not require that intentional content depends on a thinker's phenomenal properties. Taking this approach gives us a way to separate the aspectual shape of intentional states from their phenomenology, and thus retain the mechanism we need for the determinateness of content while rejecting the questionable claim that all intentional content depends on a thinker's phenomenal properties.

Concluding Remarks

There is a great deal of work to be done on the part of content internalists to complete the story about the nature of the intrinsic properties on which intentional content depends. A large portion of that story will be concerned with how we acquire our concepts of the world in a way that allows them to represent the external world even though the concepts themselves depend on only intrinsic properties for their content. An important chapter in this account will surely concern the role perception has in generating concepts and linking them to objects in the world. I bring this up to note that this chapter has merely sketched some directions to be taken in the internalist account

of intentionality, but also to point out that there is a story for the content internalist to tell about how concepts and the intentional contents that they constitute can be about objects in the world even if they are not dependent on those objects for their content. The account would need to show how the content of representations of an external world does not require continued dependence on the external environment. I think that the role of perception in concept formation will be an important part of this story, although I cannot begin to address this at present.

This essay argued that content externalism should be rejected because it is not compatible with first-person access to intentional content. Rational agents must have such access to their intentional contents if they are to be properly engaged with their intentional content. Not only did we see that we have good reason to resist the conclusions of the arguments for content externalism due to its implications for rationality, we also saw that there are ways to resist these arguments. We saw that life with content internalism is not as bad as some have thought. On the whole, the project vindicates certain key aspects of the Cartesian Theory of Mind, namely first-person access to intentional content and content internalism. We need not accept the entire Cartesian Theory of the Mind, the view that Ryle so famously critiques and which has been under fire from subsequent generations of philosophers. It seems to me that in correcting for the excesses in Descartes view, some philosophers have been too quick to distance themselves from all the features of the Cartesian Theory of Mind. Certain aspects of this theory of mind must be true if we are rational agents.

Content internalism is not a philosophical thesis of little consequence. Its truth has tremendous consequences for how we think of the relationship between the mind

and the world. If it is true, we see that there is an important sense in which the mind is independent of the world. Even if it is true that there needs to be a world (as Davidson argues) and a second person for thought, the particular nature of individual thoughts is independent of a thinker's environment in a significant way.

As we have seen, the truth of content internalism does not mean that intentional content is not communicable; first-person access to intentional content and content internalism do not entail that we cannot communicate with others. First-person access to intentional content and incommunicability are not a package deal. They may have seemed to be aspects of the Cartesian Theory of Mind that go hand in hand. Ryle offers an example of this way of thinking in the following description he offers of the Cartesian Theory of Mind.

Though minds are inaccessible to one another, they may be said to resonate, like tuning-forks, in harmony with one another, though unfortunately they would never know it. I cannot literally share your experiences, but some of our experiences may somehow chime together, though we cannot be aware of their doing so, in a manner which almost amounts to genuine communication. In the most fortunate cases we may resemble two incurably deaf men singing in tune and in time with one another. But we need not dwell on such embellishments to a theory which is radically false.³⁰

The lesson here should be that in adopting certain aspects of the Cartesian Theory of Mind we need not adopt the entire theory. The incommunicability of thoughts based on lack of knowledge of other minds only follows if we adopt questionable epistemic premises. We would need to accept the premise, for example, that to know the minds of others, we must have direct access to them. There is no reason to adopt such a premise. We know the thoughts of others indirectly. We of course are not positioned to

³⁰ (Ryle 1984), page 57.

access the minds of others as we access our own, yet it surely does not follow from that that we cannot know the minds of others. Further, we do not need a special language to describe our thoughts. We make our thoughts known to others by employing a public language. Public languages are quite capable of describing intentional contents; in fact they appear to be the only systems capable of such work. We can accept all this, and still hold that intentional contents are what they are independent of environmental factors.

