Reconciling Content-Externalism and Self-Knowledge

Two Frameworks Considered

Benjamin Sorgiovanni
Wolfson College, Oxford
DPhil Philosophy

Word count: 74,835

Reconciling Content-Externalism and Self-Knowledge: Two Frameworks Considered

Benjamin Sorgiovanni, Wolfson College, DPhil Philosophy, Trinity Term 2015

Abstract

In this thesis, I assess the prospects for reconciling content-externalism and crucial guiding intuitions about self-knowledge within two different frameworks, respectively. The first framework belongs to the prominent contemporary externalist, Tyler Burge. The second framework is built from central strands of thought in Ludwig Wittgenstein's later work.

I argue that a tension between the basic externalist intuition and crucial guiding intuitions about self-knowledge arises within a Burgean framework which does not arise within a Wittgensteinian framework. I show that given Burge's views about the individuation of mental content, switching a subject slowly between two relevantly dissimilar contexts can undermine knowledgeability of her epistemic reasons. I argue that this is a troubling result, given Burge's views about the sorts of things that epistemic reasons are. On Burge's view, epistemic reasons are rational relations between mental states. If slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of the rational relations between mental states. But the thought that the knowledgeability of the rational relations between mental states might be sensitive to changes in one's context in this way seems at odds with our intuitive picture of self-knowledge.

I argue that this tension does not arise within a Wittgensteinian framework because there is evidence that Wittgenstein rejects certain of the claims about the individuation of mental content which generate the tension in Burge's case.

The thesis examines the substantive similarities and differences between the Burgean and Wittgensteinian frameworks more generally. In doing so, it maps two contrasting ways in which the basic content-externalist intuition might be elaborated.

Acknowledgements

I would like to express my immense gratitude to my two supervisors, Bill Child and Lizzie Fricker. Lizzie provided invaluable feedback and guidance during the first year of the DPhil. In addition to being my principal supervisor during the second, third and fourth years of the DPhil, Bill supervised me for the Wittgenstein option and the thesis component for the BPhil. This thesis has its origins in work I completed during that time. I am incredibly grateful to Bill for his guidance, patience and support over the last six years.

I thank Anandi Hattiangadi and Martin Davies for their feedback on draft material which I submitted for my confirmation of status and Anil Gomes for his comments on a draft of Chapter 3.

Finally, my thanks to Clare, Mum, Dad and Chris for their patience, love and support. I do not have the words to express how much I owe them.

Table of Contents

Claims and Principles

(in order of introduction)

Misunderstanding: A subject can have thoughts involving concepts which she misunderstands.

Lay Knowledge: A subject can have thoughts involving the concept *C* even though she is ignorant of certain of the normative characterisations associated with the word '*C*'.

Deference: If the concept *C* is standardly associated with the term '*C*' by the experts, then a subject who either misunderstands *C* or lacks expert knowledge of *C*'s can, by deferring to those experts, use '*C*' to express *C*.

False Belief: If a proposition, *p*, is a meaning-giving characterisation associated with the word '*C*', then an otherwise competent speaker who believes that *p* is false can still have thoughts involving the concept *C*.

Direct Causal Contact: Direct causal relations at work in perception and in perceptually-backed demonstrative applications of an empirically applicable term, '*C*', which connect a subject with actual *C*'s, can bring it about that she uses '*C*' to express the concept *C*.

Discrimination: *S* is warranted in judging, on the basis of reflection alone, that she is thinking that *p* only if she is able to discriminate on the basis of reflection alone between instances in which she is thinking that *p* and instances in which she is thinking that *q*, where *S*'s thinking that *q* is a relevant alternative to her thinking that *p*.

Knowledgeable Reviewability: In order for one to be a critical reasoner, one's mental states, reasons and reasoning must normally be knowledgeably reviewable.

Entitlement: In order for one to be a critical reasoner, one must have an entitlement to judgments about one's mental states, reasons and reasoning.

Correctness: In order for one to be a critical reasoner, one's judgments about one's mental states, reasons and reasoning must normally be correct.

Reflection: Reflection on our mental states, reasons and reasoning adds a rational element to reasoning by giving one rational control over one's reasoning.

Having Deductive Reason: *S* has a reason to believe that *p* by way of a deduction if and only if there is some set of warranted propositional attitudes which is present in *S*'s psychology from which *p* can be deductively inferred by *S*.

Sufficient Difference: Any case in which the difference between the ways in which *A* and *B* are disposed to apply the word '*C*' is such that, first, *A* is disposed to apply '*C*' correctly but *B* is disposed to apply it incorrectly (or vice versa), and second, *B*'s disposition is ultimately to be explained in terms of her being in error about the rules governing the use of '*C*', is a case in which *A* and *B* understand the word '*C*' to mean different things, respectively.

Introduction

I

In virtue of what do our intentional mental states—our beliefs, desires, intentions, and so on—have the content they do? Traditionally, when considering questions of this sort philosophers have appealed to what we might call *internalist* views of mental content. The basic internalist thought is that the content of a subject's mental states is either determined exclusively by the subject's non-intentionally described intrinsic properties or is itself a metaphysically intrinsic property. However, over the last forty years or so a growing number of philosophers have been attracted by *externalist* views of mental content. The basic externalist thought is that the content of a subject's mental states constitutively depends at least in part on the relations which the subject bears to a context.

One common objection to externalism about mental content is that it is at odds with guiding intuitions about self-knowledge. For example, it would seem to be a guiding intuition about self-knowledge that we normally know groundlessly what we are thinking. Judgments about the mental states of another person are, on many accounts, inferences which we make on the basis of observational evidence, namely, the things which that person says or does. Our warrant for such judgments depends on their being based on such evidence. In comparison, our warrants for judgments about our own mental states do not, generally speaking, depend on evidence in this way. In order for my self-ascriptions to be warranted I generally do not need to observe my own behaviour, the things I say and do. Nor, it seems, do I need to consult any sort of *inner* evidence.

Some philosophers have argued that the intuition that we normally know groundlessly what we are thinking is inconsistent with the basic externalist thought. If the basic externalist thought is true, they ask, then how *could* we normally know groundlessly what we are thinking? If the content of our mental states constitutively depends in part on our relations to a context, then surely we need to investigate that context before we can know what we are thinking. But if

knowledge of what we are thinking must wait upon empirical investigation in this way, then such knowledge *cannot* normally be groundless. Externalists have responded in turn by, for example, challenging presumptions about self-knowledge upon which they claim such objections are based.

This thesis is about attempts to reconcile externalist views about the individuation of mental content with guiding intuitions about self-knowledge. More specifically, its principal aim is to assess the prospects for reconciling content-externalism with self-knowledge within two frameworks, respectively. The first framework belongs to the prominent contemporary externalist, Tyler Burge. The second framework is built from central strands of thought in Ludwig Wittgenstein's later work.

My thesis makes a novel contribution to the existing literature on the compatibility of content-externalism and guiding intuitions about self-knowledge in at least the following two respects. First, it situates Wittgenstein's later work with respect to the debate. There are discussions of the relation between strands in Wittgenstein's later thought and contemporary forms of content-externalism.[1] But few bring that thought to bear specifically on questions concerning the compatibility of content-externalism and guiding intuitions about self-knowledge. Fewer still compare Wittgensteinian and Burgean frameworks in connection with this concern. To my knowledge, there exists no systematic comparative analysis of these frameworks, let alone an analysis of this sort within the context of the debate between compatibilists and incompatibilists about content-externalism and self-knowledge. This is unfortunate, given the prima facie similarities between strands in Wittgenstein's later thought and Burge's externalism.[2]

One might worry that any attempt to situate Wittgenstein's views with respect to contemporary debates about content-externalism is bound to be at odds with well-documented deflationary strands in Wittgenstein's work. On Wittgenstein's view, philosophy is a primarily descriptive endeavour; philosophers are not, or at least ought not to be, in the business of putting

[1] See, for example: William Child (2006); Hanjo Glock & John Preston (1995); and Phillip Pettit (1983).
[2] Burge himself has noted these similarities. See, for example, Burge (2007b: 6, 2013a: 526).

forward theses or proffering explanations. Their task, rather, is to render perspicuous the *grammar* of our language, those rules which structure our everyday engagement with language. Wittgenstein would almost certainly regard contemporary externalists (and proponents of most other philosophical '-isms') as engaged in just the sort of philosophical theorising which his conception of philosophy is meant to guard against. Isn't it problematic, then, to seek to bring Wittgenstein into conversation with contemporary externalists?

I do not think so. I acknowledge that there are prominent deflationary or anti-explanatory strands in Wittgenstein's later work. But when thinking about the relevance of particular components of that work for contemporary philosophical debates, I think it is a perfectly legitimate approach to put these strands to one side.[3] Asking about the relevance of components of Wittgenstein's thought for contemporary debates is not, after all, the same as asking about the application which Wittgenstein himself would have made of those components within the context of those debates. I am not defending the view that Wittgenstein is *actually* an externalist. My concern is not purely, or even primarily, exegetical. It is instead to answer the following question: To what extent is an externalism which cleaves closely—more closely than does Burge's own externalism—to Wittgenstein's views about the individuation of mental states vulnerable to objections from self-knowledge?

One further, related point: I refer occasionally to 'Wittgenstein's externalism'. I use this description purely for the sake of convenience. The kind of externalism which falls under this description builds on prominent strands in Wittgenstein's later thought. I am not suggesting that it is the *only* kind of externalism which could conceivably claim origins in Wittgenstein's later work.