LIST OF REFERENCES

- Ayer, Alfred. 1963. *The Concept of a Person and Other Essays*. New York: Macmillan.
- Bach, Kent. 1987. *Thought and Reference*. Oxford: Oxford University Press.
- . 1988. "Burge's New Thought Experiment: Back to the Drawing Room." *The Journal of Philosophy* 85: 88-97.
- . 1997. "Do Belief Reports Report Beliefs?" *Pacific Philosophical Quarterly* 78: 215-241.
- Bayne, Tim and Montague, Michelle, eds. Forthcoming. *Cognitive Phenomenology*. Oxford: Oxford University Press.
- Bilgrami, Akeel. 1998. "Resentment and Self-Knowledge." In *Knowing our Own Minds*.
- . 2006. *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
- Biro, John. 1984. "What's in a Belief?" *Logique et Analyse* 27: 267-282.
- . "Individualism and Interpretation." *Acta Analytica* 6: 99-112.
- . 1992. "In Defense of Social Content," *Philosophical Studies* 67, pp. 277-293.
- . 1996. "Dretske on Phenomenal Externalism." *Philosophical Issues* 7:171-177.
- Block, Ned. 1990. "Inverted Earth" In *The Nature of Consciousness: Philosophical Debates*, Block, Flanagan, Guzeldere (eds).
- Block, Ned, Flanagan, Owen and Guzeldere, Guven, eds. 1997. *The Nature of Consciousness: Philosophical Debates*, Cambridge, MA: MIT Press.
- Boghossian, Paul. 1989. "Content and Self-Knowledge." In *Content and Justification: Philosophical Papers*.
- . 1992. "Externalism and Inference." *Philosophical Issues* 2, *Rationality in Epistemology*: 11-28.
- . 1994. "The Transparency of Mental Content." In *Content and Justification: Philosophical Papers*.
- . 1998. "What the Externalist Can Know A Priori." In *Content and Justification: Philosophical Papers*.
- . 2008. *Content and Justification: Philosophical Papers*. New York: Oxford University Press.

- Boyle, Matthew. 2011. "Transparent Self-Knowledge," *Aristotelian Society Supplementary Volume 85*: 223-241.
- Bridges, Jason. 2006. "Davidson's Transcendental Externalism." *Philosophy and Phenomenological Research* 73: 290-315.
- Brown, Jessica. 2004. *Anti-Individualism and Knowledge*. Cambridge, MA: MIT Press.
- Burge, Tyler. 1977. "Belief *de re*," *Journal of Philosophy* 74: 338-362.
- . 1979. "Individualism and the Mental." In *Externalism and Self-Knowledge*.
- . 1982. "Other Bodies." In *The Twin Earth Chronicles*, Pessin and Goldberg (eds).
- . 1986a. "Individualism and Psychology." *The Philosophical Review* 95: 3-45.
- . 1986b. "Intellectual Norms and the Foundations of Mind." *Journal of Philosophy* 83: 697-720.
- . 1988. "Individualism and Self-Knowledge." *Journal of Philosophy* 85: 649-663.
- . 1993. "Content Preservation." *Philosophical Review* 102: 457-488.
- . 1996. "Our Entitlement to Self-Knowledge." In *Externalism and Self-Knowledge*, Ludlow and Martin (eds).
- Byrne, Alex. 2011. "Transparency, Intention, and Belief." *Aristotelian Society Supplementary Volume 85*: 201-221.
- Crane, Tim. 1991. "All the difference in the world." In *The Twin Earth Chronicles*.
- . 2001. *Elements of Mind*. Oxford: Oxford University Press.
- Davidson, Donald. 1987. "Knowing One's Own Mind." In *Subjective, Intersubjective, Objective*.
- . 2001. *Subjective, Intersubjective, Objective*. New York: Oxford University Press.
- Dretske, Fred. 1996. "Phenomenal Externalism, or if Meanings Ain't in the Head, Where are Qualia." *Philosophical Issues* 7:143-158.
- Edgley, Roy. 1969. *Reason in Theory and Practice*. London: Hutchinson.
- Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.
- Falvey, Kevin and Owens, Joseph. 1994. "Externalism, Self-Knowledge, and Skepticism." *Philosophical Review* 103: 107-137.