The second respect in which my thesis contributes to the existing literature is by developing a novel problem for the Burgean framework. One common way of arguing for the

[3] There is a well-known precedent for this sort of approach. In the introduction to his *Wittgenstein on Rules and Private Language*, Saul Kripke acknowledges that he very likely presents Wittgenstein's ideas in a way which Wittgenstein himself would not approve of. But he does not take this consideration to render his project problematic. This is because, 'the present paper should be thought of as expounding neither 'Wittgenstein's' argument nor 'Kripke's': rather Wittgenstein's argument as it struck Kripke, as it presented a problem for him' (1982: 5).

claim that there is a tension between externalism and self-knowledge is by way of *slow-switching objections*. These objections purport to show that if externalism is true, then switching a subject slowly between two relatively dissimilar contexts in such a way that the subject cannot tell that she is being slowly switched can have consequences which are strongly counterintuitive. Typically, these consequences have to do with the *knowledgeability* of a subject's mental states.[4] It is argued that if externalism is true, then slow-switching can bring it about that a subject cannot know groundlessly certain of her occurrent, conscious attitudes. Since it is considered a datum that changes in one's context cannot undermine the knowledgeability of one's mental states in this way, this consequence is counted as evidence against externalism about mental content.[5]

In Chapter 3 of this thesis, I develop a different sort of slow-switching objection to Burge's views about the individuation of mental content. I argue that given Burge's views, switching a subject slowly between two relevantly dissimilar contexts can undermine knowledgeability of her epistemic reasons. I argue that this is a troubling result, given Burge's views about the sorts of things that epistemic reasons are. As the discussion will show, epistemic reasons are rational relations between mental states on Burge's view. If slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of the rational relations between mental states. But the thought that the knowledgeability of the rational relations between mental states might be sensitive to changes in one's context in this way is at odds with our intuitive picture of self-knowledge. I go on to argue that this tension does not arise within a Wittgensteinian framework because Wittgenstein rejects certain claims about the individuation of mental content which give rise to the tension in Burge's case.

I said above that one common way of arguing for an inconsistency between externalism and guiding intuitions about self-knowledge makes appeal to slow-switching objections. There exist in the literature at least *two* familiar lines of argument for an inconsistency. The first line I

[4] To clarify, 'knowledgeability' in this context means *knowability*. The consequences in question have to do with a subject's ability to know her mental states. They do not have to do with the status of her mental states as knowledge. I borrow the term from Burge (1996: 97, fn. 2).
[5] The locus classicus for this line of argument is Paul Boghossian (1998).

sketched above. It begins with the assumption that externalism is true. There is then a move, which is often substantiated by appeal to slow-switching objections, from this assumption to the claim that we cannot normally know groundlessly what we are thinking.

The second line of argument begins by assuming *both* that externalism is true and that it is true that we can normally know our mental states a priori.[6] It is then claimed that deeply counterintuitive consequences follow from these assumptions. Proponents claim that if externalism is true, then a subject can know a priori that her thinking a particular thought—her thinking that *p*, for example—depends on things in her environment being thus and so. But if a subject can know a priori that she is thinking that *p* and if she can know a priori that her thinking that *p* depends on things in her environment being thus and so, then it seems she can know a priori that things in her environment are thus and so. But this is surely an absurd result. A subject *cannot* know a priori that things in her environment are thus and so. The conclusion of this second line of argument is that it cannot be the case both that externalism is true and that it is true that we can normally know our mental states a priori.[7]

Both lines of argument have received extensive attention in the literature. My focus, however, will be exclusively on the first line of argument. My reason for focusing exclusively on the first line of argument is that my broader interest is in the details and workability of Burge and Wittgenstein's respective responses to it.

II

[6] The intuition that we normally know our mental states a priori is distinct from the intuition that we normally know our mental states groundlessly. The notions of groundless knowledge and a priori knowledge are, I take it, neither intentionally nor extensionally equivalent. Not all groundless knowledge is a priori knowledge. Plausibly, perceptual knowledge is non-inferential and so groundless. But it is not a priori knowledge. My formulation of the guiding intuitions at work in the first and second lines of argument, respectively, is in keeping with the ways in which these arguments are discussed in the literature. These different formulations serve to bring out what is at stake in each of the lines of argument.

[7] The locus classicus for this second line of argument is Michael McKinsey (1991). My characterisation of this second line of argument is informed by Brown's characterisation (2009). Brown, following Davies, refers to the conclusions of these two lines of argument as the *achievement* and c*onsequence problems*, respectively.

This thesis consists of five chapters. Here, in summary, is an overview of each (more detailed summaries are included at the beginning of each chapter).

I begin Chapter 1 by introducing and elucidating what I call *externalism's driving intuition*, the thought that at least some types of mental states are at least partly individuated by the relations in which a speaker stands to a context. I then set out and discuss several thought experiments offered by Burge and Wittgenstein, respectively, which prima facie support the intuition, noting salient similarities and differences between these two sets of thought experiments. In the second half of Chapter 1, I introduce a second intuition, which I call the *driving intuition about self-knowledge*. This is the thought that we normally know groundlessly what we are thinking. I close by outlining three categories into which accounts of how self-knowledge could normally be groundless, might fall.

In Chapter 2, I discuss slow-switching objections in connection with Burge's thought experiments which I introduced in Chapter 1. I conclude that slow-switching objections have initial plausibility as objections to two of those thought experiments, which I call the *arthritis* and *water cases*, respectively. I then go on to consider Burge's response to slow-switching objections. This response culminates in the claim that slow-switching cannot undermine the knowledgeability of one's mental states. I close by unpacking Burge's views about the source of our warrant to first-person judgments.

In Chapter 3, I develop a novel slow-switching objection to Burge's views. I argue that given assumptions at work in Burge's arthritis case, slow-switching can undermine knowledgeability of a subject's epistemic reasons. This is a prima facie troubling result I claim, given that on Burge's view epistemic reasons are rational relations between mental states. I go on to consider whether the objection can be applied to the two further thought experiments belonging to Burge which I considered in Chapter 1.

In Chapter 4, I situate Wittgenstein's later work with respect to the debate between compatibilists and incompatibilists about externalism and self-knowledge. I begin by showing that

Wittgenstein would accept the driving intuition about self-knowledge. I go on to consider whether slow-switching objections have initial plausibility as objections to the two thoughts experiments belonging to Wittgenstein which I discussed in Chapter 1, given Wittgenstein's views about the way in which mental content is individuated. I conclude that they do not. I then consider whether slow-switching objections have initial plausibility as objections to the arthritis case, given Wittgenstein's views. Again, my conclusion is that they do not.

I begin Chapter 5 by considering how Wittgenstein might respond should it turn out that there is some further slow-switching scenario with respect to which slow-switching objections *do* have initial plausibility, given a Wittgensteinian framework. I defend the view that Wittgenstein's response is likely to be similar in form to Burge's response, as outlined in Chapter 2. I go on to consider whether Wittgenstein offers a positive account of one's warrant for judgments about one's mental states and if so how best to interpret that account. I defend John McDowell's interpretation of Wittgenstein on self-knowledge, according to which Wittgenstein declines to give a positive account. I close by defending the view that the slow-switching objection which I raised against the Burgean framework in Chapter 3 does not have purchase against the Wittgensteinian framework. This is because there are strands of thought in Wittgenstein's later work that are at odds with those assumptions which give rise to the objection in Burge's case.

Chapter 1

EXTERNALISM AND SELF-KNOWLEDGE

INTRODUCTION

My aims in this chapter are essentially twofold: first, to introduce externalism's driving intuition and present several thought experiments offered by Burge and Wittgenstein, respectively, which prima facie support the intuition; and second, to introduce what I will call the driving intuition about self-knowledge. In Subsection 1.1 I introduce and elucidate externalism's driving intuition. In Subsection 1.2 I outline three arguments which Burge offers in support of this intuition. The discussion will make it clear that these thought experiments complement one another, each one extending the scope of Burge's externalism to a particular group of subjects. In Subsection 1.3 I introduce two arguments taken from Wittgenstein's work which prima facie also support externalism's driving intuition. The discussion will aim at highlighting some of the points of similarity and difference between Wittgenstein's thought experiments and Burge's thought experiments which we considered in Subsection 1.2.

In Subsection 2.1 I switch focus from externalism's driving intuition to the driving intuition about self-knowledge, the thought that we normally know groundlessly what we are thinking. Generally speaking, a judgment's being groundless is a consideration which counts against its constituting knowledge. So we might well ask why first-person judgments should be any different. What is it about first-person judgments which explain how they *could* normally constitute knowledge? In Subsection 2.2 I outline three different categories into which accounts of self-knowledge that attempt to provide an answer to this question might fall.

1. EXTERNALISM'S DRIVING INTUITION

1.1 Introducing the Intuition

The debate between internalists and externalists in the philosophy of mind is a debate about whether a subject's relations to a context play a role in individuating her mental states. Internalism, as I shall understand it, is the view that a subject's mental states either are individuated in such a way that if the subject's non-intentionally described intrinsic properties are the same, then their mental states are the same, or are themselves metaphysically primitive intrinsic properties of a subject.[1] A proponent of a view of this kind denies that a subject's relations to a context play a role in individuating her mental states, either because she denies that mental states are individuated by more basic properties or because she insists that only more basic *intrinsic* properties of a subject can play an individuating role.

Externalists take the opposing view. They deny, on the one hand, that mental states are metaphysically primitive intrinsic properties of a speaker and on the other, that the only things that play a role in individuating mental states are a subject's more basic intrinsic properties. The basic externalist thought is that at least some types of mental states are at least partly individuated by the relations in which a speaker stands to a context. I shall refer to this basic thought hereafter as *externalism's driving intuition*.

This statement of externalism's driving intuition stands in need of clarification. Specifically, it will be useful to clarify: the kinds of mental states at issue in the debate; the kinds of relations which play an individuating role according to externalists; and what it means to say that relations to a context individuate a mental state. I shall discuss each of these points in turn.

[1] One can endorse internalism without thereby committing oneself to any particular account of what it is for a property to be intrinsic. For two such accounts see David Lewis (1983) and Rae Langton & David Lewis (1998).

The debate between internalists and externalists typically centres on those mental states which are paradigmatically intentional—states such as believing, desiring, intending, and so on. Some philosophers have argued that externalism's driving intuition is true of other sorts of intentional states, for example, some perceptual states.[2] However, I will understand externalism's driving intuition as in the first instance an intuition about the individuation of the propositional attitudes. Hereafter, unless stated otherwise, I will use 'mental state' to refer to intentional mental states of this specific kind.

As I have characterised it, externalism's driving intuition is an intuition about the individuation of mental states by type as opposed to token. If it were about the individuation of states by token and if the intuition were sound, then the relations in which a subject stands to a context would bear on the question of whether he was so much as tokening a mental state.[3] The intuition takes it for granted that a subject is tokening a state but maintains that the relations in which the subject stands to a context sometimes bear on the question of whether the state is of the type, say, *belief that p* as opposed to, say, *belief that q*. To be clear, it will suffice to vindicate the intuition if it should turn out that on at least *some* occasions, the relations that a subject bears to a context play an individuating role with respect to her believing that *p*. It is not part of the intuition that relations to a context play an individuating role in *every* case in which a subject believes that *p*.

Mental states are individuated by type partly in virtue of their content—a proposition or thought—and partly in virtue of the attitude which the subject takes to that content. A belief that *p* is the type of state that it is partly in virtue of the fact that it is a *belief* (as opposed to, say, a desire or an intention) and partly in virtue of the fact that it is a belief *that p* (as opposed to, say, a belief that *q* or a belief that *y*). Generally speaking, internalists and externalists disagree about whether a

[2] See, for instance, Tyler Burge (2007a and 1986a).
[3] In fact, I think that both Burge and Wittgenstein endorse (or would endorse) the claim that a subject's relations to her context bear on whether she is so much as tokening a mental state. But this is not a component of their respective views which I shall discuss in this thesis. Other philosophers who endorse this claim include Donald Davidson. See, for example, the discussion of swampman in Davidson (2001c).

subject's relations to a context play a role in individuating her mental states because they disagree about whether those relations play a role in individuating the content of those states. Questions concerning the individuation of the attitudinal component of mental states are standardly left to one side. Nevertheless, defending the claim that the attitudinal components of some types of mental states are individuated partly in virtue of the relations in which a speaker stands to a context is one perfectly good way of defending externalism's driving intuition.[4] If the claim is true then the intuition is sound. My focus, however, will be on thought experiments which look to defend externalism's driving intuition by defending a claim about the individuation of the content of certain mental states.

I characterised the content of a mental state as a proposition or thought. Externalism's driving intuition is neutral, however, with respect to the sorts of things propositions or thoughts are. One can endorse that intuition without thereby committing oneself to a theory of propositions as, for example, Fregean senses, structured entities with objects or properties as constituents or sets of possible worlds.

The content of mental states is *conceptual* content, content that involves concepts that a speaker must understand sufficiently well to exercise in thought and reasoning in order to have mental states involving that content. For example, in order for someone to believe that Madrid is the capital of Spain they must understand certain concepts, namely, the concept Madrid, the concept capital and the concept Spain. As in the case of propositions, externalism's driving intuition is non-committal with respect to the sorts of things concepts are.[5]

[4] Sven Bernecker has argued that Burge's views about the individuation of content support the claim that the attitudinal component of a subject's mental state may depend in part on the relations which she bears to her social context. See Bernecker (1996). Bernecker argues that it follows that Burge's views about the individuation of mental content are incompatible with our intuitions about self-knowledge.

[5] Someone might object that externalism's driving intuition is at least inconsistent with the mental representation theory of concepts. It depends how one understands such a theory. As Burge himself has pointed out, externalism is not inconsistent with either the idea that occurrent propositional attitudes are associated with token mental representations or the idea that such tokens enter into causal relations in virtue of their syntax and not their intentional content (Burge, 1982: 185). If, however, one takes the mental representation theory of concepts to involve a commitment to the claim that the relations between concepts thus conceived give an exhaustive explanation of how propositional attitudes come to have the intentional content they do, then clearly there *is* a conflict between that theory and externalism's driving intuition.

We have gone some way towards clarifying the notion of a mental state at work in my earlier characterisation of externalism's driving intuition. What are the kinds of relations which play a role in individuating mental states thus conceived, according to externalists? Those relations which externalists commonly identify as playing such a role fall into two broad types. Relations of the first type hold between a subject and a socio-linguistic context, that is, a community of speakers. Relations of the second type hold between a subject and a physical context, which consists of kinds and properties. The externalist need not identify only one type of relations as playing an individuating role; she can acknowledge that individuation depends on the complex interaction of relations of both types.

To be clear, the relevant relations need not be to an *occurrent* context. Obviously, it would suffice to vindicate externalism's driving intuition if it should turn out that *S*'s thinking that *p* at some time, *t*, depends on her relations to a context which obtains at *t*. But that intuition would also be vindicated if it should turn out that *S*'s thinking that *p* at *t* depends on relations that she bears to a context which obtained, for example, at some time prior to *t*. In other words, when thinking about externalism's driving intuition, we should understand 'context' to mean *spatio-temporal* context. We will see the significance of this point when considering Wittgenstein's views about the individuation of mental content in Chapter 1, Subsection 1.3.

Let us now consider the notion of *individuation*. When externalists claim that a subject's intentional mental states are partly individuated by the relations they bear to a context they are advancing a metaphysical thesis. They are claiming that the mental states in question depend constitutively on those relations. This claim may seem to commit the externalist to the seemingly bizarre view that some mental states are partly located outside the physical boundaries of a subject's body. One way in which the externalist can deny that she is so committed is by insisting on a distinction between *x*'s constitutively depending on *y* and *x*'s being constituted (either in part or in whole) by *y*. This is the tack which Burge himself takes. He writes:

> … [externalism] certainly does not entail that thoughts are 'outside' the head or are themselves relations to something external. Neither thoughts themselves nor their representational contents are

> relations to something outside the individual. Their natures constitutively *depend* on relations that are not reducible to matters that concern the individual alone. But the natures are not themselves relations, and their representational contents are not themselves (in general) relational. (Burge, 2007e: 154)

Someone might object to this tack on the grounds that it muddies the distinction between constitutive dependence and weaker kinds of dependence, causal dependence for example. One obvious way of drawing the distinction between *x*'s constitutively depending on *y* and *x*'s causally depending on *y* is on the grounds that the former sort of dependence, but not that latter, entails that *x* is at least partly constituted by *y*. Once we put aside the idea that *x* constitutively depends on *y* only if *y* is partly constitutive of *x*, what is it that grounds the distinction? That we do distinguish the two kinds of dependence is important, at least from the externalist's perspective. As Burge himself points out, 'It is trivial that many mental states causally depend on relations between environment and individual. Acquiring such states depends on being caused to have them' (Burge, 2010: 64). But this fact alone does not suffice to vindicate externalism's driving intuition. To claim that some intentional states are partly individuated in virtue of relations that a subject bears to a context is to claim a kind of dependence between those states and relations to a context which is more than merely causal.

On Burge's view, there is still room for a distinction between constitutive and causal dependence even once we give up the idea that *x*'s constitutively depending on *y* implies that *x* is (at least partially) constituted by *y* (Burge, 2010: 66-67). An organ's being a heart depends necessarily on its having the function of pumping blood through a circulatory system. Any organ that lacks this function is not a heart. An organ can have this function only if it bears certain relations to things outside of its physical boundaries—blood, blood vessels, and so on. These relations are not properly speaking causal relations; it is not that an organ's bearing these relations *causes* it to be a heart. But by the same token, there is no sense in which these relations are *constitutive* of a heart. Even though something is a heart only if it bears certain relations to things outside itself, a heart is not, either in part or in whole, a relation to things outside of itself.

I think that Burge is right to draw a distinction between *x*'s constitutively depending on *y* and *x*'s being constituted by *y*. Moreover, I think that this distinction grounds a compelling explanation of why the externalist is not committed to the claim that some mental states are located in part outside the physical boundaries of a subject's body.[6] Some mental states constitutively depend on the obtaining of certain relations to a context. But these relations are in no way constitutive of those states.

1.2 Three Burgean Arguments in Support of Externalism's Driving Intuition

In this subsection and the next I will consider several thought experiments, offered by Burge and Wittgenstein, respectively, which prima facie support externalism's driving intuition. To be clear, it will not be my concern to argue that these thought experiments do *in fact* support externalism's driving intuition or otherwise show what Burge and Wittgenstein take them to show. The success of the thought experiments is something that I will take for granted, not because I believe they are necessarily beyond dispute but because addressing even the most salient objections would lead the discussion too far afield (I will, however, refer the reader where appropriate to literature in which these objections are raised).

In this subsection, I want to consider three thoughts experiments offered by Burge in support of externalism's driving intuition. The discussion will make it clear that these thought experiments complement one another, each one extending the scope of Burge's externalism to a particular group of subjects.

The first thought experiment is set out by Burge in 'Individualism and the Mental'. The experiment has three phases. In the first phase we are asked to consider Alf, who has a set of beliefs which he expresses using the term 'arthritis'. Alf expresses these beliefs by means of utterances such as 'I've had arthritis for years now' and 'The arthritis in my fingers is more painful

[6] Another way in which the externalist can deny that they are committed to this claim is by denying that mental states have *any* spatial location. Colin McGinn (1989) has defended this strategy.

than the arthritis in my ankles'. Interpreted as expressing beliefs about *arthritis*, these utterances are true. Alf *has* had arthritis for years now and the arthritis in his fingers *is* more painful than the arthritis in his ankles. During a visit to his doctor Alf reports 'My arthritis has spread to my thigh'. Now, interpreted as expressing a belief about arthritis, this utterance is clearly false. Since arthritis is an illness which only afflicts the joints, it cannot have spread to Alf's thigh. When Alf reports 'My arthritis has spread to my thigh' is he expressing a false belief about arthritis, or is he expressing a belief about some other ailment altogether? Burge's own view is that standard practice is to attribute to Alf a belief about arthritis.[7] Because Burge believes that standard practice regarding belief attributions is instructive with respect to a person's actual mental content, he thinks, and invites the reader to agree, that the content of the belief which Alf expresses when he utters 'My arthritis has spread to my thigh' is *that my arthritis has spread to my thigh*. Because the content of a belief is conceptual content—content which involves concepts which the speaker must understand well enough to exercise in thought and reasoning if they are to have mental states involving that content—we can draw a conclusion about the concept which Alf uses the term 'arthritis' to express, namely, that he uses the term to express the concept arthritis.

In the second phase of the thought experiment we imagine a physical duplicate of Alf who inhabits a Twin Earth. Twin Alf, as we will call him, is indiscernible from Alf in every non-intentional intrinsic respect. Like Alf, Twin Alf is disposed to express a set of beliefs using the term 'arthritis'. And like Alf, he reports to his doctor that 'My arthritis has spread to my thigh'. On Twin Earth, however, the term 'arthritis' refers not to an ailment that only afflicts the joints but to any rheumatoid ailment whatsoever.