- Fodor, Jerry. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
- Farkas, Katalin. 2003. "What is Externalism?" *Philosophical Studies* 112: 187-2003.
- . 2008a. "Phenomenal Intentionality Without Compromise." *The Monist* 91: 273-293.
- . 2008b. *The Subject's Point of View*. New York: Oxford University Press.
- Frege, Gottlieb. 1892. "Sense and Reference." In *Readings in the Philosophy of Language*, Ludlow (ed).
- . "The Thought: A Logical Inquiry." In *Readings in the Philosophy of Language*, Ludlow (ed).
- Gallois, Andre. 1996. *The World Without, The Mind Within*. Cambridge: Cambridge University Press.
- Gertler, Brie, ed. 2003. "Introduction: Philosophical Issues about Self-Knowledge." In *Privileged Access: Philosophical Accounts of Self-Knowledge*, Gertler (ed).
- . 2011. *Self-Knowledge*. New York: Routledge.
- . Forthcoming. "Mental Content: Externalism and Internalism."
- Gibbons, John. 1996. "Externalism and Knowledge of Content." *Philosophical Review* 105: 287-310.
- Goldberg, Sanford, ed.. 2007. *Internalism and Externalism in Semantics and Epistemology*. New York: Oxford University Press.
- Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." In *The Nature of Consciousness: Philosophical Debates*, Block, Flanagan, Guzeldere (eds).
- Heil, John. 1988. "Privileged Access." In *Externalism and Self-Knowledge*, Ludlow and Martin (eds).
- Horgan, Terry and Kriegel, Uriah, eds. Forthcoming. *Phenomenal Intentionality: New Essays*, New York: Oxford University Press.
- Horgan, Terry and Kriegel, Uriah. Forthcoming. "The Phenomenal Intentionality Research Program." In *Phenomenal Intentionality: New Essays*, Horgan and Kriegel (eds).
- Horgan, Terry, Tienson, John, and Graham, George. 2004. "Phenomenal Intentionality and the Brain in a Vat." In *The Externalist Challenge*, Schantz (ed).
- Jackson, Frank. 2003a. "Narrow Content and Representation, or Twin Earth Revisited." *Proceedings and Addresses of the American Philosophical Association* 77: 55-70

- . 2003b. "Representation and Narrow Belief," *Philosophical Issues* 13: 99-112.
- . 2004. "On an Argument from Properties of Words to Broad Content." In *The Externalist Challenge*, Schantz (ed.),
- Jeshion, Robin, ed.. 2010. *New Essays on Singular Thought*. New York: Oxford University Press.
- Kaplan, David. 1989. "Demonstratives." In *Themes from Kaplan*, Almong, Perry and Wettstein (eds). Oxford University Press.
- Kim, Jaegwon. 1996. "Dretske's Qualia Externalism." *Philosophical Issues* 7:159-165.
- Kripke, Saul. 1979. "A Puzzle About Belief." In *Readings in Philosophy of Language*, Ludlow (ed).
- . 1972/1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lau, Joe and Deutsch, Max. 2008. "Externalism About Mental Content." In *The Stanford Encyclopedia of Philosophy*, ed. Edward N Zalta.
- Lewis, David. 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88: 513-43.
- Loar, Brian. 1988. "Social Content and Psychological Content." In *The Twin Earth Chronicles*, Pessin and Goldberg (eds).
- . 2003. "Phenomenal Intentionality as the Basis for Mental Content." In *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Hahn and Ramberg (eds). MIT Press.
- Ludlow, Peter. 1995. "Externalism, Self-Knowledge, and the Prevalence of Slow Switching." In *Self-Knowledge and Externalism*, Ludlow and Martin (eds).
- . *Readings in the Philosophy of Language*. Cambridge, MA: MIT Press.
- Ludlow, Peter and Martin, Norah, eds. 1998. *Externalism and Self-Knowledge*. Palo Alto: CSLI Publications.
- Ludwig, Kirk. 1994. "First Person Knowledge and Authority." In *Language, Mind, & Epistemology: On Donald Davidson's Philosophy*, Preyer, Siebelt, and Ulfing (eds). Kluwer Academic Publishers.
- . 1996a. "Explaining Why Things Look the Way They Do." In *Perception, Vancouver Studies in Cognitive Science*, v. 5, Aikens (ed). Oxford University Press.

- . 1996b. "Singular Thought and the Cartesian Theory of Mind." *Noûs* 30: 434-460.
- Luntley, Michael. 1999. *Contemporary Philosophy of Thought: Truth, World, Content*. Malden, MA: Blackwell.
- Lycan, William. 2001. "The Case for Phenomenal Externalism." *Philosophical Perspectives* 15: 17-35
- McDowell, John. 1984. "De re Senses" *Philosophical Quarterly* XXXIV: 283-94.
- . "Singular Thought and the Extent of Inner Space." In *Subject, Thought, and Context*, McDowell and Pettit (eds).
- . 1994. *Mind and World*. Cambridge, MA: Harvard University Press.
- McDowell, John. and Pettit, Philip, eds. 1986. *Subject, Thought, and Context*, Oxford: Oxford University Press.
- McGinn, Colin. 1977. "Charity, interpretation, and belief," *Journal of Philosophy* 74:521-535.
- . "Consciousness and Content." In *The Nature of Consciousness: Philosophical Debates*, Block, Flanagan, and Guzeldere (eds).
- McKinsey, Michael. 1991. "Anti-Individualism and Privileged Access." In *Externalism and Self-Knowledge*, Ludlow and Martin (eds).
- . 2003. "Transmission of Warrant and Closure of Apriority." In *New Essays on Semantic Externalism and Self-Knowledge*, Nuccetelli (ed).
- McLaughlin, Brian. and Tye, Michael. 1998. "Externalism, Twin Earth, and Self-Knowledge." In *Knowing Our Own Minds*, Wright, Smith and Macdonald (eds).
- . 1998. "Is Content-Externalism Compatible with Privileged Access?" *The Philosophical Review* 107: 349-80.
- Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.
- Nozick, Robert. 1981. *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Nuccetelli, Susana. 1999. "What Anti-Individualists Cannot Know A Priori." *Analysis* 59: 48-51.
- Nuccetelli, Susana, ed. 2003. *New Essays on Semantic Externalism and Self-Knowledge*. Cambridge, MA: MIT Press, 2003.