The claim in the third phase of the thought experiment is that we would not attribute to Twin Alf a belief about arthritis: 'The word 'arthritis' [on Twin Earth] does not mean arthritis ... However we describe the patient's attitudes [on Twin Earth], it will not be with a term or phrase

[7] On Burge's view, attributing a belief about, say, arthritis to someone consists in ascribing to that person a content clause with the term 'arthritis' in what Burge calls *oblique position*. A term functions in oblique position if it cannot automatically be substituted for an extensionally equivalent term without altering the truth-value of the ascription. See Burge (1979).

extensionally equivalent with 'arthritis'' (Burge, 1979: 79). Instead, Burge thinks, we would attribute to Twin Alf a belief about *tharthritis*, where 'tharthritis' is a term which refers to any rheumatoid ailment whatsoever. Given that practices of attribution are instructive with respect to actual mental content, we should conclude that the content of the belief which Twin Alf expresses when he says 'My arthritis has spread to my thigh' is not *that my arthritis has spread to my thigh* but *that my tharthritis has spread to my thigh*. Because the content of a belief is conceptual content, we should conclude that Twin Alf uses 'arthritis' to express, not the concept arthritis, but the concept tharthritis.[8]

The conclusion of the thought experiment, then, is that Alf and Twin Alf express different beliefs when they report 'My arthritis has spread to my thigh'. Alf uses the utterance to express a belief about arthritis, while Twin Alf uses it to express a belief about tharthritis. Moreover, they use the term 'arthritis' to express two different concepts. Alf uses it to express the concept arthritis, while Twin Alf uses it to express the concept tharthritis. Crucially though, these differences cannot be attributed to any non-intentional intrinsic difference between Alf and Twin Alf, for ex hypothesi they are exactly similar in this respect. The only relevant difference, it seems, is a difference in the practices of the socio-linguistic communities to which Alf and Twin Alf are respectively related, and it is this, Burge thinks, to which we ought to attribute the difference in mental states. In this way, the thought experiment supports externalism's driving intuition.

Hereafter I shall refer to this thought experiment as the *arthritis case*. One salient feature of the arthritis case is that it relies on a claim which hereafter I am going to call *Misunderstanding*:

Misunderstanding: A subject can have thoughts involving concepts which she misunderstands.

[8] For an objection to this conclusion, see Georgalis (1999).

Why think that the arthritis case relies on *Misunderstanding*? Let us approach this question by way of another: What is it to misunderstand a concept? Burge himself has little to say in response to this latter question. He writes:

> The notions of misconception, incomplete understanding, conceptual or linguistic error, and ordinary empirical error are to be taken as carrying little theoretical weight. I assume that these notions mark defensible, common-sense distinctions. But I need not take a position on available philosophical interpretations of these distinctions. (Burge, 1979: 88)

(I assume that these notions are to be taken as carrying little theoretic weight in the sense that they are not to be understood in terms of any theory. They play a critical role in Burge's views about the individuation of mental content, and so carry theoretical weight in that sense). I am not going to attempt to specify a set of necessary and sufficient conditions for misunderstanding a concept. I do, however, want to say a little more than Burge about what is distinctive about paradigmatic cases of misunderstanding, given a Burgean framework.

Paradigmatic cases of misunderstanding the concept <u>*C*</u> seem to be cases in which the subject makes two different sorts of errors.[9] The first sorts of errors are errors about what is necessarily true or necessarily false of things which come under the extension of <u>*C*</u>. The second sorts of errors are errors about what the experts or fully competent subjects hold to be true about *C*'s. I will discuss each sort of error in turn. With regards to the first sort, the subject might believe that some proposition, *p*, expresses a necessary truth about *C*'s when it in fact does not (for example, she might believe that the proposition 'Dragonflies are small birds' expresses a necessary truth about dragonflies). Alternatively, she might falsely believe that *p* expresses a necessary falsehood about *C*'s (for example, she might believe that the proposition 'Dragonflies have wings' expresses a necessary falsehood about dragonflies).[10] In both kinds of cases, the subject's error brings into question her grasp of the concept <u>*C*</u>. Compare such cases with cases in which the subject's error has to do only with what is contingently true or false of *C*'s. Suppose, for example,

[9] To be clear, the thought is that it is only a paradigmatic case of misunderstanding if the subject makes *both* sorts of errors.

[10] We can imagine more nuanced cases still. For instance, a subject might believe that *p* expresses a contingent truth about *C*'s when it in fact expresses a necessary truth.

that I believe that the proposition 'There are no dragonflies in the southern hemisphere' expresses a contingent truth about dragonflies. In this case, I have a false belief about dragonflies. But my error is not such that I am wrong about what must be true or false of something in order for it to come under the extension of the concept dragonfly. My error does not impugn my understanding of the concept.

I said that the first sorts of errors are errors about what is necessarily true or necessarily false of things which come under the extension of *C*. In fact, I think that only *certain kinds* of errors about what is necessarily true or necessarily false of things which come under the extension of *C* are involved in paradigmatic cases of misunderstanding. Burge distinguishes two sorts of necessarily true propositions, *normative characterisations* and *meaning-giving characterisations*. Normative characterisations are 'statements about *what Xs are* that purport to give basic, "essential," and necessary true information about *Xs*' (Burge, 1986b: 703). The sentence 'The atomic number of gold is 79' is an example. It purports to give necessarily true information about gold. Meaning-giving characterisations, which are a subset of normative characterisations, are statements which are central to giving the linguistic meaning of an expression and establishing norms for conventional linguistic understanding. In order to fully grasp the linguistic meaning of an expression, one must grasp the meaning-giving characterisations associated with that expression. Not all normative characterisations are meaning-giving in the relevant sense. Although 'The atomic number of gold is 79' purports to give necessarily true information about gold, it does not contribute to the linguistic meaning of the term 'gold'; someone could fully grasp that meaning without knowing that the atomic number of gold is 79. Compare 'Fatigue is tiredness brought on by exertion' or 'To walk is to move on foot at a natural, unhurried pace'. A person who does not hold these characterisations true does not fully understand the linguistic meaning of 'fatigue' or 'walk' (which is not to say that they do not understand the meaning well enough to have thoughts involving the relevant concepts).

The kinds of errors about what is necessarily true or necessarily false of things that come under the extension of *C* which are involved in paradigmatic cases of misunderstanding are errors

involving the meaning-giving characterisations associated with the term '*C*' (as opposed to errors involving those normative characterisations associated with '*C*' which are not meaning-giving). If, for example, someone believes that 'To walk is to jog quickly' expresses a necessary truth about walking, or that 'To walk is to move on foot at a natural, unhurried pace' expresses a necessary falsehood about walking, then they misunderstand the concept walk. In contrast, someone who believes that 'The atomic number of gold is 47' expresses a necessary truth about gold, or that 'The atomic number of gold is 79' expresses a necessary falsehood about gold, may nevertheless understand the concept gold perfectly well.

The second sorts of errors involved in paradigmatic cases of misunderstanding *C* are errors about what the experts or fully competent subjects hold to be true about *C*'s. Normally, a subject who misunderstands the concept *C* incorrectly believes that her beliefs about *C*'s are shared by the experts or those speakers who are fully competent. Normally, a subject who believes that the statement 'Dragonflies are small birds' expresses a necessary truth about dragonflies will count as misunderstanding the concept dragonfly only if she believes that dragonfly experts also believe that the statement expresses a necessary truth about dragonflies. If she does not have some further belief to this effect, then she is not fairly described as misunderstanding the concept.

To summarise, paradigmatic cases of misunderstanding a concept *C* are cases in which the subject makes two different sorts of errors. The first sorts of errors are errors, involving the meaning-giving characterisations associated with '*C*', about what is necessarily true or false of those things which fall under the extension of *C*. The second sorts of errors are errors about what the experts hold to be true of *C*'s. In paradigmatic cases of misunderstanding, the subject believes that her beliefs about *C*'s are shared by those speakers who are fully competent.[11]

Why think that the arthritis case relies on *Misunderstanding*? The conclusion of the first phase of the thought experiment is that when Alf says 'My arthritis has spread to my thigh' he is

[11] Typically, the experts will constitute a subset of all speakers. But this will not be so in every case. We can expect there to be at least some cases in which all or almost all participants are experts in the relevant sense (that is, cases in which all or almost all participants have a reflective grasp of the meaning-giving characterisations associated with the relevant term).

expressing the belief that his *arthritis* has spread to his thigh. But if this is indeed Alf's belief, then he must believe that it is true that arthritis can afflict one's thigh. But it *isn't* true that arthritis can afflict one's thigh. In fact, it is necessarily false. 'Arthritis is an inflammation of the joints' is a meaning-giving characterisation associated with the term 'arthritis'. So if Alf really does believe that his arthritis has spread to his thigh, then Alf must be in error about the meaning-giving characterisations associated with the term 'arthritis'. Moreover, it seems that Alf believes that medical professionals agree that it is true that arthritis can afflict one's thigh. Alf defers to his doctor when he uses the term 'arthritis'. He intends to use the term as his doctor uses it.[12] He does not take himself to using the word in an idiosyncratic way. In light of these considerations, and given the characterisation of paradigmatic cases of misunderstanding I outlined above, we should say that Alf misunderstands the concept arthritis. Consequently, the conclusion of the first phase of the arthritis case—that Alf believes that his arthritis has spread to his thigh—can only be true if misunderstanding a concept is consistent with understanding it sufficiently well to exercise it in thought and reasoning. So the conclusion of the first phase of the arthritis case presupposes that *Misunderstanding* is true.

To be clear, the arthritis case relies not just on the *possibility* of Alf's misunderstanding the concept arthritis, but on his *actually* doing so. If Alf did not misunderstand the concept then he would not have reported to his doctor that 'My arthritis has spread to my thigh'. If he had not made this report, then Alf and Twin Alf would no longer be exactly similar in every non-intentional intrinsic respect and hence the thought experiment would not show that two individuals who were exactly similar in every non-intentional intrinsic respect could have different beliefs. Consequently, it would not support externalism's driving intuition.

Because the thought experiment relies on Alf's misunderstanding the concept arthritis, however, it has limited scope. Specifically, it does not give us a reason to accept externalism about

[12] That Alf does defer to the experts when he uses the term 'arthritis' is borne out by the way in which he responds to correction: 'In examples like ours, [the subject] typically admits his mistake, changes his views, and leaves it at that'. (Burge, 1979: 95)

subjects who belong to one of two groups. The first group consists of subjects who grasp the meaning-giving characterisations associated with a particular term but who lack expert knowledge about the relevant kind. In paradigmatic cases of misunderstanding, a subject is in error concerning the meaning-giving characterisations associated with the relevant term. In paradigmatic cases where a subject lacks expert knowledge, in comparison, the subject is merely ignorant of certain normative characterisations associated with the term.[13] A subject who is merely ignorant of the fact that the atomic number of gold is 79 does not thereby misunderstand the concept gold. Similarly, a subject who is merely ignorant of, for example, the fact that some dragonflies can live beyond five years does not thereby misunderstand the concept dragonfly. Nevertheless, their knowledge about gold and dragonflies, respectively, is clearly limited. The second group to which the arthritis case does not apply consists of speakers who are fully competent with respect to the relevant concept. Necessarily, such speakers do not misunderstand the concept.

Burge has two further thought experiments, both structurally identical to the arthritis case, which broaden the scope of his externalism to include some members of these two groups of speakers. The first of these arguments which I want to consider is set out by Burge in his article 'Other Bodies'. Suppose that Carl is a competent speaker with respect to the term 'water'. He is familiar with many, but not all, of the normative characterisations associated with the term (and endorses as true those characterisations with which he is familiar). One of the normative characterisations with which Carl is not familiar is 'The molecular structure of water is H_2O'. It is not that Carl believes that the molecular structure of water is *not* H_2O. He does not have an opinion either way. Some of the speakers in the linguistic community to which Carl belongs are water 'experts' who know that the molecular structure of water is indeed H_2O. Carl is aware that water 'experts' exist and he defers to these speakers when he uses the term 'water'. Because Carl defers to the 'experts' when he says, for instance, 'I feel like a drink of water', it is natural, Burge

[13] Of course, cases in which a subject lacks expert knowledge are likely to be cases in which the subject is also ignorant of a range of contingently true statements about the relevant kind.

wants to say, to attribute to him a belief about water, in spite of his ignorance about water's molecular structure. Given that we find this attribution natural, we can conclude that the content of the belief which Carl expresses when he asserts 'I feel like a drink of water' is *that I feel like a drink of water*, from which we can infer that Carl uses the term 'water' to express the concept water.

Now imagine Twin Carl, an inhabitant of a Twin Earth, who is exactly similar to Carl in every non-intentional intrinsic respect. On Twin Carl's Twin Earth there is a substance that the inhabitants refer to as 'water', which possesses all the immediately observable qualities associated with water on Earth but a different molecular structure. The molecular structure of the substance referred to as 'water' on Twin Earth is not H_2O but, say, XYZ. Let us assume that it is correct to say that the substance on Twin Earth is not water but something else, *twater*, say. Twin Carl is ignorant of the molecular structure of twater. Nevertheless, some of the speakers in the linguistic community to which Twin Carl belongs have determined that twater's molecular structure is XYZ. Twin Carl is aware that such twater 'experts' exist and he defers to them when he uses the term 'water'. Burge contends that when Twin Carl says 'I feel like a drink of water' we would not attribute to him a belief about water, but a belief about twater. The content of the belief which Alf expresses when he asserts 'I feel like a drink of water' is *that I feel like a drink of twater*. Unlike Carl, Twin Carl uses the term 'water' to express the concept twater

Since Carl and Twin Carl are physical duplicates, these differences cannot be attributed to any non-intentional intrinsic difference between them. The conclusion of the thought experiment is that they are due, as in the arthritis case, to differences in the contexts to which Carl and Twin Carl are respectively related. In the arthritis case it was differences between the practices of Alf and Twin Alf's respective socio-linguistic communities that accounted for the difference in their mental states. In comparison, one of the lessons of the thought experiment we have just considered—hereafter the *water case*—is that differences in respective physical contexts can be equally significant. The socio-linguistic communities to which Carl and Twin Carl respectively belong are obviously dissimilar and this dissimilarity plays an important role in the thought

experiment. The experts on Earth mean something different by 'water' than do the experts on Twin Earth. This difference accounts for the difference in Carl and Twin Carl's mental states. But what accounts for the fact that the experts on Earth and Twin Earth mean different things by the term 'water' are differences between their respective *physical* contexts, namely, the fact that on Earth there is water but no twater, while on Twin Earth there is twater but no water. In Burge's own words, 'The difference in [Carl and Twin Carl's] mental states and events seems to be a product primarily of differences in their physical environments, mediated by differences in their social environments—in the mental states of their fellows and conventional meanings of words they and their fellows employ' (Burge, : 87).

The point of difference between the two thought experiments which I want to emphasise, however, is that unlike the arthritis case, the water case does not rely on *Misunderstanding*. Carl's knowledge of water is limited, but he does not misunderstand the concept water. Unlike Alf, Carl is not in error concerning the meaning-giving characterisations associated with the term 'water'. He is merely ignorant of some of the normative characterisations associated with the term. Moreover, unlike Alf, Carl does not have false beliefs about which statements fully competent speakers regard as meaning-giving with respect to the term 'water'. Carl acknowledges that there are such subjects. But he is agnostic about what it is that they believe.

Although it does not rely on *Misunderstanding*, the water case does rely on a similar principle, which we might call *Lay Knowledge*:

> ***Lay Knowledge****:* A subject can have thoughts involving the concept *C* even though she is ignorant of certain of the normative characterisations associated with the word '*C*'.

Misunderstanding and *Lay Knowledge* make claims about the degree of understanding or knowledge which is consistent with exercising a particular concept in thought and reasoning.

Specifically, they claim that misunderstanding a concept C or lacking expert knowledge about *C*'s is consistent with having thoughts involving C. Neither principle, however, tells us anything about *how it is* that subjects who either misunderstand C or lack expert knowledge of *C*'s are able to have thoughts involving C. This task falls to a further claim—*Deference*—on which both the arthritis and water cases rely:

> ***Deference***: If the concept C is standardly associated with the term '*C*' by the experts, then a subject who either misunderstands C or lacks expert knowledge of *C*'s can, by deferring to those experts, use '*C*' to express C.

In 'The Meaning of 'Meaning'', Hilary Putnam claims that a subject who has less than full competence with a term can nevertheless use it to mean what the experts use it to mean. For example, a subject who cannot discriminate between samples of gold and samples of fool's gold can nevertheless use the term 'gold' to mean gold (as opposed to *gold or fool's gold*). They can manage this, Putnam claims, by deferring to the experts when they use the term 'gold', that is, by using 'gold' with the intention of picking out an object which has the same essential nature as those objects which are picked out by the term 'gold' as it is standardly used by the experts. Burge thinks that the analogous thing is true of concepts. A speaker who cannot discriminate between samples of gold and samples of fool's gold can nevertheless use the term 'gold' to express the concept gold, on Burge's view, by deferring to the experts.[14]

Alf and Carl defer to the experts when they use the terms 'arthritis' and 'water', respectively. The fact that they do so explains how it is that they are able to use these terms to express the concepts which the experts use them to express. To this extent, the arthritis and water

[14] Putnam and Burge disagree about the level of understanding which a speaker (who defers to the experts) must have in order to mean by a term what the experts commonly use it to mean. The level of understanding which Putnam thinks is necessary is higher than the level of understanding which Burge thinks is necessary. For example, on Putnam's view, in order to use the term 'tiger' to mean tiger, a speaker 'is required to know that stereotypical tigers are striped' (Putnam, 1975b: 248). On Burge's view, in comparison, 'A user of the term 'tiger' who is acquainted only with albino tigers and never learns that most tigers have stripes can use the term 'tiger' competently, and synonymously with our term 'tiger'' (Burge, 2013f: 236).

cases rely, not just on the possibility of Alf and Carl deferring to the experts, but on their *actually* doing so. If they did not defer to the experts, then there would not be the reason Burge thinks that there is for concluding that Alf and Carl have thoughts involving the concepts arthritis and water, respectively. Because the thought experiments are so reliant, their scope is limited. Specifically, neither case applies to subjects who are fully competent with respect to the relevant term. Such speakers do not defer to the experts; they *are* the experts.[15]

Burge has a third thought experiment which widens the scope of his externalism so that it encompasses the mental states of some experts. This thought experiment, which Burge sets out in 'Intellectual Norms and Foundations of Mind', is as follows. We begin by imagining an individual, Tom, an expert speaker with regard to the term 'sofa' who at some point comes to question the meaning-giving characterisations associated with the term 'sofa'. Specifically, he comes to question whether the characterisation 'Sofas are items of furniture of such-and-such construction meant primarily for sitting' is true. If you were to ask him, Tom would put his position thus: 'Sofas are not items of furniture, but religious artefacts'. Tom readily concedes that most people, including sofa 'experts', understand sofas to be items of furniture of such-and-such construction meant primarily for sitting. He just thinks that these people are systematically wrong. As it turns out, the sofa experts are not wrong. Sofas really *are* items of furniture, not religious artefacts. The question is, when Tom declares 'Sofas are not items of furniture, but religious artefacts' is he expressing a belief about sofas, or about something else altogether? Burge's contention is that Tom is doing the former. The content of the belief Tom expresses when he declares that 'Sofas are not items of furniture, but religious artefacts' is *that sofas are not items of furniture, but religious artefacts*. Burge thinks that in spite of his strange views about sofas, Tom uses the term 'sofa' to express the concept sofa.

[15] Of course, the scope of the arthritis and water cases is limited in certain other respects. Because they rely on *Deference*, they will not apply to *every* subject who misunderstands the relevant concept or lacks the relevant expert knowledge. Specifically, they will not apply to those subjects who misunderstand or lack expert knowledge but who do not defer to the experts when they use of the relevant term.

Next we imagine Twin Tom, who is exactly similar to Tom in every non-intentionally described intrinsic respect. Twin Tom inhabits a Twin Earth on which 'sofa' refers to objects which actually are, and are widely acknowledged to be, religious artefacts. There are no sofas on Twin Earth. Let us call the objects to which 'sofa' refers on Twin Earth *safos*. Whilst Twin Tom has the true belief that safos are religious artefacts, he also has the false belief that others incorrectly believe them to be items of furniture of such-and-such construction meant primarily for sitting. Perhaps he has misinterpreted jokes people have made about safos being items of furniture as genuine expressions of belief. Like Tom, if we were to ask Twin Tom, he would express his views by telling us that 'Sofas are not items of furniture, but religious artefacts'. Burge's contention is that, unlike Tom, when Twin Tom utters this sentence he is expressing a belief about safos. The content of the belief which the utterance expresses on Twin Tom's lips is not *that sofas are not items of furniture, but religious artefacts* but *that safos are not items of furniture, but religious artefacts*. While Tom uses the term 'sofa' to express the concept <u>sofa</u>, Twin Tom uses the term to express the concept <u>safo</u>.