- Peacocke, Christopher. 1983. *Sense and Content*. Oxford, Oxford University Press.
- Perry, John. 1979. "The Essential Indexical." *Nous* 13: 3-21.
- Pessin, Andrew. and Goldberg, Sanford, eds. 1996. *The Twin Earth Chronicles*. New York: M.E. Sharpe.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning' ." In *The Twin Earth Chronicles*, Pessin and Goldberg (eds).
- . 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.
- Reed, Baron. 2010. "Self-Knowledge and Rationality." *Philosophy and Phenomenological Research* 81: 165-181.
- Rowlands, Mark. 2003. *Externalism: Putting Mind and World Back*. Ithaca: McGill-Queen's University Press.
- Ryle, Gilbert. 1949/1984. *The Concept of Mind*, Chicago: University of Chicago Press.
- Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge, MA: MIT Press.
- Sawyer, Susan. 1999. "An Externalist Account of Introspective Self-Knowledge." *Pacific Philosophical Quarterly* 80: 358-78
- Schantz, Richard, ed. 2004. *The Externalist Challenge*. Berlin: Walter de Gruyter GmbH & Co.
- Searle, John. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- . *Rediscovery of the Mind*, Cambridge. MA: MIT Press.
- Segal, Gabriel. 1989. "The Return of the Individual." *Mind* XCVIII: 39-57.
- . 2000. *A Slim Book about Narrow Content*. Cambridge, MA: MIT Press.
- Shoemaker, Sydney. 1996. *The First Person Perspective and Other Essays*. New York: Cambridge University Press.
- . 2003. "Moran on Self-Knowledge." *European Journal of Philosophy* 11: 391-401.
- . 2009. "Self-Intimation and Second Order Belief." *Erkenntnis* 71: 35-51.
- Siewert, Charles. 1998. *The Significance of Consciousness*. Princeton, NJ: Princeton University Press.

- . 2003. "Self-Knowledge and Rationality: Shoemaker on Self-Blindness." In *Privileged Access*, Gertler (ed).
- . 2006. "Consciousness and Intentionality." *The Stanford Encyclopedia of Philosophy*, ed. Edward N Zalta.
- . Forthcoming. "Phenomenal Thought." In *Cognitive Phenomenology*, Bayne and Montague (eds).
- Soames, Scott. 1987. "Direct Reference, Propositional Attitudes, and Semantic Content." *Philosophical Topics* 15: 47-87.
- Stalnaker, Robert. 1999. *Context and Content*. Oxford: Oxford University Press.
- Strawson, Peter. 1959. *Individuals*. London: Methuen.
- Warfield, Ted. 1992. "Privileged Self-Knowledge and Externalism are Compatible." In *Externalism and Self-Knowledge*, Ludlow and Martin, (eds).
- . 1997. "Externalism, Self-Knowledge and the Irrelevance of Slow-Switching." In *Externalism and Self-Knowledge*, Ludlow and Martin (eds).
- Woodfield, Andrew, ed. 1982. *Thought and Object: Essays on Intentionality*. Oxford: Oxford University Press.
- Wright, Crispin, Smith, Barry, and Macdonald, Cynthia, eds. 1998. *Knowing Our Own Minds*. Oxford: Oxford University Press.
- Zemach, Eddy. 1976. "Putnam's Theory on the Reference of Substance Terms." In *The Twin Earth Chronicles*, Pessin and Goldberg (eds).

BIOGRAPHICAL SKETCH

Casey Woodling graduated from the University of Central Florida cum laude with a double major in Philosophy and English in 2002. He began his graduate studies at the University of Florida in 2003, earning his M.A. in 2006. He took a two-year leave of absence to serve in the Peace Corps in Madagascar (2007-2009), and earned his Ph.D. in the fall of 2011.