Since Tom and Twin Tom are physical duplicates, it seems these differences cannot be attributed to any non-intentional intrinsic difference between them. Rather, they appear to be due, as they are in the water case, to differences in the social and physical contexts to which Tom and Twin Tom are respectively related. Unlike the water case, however, the thought experiment we have just considered—hereafter the *sofa case*—does not depend on *Lay Knowledge*. Nor does it depend on *Misunderstanding*. Tom neither misunderstands the concept <u>sofa</u> nor lacks expert knowledge about sofas. He simply has an unorthodox theory about sofas.

The sofa case does, however, depend on the following claim:

> ***False Belief***: If a proposition, *p*, is a meaning-giving characterisation associated with the word '*C*', then an otherwise competent speaker who believes that *p* is false can still have thoughts involving the concept <u>*C*</u>.

Why think that the sofa case depends on *False Belief*? The proposition that sofas are items of furniture of such-and-such construction meant primarily for sitting is, Burge wants to say, a meaning-giving characterisation associated with the term 'sofa'. Tom believes that this proposition is false. Nevertheless, when he asserts 'Sofas are not items of furniture but religious artefacts' he is, according to Burge, expressing a belief involving the concept sofa. This claim holds only if an otherwise competent subject who believes that a proposition, which is a meaning-giving characterisation associated with the term 'sofa', expresses a falsehood about sofas, can have thoughts involving the concept sofa. In other words, the claim holds only if *False Belief* is true.

False Belief does not tell us *how it is* that a subject who believes that a proposition, which is a meaning-giving characterisation associated with '*C*', expresses a falsehood about *C*'s, can have thoughts involving the concept *C*. Presumably, there are going to be cases in which it is the fact that the subject defers to the experts when they use the term '*C*' which explains how they can come to use the term to express *C*, in spite of their aberrant beliefs. But obviously, the sofa case is not among them. Tom is an expert. Necessarily, he does not defer to more competent speakers when he uses the term 'sofa'.

Burge thinks that direct causal relations—causal relations not mediated by deferential relations—linking a subject with an objective subject matter can play a role in individuating a subject's mental states. The direct causal relations which are of primary interest to Burge are those at work in perception and in perceptually backed, demonstrative applications of empirically applicable terms (for example, 'This book, 'That chair') (Burge, 2003a: 347). Both perception and perceptually backed demonstrative applications depend for their success on the obtaining of causal relations connecting the subject to an objective subject matter. On Burge's view, these relations can play an individuating role with respect to a subject's mental content. More precisely, Burge endorses the following claim:

Direct Causal Contact: Direct causal relations at work in perception and in perceptually-backed demonstrative applications of an empirically applicable

> term, '*C*', which connect a subject with actual *C*'s, can bring it about that she uses '*C*' to express the concept <u>*C*</u>.

It is the fact that Tom makes regular perceptually-backed, demonstrative applications of the term 'sofa' to actual sofas which ultimately explains, on Burge's view, how it is that he has thoughts involving the concept <u>sofa</u>, in spite of his nonstandard theory about sofas.

Before concluding this subsection, I want to say something very briefly about the connection, on Burge's view, between *Deference* and *Direct Causal Contact*. On Burge's view, the individuating role played by direct causal relations is in an important sense more fundamental than the individuating role played by deferential relations. According to Burge:

> The ultimate source of [externalism], even in cases of social dependence, is the cognitive distance between individual and subject matter in any objective representation. This distance must be bridged by non-representational relations between individual and objective subject matter. (Burge, 2007b: 24)

The non-representational relations which Burge is referring to here are the causal relations at work in perception and in perceptually backed, demonstrative applications of empirically applicable terms. Burge thinks that causal relations of this sort are ultimately responsible for individuating a subject's mental content, even in cases of social dependence.[16] To play an individuating role, the causal relations at work in perception and perceptually backed, demonstrative applications of a term need not connect the subject to the subject matter directly. Indeed, 'they need not occur in the individual's own life. The relations to the environment can route through other people, or through the evolution of an individual's perceptual system' (Burge, 2007b: 3). According to *Deference*, a subject can use a term to express the concept which the experts standardly use it to express by deferring to the experts. But they can do this, Burge thinks, only insofar as there already exist causal connections linking the experts to the subject matter directly (I will have more to say about

[16] Why does Burge think that the gap between subject and subject matter must be bridged by *non-representational* relations? In summary, Burge thinks that non-representational relations are necessary for certain non-conceptual abilities—including the ability to make perceptually backed, demonstrative applications of an empirically applicable term—which are in turn necessary (he thinks) for conceptual thought. This line of reasoning is expanded upon in Burge (1977, 2007d).

the roles which direct causal and deferential relations play in individuating a subject's mental states, on Burge's view, in Chapter 2, Section 1).

Taken together, the thought experiments we have considered in this subsection support externalism's driving intuition with respect to the mental states of three groups of speakers. The first group consists of speakers who misunderstand the relevant concept and who defer to more competent speakers when they use the relevant term. The second group consists of speakers who lack expert knowledge of the relevant kind and defer to more competent speakers when they use the relevant term. The third group consists of fully competent speakers who have extensive and prolonged direct causal contact with *C*'s and believe that certain meaning-giving characterisations associated with the word '*C*' are false. The first thought experiment we considered, the arthritis case, applies to speakers in the first group, but not to speakers in the second or third. The second thought experiment we considered, the water case, applies to speakers in the second group, but not to those in the first or third. The third thought experiment we considered, the sofa case, applies to speakers in the third group, but not to those in the first or second. In this way, the arthritis, water and sofa cases complement one another and in so doing serve to broaden the scope of Burge's externalism.

1.3 Wittgenstein and Externalism's Driving Intuition

We have been discussing the relations between three different thought experiments which Burge offers in support of externalism's driving intuition. In this subsection I want to consider two thought experiments drawn from Wittgenstein's later work which prima facie also support the intuition. The discussion will aim at highlighting some of the points of similarity and difference between Wittgenstein's thought experiments and the thought experiments considered in Subsection 1.2. As I made clear in the introduction to the thesis, I am not defending the view that Wittgenstein is *actually* an externalist or that he sets out the thought experiments I am going to consider with the intention of supporting something like externalism's driving intuition. My concern is not purely,

or even primarily, exegetical. It is instead to answer the following question: To what extent is an externalism which cleaves closely—more closely than does Burge's own externalism—to Wittgenstein's views about the individuation of mental states vulnerable to objections from self-knowledge? Showing that certain of Wittgenstein's thought experiments can at least be interpreted in such a way so as to support externalism's driving intuition is an important first step in answering this question.

The thought experiments expressed in the following two passages are representative of the sorts of thought experiments I have in mind:

> Let us imagine a god creating a country instantaneously in the middle of the wilderness, which exists for two minutes and is an exact reproduction of a part of England, with everything that is going on there in two minutes. Just like those in England, the people are pursuing a variety of occupations. Children are in school. Some people are doing mathematics. Now let us contemplate the activity of some human being during these two minutes. One of these people is doing exactly what a mathematician in England is doing, who is just doing a calculation.—Ought we to say that this two-minute-man is calculating? Could we for example not imagine a past and a continuation of these two minutes, which would make us call the processes something quite different? (*Remarks on the Foundations of Mathematics*, hereafter RFM, VI, §35)

> Now suppose I sit in my room and hope that N.N. will come and bring me some money, and suppose one minute of this state could be isolated, cut out of its context; would what happened in it then not be hoping? –Think, for example, of the words which you may utter in this time. They are no longer part of this language. And in different surroundings the institution of money doesn't exist either. (*Philosophical Investigations*, hereafter PI, §584)[17]

There are obvious similarities between the thought experiments which are set out in these passages and the arthritis, water and sofa cases which we considered in Subsection 1.2. Wittgenstein's thought experiments encourage the thought that two subjects who are in certain important respects indiscernible, but who happen to be situated in contexts which are relevantly different, could differ with respect to their mental states. But in precisely what sense are the subjects in Wittgenstein's thought experiments indiscernible? Burge's arthritis, water and sofa cases each involve us imagining two subjects who are indiscernible in every non-intentionally

[17] Unless otherwise stated, passages cited from *Philosophical Investigations* are from the revised 4th edition.

described intrinsic respect. Are the subjects in Wittgenstein's thought experiments indiscernible in this same sense?

The answer to this question is 'No'. Consider, for instance, the thought experiment set out in *Remarks on the Foundations of Mathematics* VI, §35. It is not part of this thought experiment that we are to imagine two subjects who for a two minute period are indiscernible in *every* non-intentionally described intrinsic respect. For example, it is not part of the thought experiment that we are to imagine two subjects who are indiscernible with respect to their non-intentionally described *dispositions*. We are to imagine that the two-minute man is doing exactly what the mathematician in England is doing for a period of two minutes. But the mathematician's non-intentionally described dispositions are not among the things he is *doing* during the two minutes. Wittgenstein would be happy to include differences in their dispositions, non-intentionally described, among the differences between the mathematician and the two-minute man.

Wittgenstein asks whether we could imagine a past and a continuation of the two minutes which would make us call what the two-minute man is doing something other than calculating. Clearly, the answer he wishes the reader to give is 'Yes'. But his basic point is not a point about whether we would *call* what the two-minute man is doing calculating (given such-and-such a past and continuation). Rather, it is a point about whether what the two-minute man is doing *is* calculating (as we shall see, on Wittgenstein's view there is a constitutive relation between those things which warrant our describing *S* as, say, calculating and *S*'s actually calculating). Wittgenstein's basic point is that two subjects who, for a period of two minutes, have the same things come before their minds, who engage in the same verbal and non-verbal behaviour—where these things are described in terms which do not take it for granted that the subjects are indeed calculating—might nevertheless differ with respect to whether they are engaged in the activity of calculating. This point is not threatened by our imagining two subjects who differ with respect to their non-intentionally described dispositions.

Nor is it threatened by our imagining two subjects who are indiscernible with respect to certain of their *intentionally* described properties. It is essential for Wittgenstein's thought experiment that we describe the subjects' verbal and non-verbal behaviour, the things that come before their minds, and so on, in terms which do not take it for granted that the subjects are indeed calculating. But it is *not* essential that we describe these things in non-intentional terms. After all, there are going to be a range of intentional descriptions which do not take it for granted that the subjects are indeed calculating. For example, we could describe what the mathematician is doing as writing down marks on a piece of paper. This description leaves it open whether the mathematician is calculating, but it is nevertheless an intentional description.

These conclusions about the way to understand the first passage should inform our reading of the second. In the second passage I cited, we are asked to imagine one minute of Wittgenstein's state of hoping that N.N. will come and give him some money cut out of its actual context. If we were being asked to imagine the one-minute extract under a description which takes it for granted that it is an extract of Wittgenstein's state of hoping that N.N. will come and give him some money, then the thought experiment would not make any sense. Rather, what we are being asked to imagine is a one-minute extract of Wittgenstein's state described in terms which do not take it for granted that it is a state of hoping. Could we imagine some other context within which this extract would not constitute hoping?[18] Again, the answer which Wittgenstein clearly wishes the reader to give is 'Yes'.

One might think that, unlike the first thought experiment, it is a feature of the second thought experiment that we are to imagine two subjects who are indiscernible with respect to their non-intentionally described dispositional properties. After all, we are being asked to imagine a one-minute extract of Wittgenstein's state. One might think that such an extract will obviously include facts about Wittgenstein's non-intentionally described dispositions. In fact, I do not think that this is obviously what Wittgenstein has in mind. I think it is plausible that the extract is simply

[18] That Wittgenstein wishes us to imagine the one-minute extract in some other context—as opposed to in isolation from *any* context—is strongly suggested by the passage's final line.

to include the things which come before Wittgenstein's mind and his verbal and non-verbal behaviour, where these things are described in terms which do not beg the question at issue. But even if it is a feature of this particular thought experiment that we are to imagine two subjects who are indiscernible with respect to their non-intentionally described dispositional properties, it is not an *essential* feature. Wittgenstein would be happy to run a version of the thought experiment in which the two subjects differed with respect to their non-intentionally described dispositional properties. Such a version would, on his view, still succeed in making his basic point. It would still succeed in showing that two subjects who, for a period of a minute, have the same things come before their mind, who engage in the same non-intentionally described verbal and non-verbal behaviour (where these are described in terms which do not take it for granted that the subjects are hoping) might nevertheless differ with respect to whether they are hoping.

This is a point of difference between Burge's thought experiments and Wittgenstein's. Burge is explicit that when considering the arthritis, water and sofa cases we are to imagine two subjects who are indiscernible with respect to their non-intentionally described dispositions.[19] Moreover, I think that Burge regards our doing so to be an essential feature of these cases. He would no doubt be happy to run versions of the thought experiments where there are dissimilarities between the subjects' non-intentionally described dispositions. But I do not think that he would consider these versions to succeed in making the same basic point which he takes to be made by the versions of the thought experiments which he actually gives.

Clearly, then, this difference between Burge and Wittgenstein's respective thought experiments is to be explained in terms of a difference between the basic point which Burge and Wittgenstein understand those thought experiments to be making. Burge takes the basic point of the arthritis, water and sofa cases to be an essentially positive one. On his view, they show that the

[19] For example: 'We further suppose that both have the same qualitative perceptual intake and qualitative streams of consciousness, the same movements, the same behavioural dispositions and inner functional states (non-intentionally and individualistically described)' (Burge, 2007c: 86); 'As a second step, imagine a person B (or A in nonfactual circumstances) who is, for all intents and purposes, physically identical to A. He has the same physical dispositions, receives substantially the same physical stimulations, produces the same motions, utters the same sounds' (Burge, 1986b: 707-708).

relations in which a subject stands to a social or natural context play a role in individuating her mental states. In comparison, Wittgenstein takes his thought experiments to be making an essentially negative point. Wittgenstein takes those thought experiments to show that whether a person is engaged in a particular activity or instantiating a particular propositional attitude does not depend exclusively on what the subject is doing or what is going on within the subject at the time, where these things are described in terms which do not beg the question at issue. It will suffice to make Wittgenstein's point—but not Burge's—to show that two subjects, who are engaged in the same behaviour, who have the same things come before their minds (where these things are described in terms which do not beg the question at issue), but who differ with respect to their non-intentionally dispositions, might differ with respect to their mental states. It will not suffice to make Burge's point because the difference in the subjects' mental states could be ultimately traceable to differences in their non-intentionally described dispositions.

I want to mention three further points of difference between Burge's thought experiments and Wittgenstein's. The first further point of difference concerns the histories of the subjects in the thought experiments. Burge is explicit that when we are thinking through his thought experiments, we are to imagine individuals who have indiscernible histories, non-intentionally described.[20] In comparison, neither of Wittgenstein's thought experiments involve us imagining subjects who are indiscernible in this respect. Rather, we are to imagine subjects who have certain commonalities at the non-intentional (and intentional) level of description for a specified period of time—two minutes in the case of the first thought experiment, one minute in the case of the second. It is not part of Wittgenstein's thought experiments that we are to imagine the subjects having led indistinguishable lives non-intentionally described leading up to this period.

[20] For example: 'The second step of the thought experiment consists of a counterfactual supposition. We are to conceive of a situation in which the patient proceeds from birth through the same course of physical events that he actually does, right to and including the time at which he first reports his fear to his doctor. Precisely the same things (non-intentionally described) happen to him' (Burge, 1979: 77-78); 'The conclusion is that *A* and *B* are physically identical until the time when they express their views. But they have different mental states and events' (Burge, 1986b: 708).

The second further point of difference is that Wittgenstein's thought experiments do not rely on any of the three claims which we discussed in connection with the arthritis, water and sofa cases—*Misunderstanding*, *Lay Knowledge* or *False Belief*, respectively. Neither of Wittgenstein's thought experiments presuppose that either the subject who is normally contextualised or the subject who is situated in the context which differs from the normal one: can have thoughts involving concepts which they misunderstand; can have thoughts involving the concept <u>*C*</u> even though they are ignorant of certain normative characterisations associated with '*C*'; or can have thoughts involving the concept <u>*C*</u>, even though they believe that *p* is false, where *p* is a meaning-giving characterisation associated with '*C*'. As the following discussion will make clear, there are claims about the individuation of mental states on which Wittgenstein's thought experiments *do* rely. But none of these claims restrict the scope of Wittgenstein's thought experiments to particular sub-groups of speakers. Both of Wittgenstein's thought experiments apply to *any* speaker who is sufficiently competent to have a mental state with the relevant attitude or to be engaging in the relevant activity.

The third further point of difference is that unlike the thought experiments considered in Subsection 1.2, neither of Wittgenstein's thought experiments is addressed primarily to the content of the speaker's mental states. The first thought experiment is addressed primarily to the activity of calculating, the second to the propositional attitude of hoping. I say that neither thought experiment is addressed *primarily* to the content of a speaker's mental states. The final line of the second passage is clearly meant to encourage the reader towards a conclusion about the content of the one-minute extract of Wittgenstein's state. The conclusion is that if the context within which we are situating this extract is one within which the institution of money does not exist, then, regardless of whether the state would be correctly described as a state of hoping, the content of the state would not be *that N.N. would come and give me some money*. There is a claim about the individuation of mental states at work here. On Wittgenstein's view, in order to have thoughts involving, say, the concept <u>chess</u>, there must exist a practice of playing chess (or relevantly similar games) (PI §337). In the absence of such a practice, Wittgenstein thinks, there is not anything that

could count as having a thought about chess. The analogous thing is true, Wittgenstein thinks, of thoughts involving the concept money. In the absence of the institution of money (or relevantly similar institutions), there is not anything that could count as having a thought about money.

The more general claim at work here—the claim that, for a wide range of *C*'s, a subject can have thoughts involving the concept *C* only if there exists a practice of talking about *C*'s (or about things which are relevantly similar)—is but one component in Wittgenstein's positive view about the individuation of mental content. In the remainder of this subsection, I want to consider this view in more detail. The discussion will show that there is no principled reason why Wittgenstein does not give arguments structurally identical to the arguments put forward in the passages we have considered, but which are addressed primarily to mental content instead of mental attitudes and activities.

Throughout Wittgenstein's later work there is a concerted attempt to displace a particular picture of what understanding (meaning, intending, and so on) that *p* consists in. According to the picture Wittgenstein wishes to displace, what make it the case that a person understands that *p*, say, (as opposed to that *q* or that *r*) are mental events, processes or states which go on in the mind of the person when they understand (PI §152). Suppose I understand your words 'Smith is coming tonight' to mean that Smith *A* is coming tonight (as opposed to Smith *B*). In virtue of what is that the content of my understanding? According to the proponent of the picture Wittgenstein disputes, the answer is 'Solely in virtue of events, processes and states which went on within you when you understood. Perhaps you had a sensation of a particular kind, or perhaps a particular image or picture came before your mind when you understood'.[21]

Wittgenstein explicitly rejects this picture:

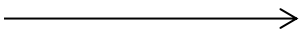
[21] Someone might think that Wittgenstein's objections to this view must be misplaced in the case of meaning. What makes it the case that I mean that *p* (as opposed to that *q* or that *r*), one might think, is that I *intend* to mean that *p*, and, one might go on, surely intending to mean that *p* is a mental event, process or state that goes on in the mind if anything is. As the discussion will make clear, this is not the case on Wittgenstein's view. The same considerations which Wittgenstein thinks show that meaning is not a mental event, process or state also show, on his view, that intending is not a mental event, process or state.

> If God had looked into our minds, he would not have been able to see there whom we were speaking of. (PPF §284)

There is an assumption at work in this passage, namely, the thought that if God *had* looked into our minds, he would only have been able to see the events, processes and states going on therein. Wittgenstein's thought is that God could not discern the content of our thoughts or utterances on the basis of those events, processes and states.

Wittgenstein's main line of objection to the idea that it is events, processes and states which went on within me when I understood that determine the content of my understanding consists of two moves. The first move is to reject the idea that whenever one understands, there is always *some* event, process or state going on within one that accompanies the understanding and which could be identified as the thing in virtue of which my understanding has the content it does. When I understood your words to mean that Smith *A* is coming tonight, perhaps an image did come before my mind. But, Wittgenstein rightly insists, it need not have. I could just as easily have understood your words in the way that I did, even if the image had not come before my mind. Wittgenstein's point is not merely that for any particular image, it is possible that I may have understood your words in the way that I did had that particular image not come before my mind. It is that I could have understood your words in the way that I did even if there was not *any* image that came before my mind when I understood. And what holds true for understanding also holds true, Wittgenstein thinks, for meaning, intending, and so on.

The second move is to insist that even if some event, process or state is going on within me when I understand, it cannot be that in virtue of which my understanding has the content which it does. Suppose, for example, that upon hearing the words 'right-pointing arrow' the following image comes before my mind:



Can it be in virtue of this image coming before my mind that I understand the words 'right-pointing arrow' to mean right-pointing arrow? Wittgenstein's response is 'No'. A recurring theme throughout *Investigations*, and Wittgenstein's later work more generally, is that considered in isolation—that is, apart from its application—any image is variously interpretable.[22] The vast majority of us find it natural to interpret the image above as an image of an arrow pointing to the right. However, according to Wittgenstein, there is nothing about the image considered in isolation in virtue of which it is correct to interpret it this way. Suppose someone found it natural to interpret the image as an 'arrow' pointing to the left. On Wittgenstein's view, there is no feature of the image considered in isolation, and described in terms which do not already presuppose that it is an image of an arrow pointing to the right, to which we can appeal to show that our way of interpreting the image is correct and their way incorrect.

Someone might object that this line of reasoning overlooks a crucial difference between mental images and images depicted in some non-mental medium. They might agree that nothing about an image drawn on a piece of paper and considered in isolation suffices to determine which of the various ways of interpreting the image are correct and which incorrect. But, they might insist, this is not true of mental images for mental images are *self-interpreting*. To support their claim they might appeal to the fact that it makes no sense for a person to ask whether the image that comes before their mind is an image of an arrow pointing to the right or an image of an 'arrow' pointing to the left.

Wittgenstein would certainly agree that it makes no sense for a person to ask these sorts of questions. But he would reject the idea that it makes no sense because mental images are self-interpreting. For Wittgenstein, the idea of a self-interpreting image is a piece of philosophical fiction, a corollary of a misleading picture of mental phenomena:

> "A mental image must be more like its object than any picture. For however similar I make the picture to what it is supposed to represent, it may still be the picture of something else. But it is an

[22] See, for instance, PI §§ 139, 663.

> intrinsic feature of a mental image that it is the image of *this* and of nothing else." That is how one might come to regard a mental image as a super-likeness. (PI §389)

The line of reasoning I have just sketched purports to show that it cannot be in virtue of a mental image coming before my mind that my understanding has the content which it does. To be clear, however, the reasoning is an example of a more general argument which is intended to apply to *any* mental event, state or process.

We commonly refer to understanding, meaning and intending as mental states. On Wittgenstein's view, however, this way of talking is, properly speaking, misconceived. We have seen that it is not in virtue of states (events or processes) going on within one that one understands that *p*, on Wittgenstein's view. And, as the following discussion will make clear, Wittgenstein does not think that understanding that *p* is a metaphysically intrinsic state (event or process). So on Wittgenstein's view, understanding that *p* is not properly thought of as a mental state (event or process) (*Zettel* §§26, 45).[23] For ease of discussion, however, I am going to continue to use 'mental state' to describe mental phenomena like meaning, understanding, intending, and so on.

If the picture according to which mental states like meaning, intending and understanding have the content which they do solely in virtue of events, processes or states which go on in the mind of the subject is a bad picture, then what is the picture with which Wittgenstein thinks it should be replaced? Consider the following passage:

> What happens is not that this symbol cannot be further interpreted, but: I do no interpreting. I imagine N. No interpretation accompanies this image; what gives the image its interpretation is the path on which it lies. (*Philosophical Grammar,* hereafter PG, 23)

Here Wittgenstein is talking specifically about the individuation of a mental image. But his point generalises to understanding, meaning, intending, and so on. What makes it the case that I

[23] Another line of thought leading to this conclusion is the idea that understanding, meaning, intending, and so on lack what Wittgenstein calls genuine duration. Mental states run a course (*Zettel* §488), and one can observe, uninterruptedly, the course's progression (*Zettel* §§76-77). We can observe with continuity the onset of a pain, its intensification and its dissipation. This is not the case with understanding, meaning and intending. We may certainly be able to test for understanding and our tests may reveal changes with respect to it. But we cannot observe the course of understanding as we may, say, the course of a subject's anxiety (*Zettel* §§77, 82).

understand that *p* (as opposed to that *q* or that *r*) is, Wittgenstein thinks, the path on which my understanding lies. But what might it mean to talk about the 'path on which one's understanding lies'? Wittgenstein's idea is that what makes it the case that I understand the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight are not facts about my non-intentionally described intrinsic state at the time—facts about what came before my mind or about my verbal or non-verbal behaviour, non-intentionally described. On Wittgenstein's view, facts about the broader context may also play an individuating role:

> … we refer by the phrase "understanding a word" not necessarily to that which happens while we are saying or hearing it, but to the whole environment of the event of saying it. (*The Blue and Brown Books*, hereafter BB, 157).[24]

We have already noted one way in which Wittgenstein thinks that the content of my mental states depends on the broader context. In order to have thoughts involving the concept chess, Wittgenstein thinks, there must be an extant practice of playing chess (or relevantly similar games). To put the point in a way that Wittgenstein himself never does, the existence of a practice of playing chess (or relevantly similar games) is a necessary condition for my having thoughts about chess. But clearly it cannot be a sufficient condition, for I can have thoughts about things other than chess even if there is an extant practice of playing chess.

If the existence of a practice of playing chess is not sufficient for my having a thought about chess, then what else is required, on Wittgenstein's view? On Wittgenstein's view, questions like 'What makes it the case that *S* understands that *p*?' ought to be answered by calling to mind the *criteria* for *S*'s understanding that *p*, that is, those things which warrant someone else's saying of *S* that she understands that *p*. This is because on Wittgenstein's view, there is a constitutive relation between the sorts of things which count as criteria for *S*'s understanding that *p*

[24] '… "Understanding" is not the name of a single process accompanying reading or hearing, but of more or less interrelated processes against a background, or in a context, of facts of a particular kind, viz. the actual use of learnt language or languages'. (PG 74)

and *S*'s understanding that *p* (PI §353). So what *are* the criteria for *S*'s understanding the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight (as opposed to Smith *B*)?

Wittgenstein refrains from answering questions of this kind by giving a systematic account of the sorts of contextual features which constitute criteria for the relevant state. The main reason is that he does not think that such an account can be given. What counts as a criterion for *S*'s understanding that *p*, meaning that *q*, intending to φ, is, Wittgenstein thinks, going to vary from case to case. But we can say something about the sorts of contextual features which may, in any particular case, constitute criteria for *S*'s understanding the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight. These may include facts about the immediate context (for instance, the fact that we were having a conversation about Smith *A*) or facts about an earlier context (for example the fact that earlier we had been talking about whether Smith *A* would come tonight). They will almost certainly include facts about *S* herself, facts about her abilities, for instance (for example, the fact that *S* can speak English, the fact that she is able to identify Smith *A*, and so on). They may include facts about the things that *S* said or did at the time (for instance, the fact that *S* said that she looked forward to seeing Smith *A*) or facts about what *S* had said or done earlier (for instance, the fact that she had asked whether Smith *A* was coming). Wittgenstein is clear that they may also include facts about *S*'s *dispositions*, facts about what *S would* have said and done had the circumstances been different:

> "What makes this sentence a sentence that has to do with *him*"? "The fact that we were speaking about him."—"And what makes our conversation a conversation about *him*?"—Certain transitions we made or *would make*'. (*Last Writings on the Philosophy of Psychology* ,Vol. 1 §308. My emphasis in final line. See also PI §§684, 187)

What sorts of facts about *S*'s dispositions does Wittgenstein think may constitute criteria for *S*'s understanding the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight? As Budd makes clear:

> A special significance is assigned [by Wittgenstein] to how someone would express his psychological state in words: a criterion for which person my picture, image, sentence, or thought is of, or for what I intended to say, or for what I meant by what I said, is the answer I would have given to a question about my meaning. (Budd, 1984: 140)

On Wittgenstein's view, the fact that *S* is disposed to respond to questioning about the content of her understanding by saying that she understands the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight is a criterion for her having understood those words in that way.

In Subsection 1.1 I pointed out that the relations which externalists think can play an individuating role with respect to a subject's mental states include relations to a context which obtained at some prior time. Wittgenstein's view exemplifies this point. On Wittgenstein's view, facts about what went on before *S* heard the words 'Smith is coming tonight'—for example, the fact that earlier we had been talking about whether Smith *A* would be coming tonight—can play an individuating role with respect to the content of *S*'s occurrent understanding.

I said that on Wittgenstein's view, facts about *S*'s dispositions may constitute criteria for *S*'s understanding the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight. One might wonder whether this component of Wittgenstein's view is distinctly externalist. One might think that *S*'s dispositions to respond in particular ways to questions about the content of her understanding are intrinsic properties of *S*. But if they are intrinsic properties of *S*, then there does not appear to be anything distinctly externalist about the thought that *S*'s understanding is individuated in part by the fact that she is disposed to respond to questioning about the content of her understanding by saying that she understands the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight. If the relevant disposition is an intrinsic property of *S*, then this is a thought with which the internalist can readily agree.

In fact, the relevant disposition is not an intrinsic property of *S*, on Wittgenstein's view. On that view, the facts about the broader context which play a role in individuating the content of a subject's mental states play that role under an intentional characterisation. For example, what makes it the case that *S* understands the words 'Smith is coming tonight' to mean that Smith *A* is coming tonight is the fact that we were *talking about* Smith *A*, the fact that she can *identify* Smith *A* and so on. The same is true, on Wittgenstein's view, of facts about how *S* is disposed to respond to questioning about the content of her understanding. It is the fact that *S* is disposed to respond to

questioning by *saying* that she understood the words to mean that Smith *A* is coming tonight—and not the fact that she is disposed to respond by speaking the sentence 'I understood the words to mean that Smith *A* is coming tonight'—that has a role in individuating her understanding, according to Wittgenstein. Crucially though, whether or not *S*'s disposition is correctly intentionally characterised as a disposition to say that she understood the words to mean that Smith *A* is coming tonight will itself depend on facts about the broader context, on Wittgenstein's view (on the fact that we were having a conversation about Smith *A*, on the fact that *S* is able to identify Smith *A*, and so on). To the extent that the relevant disposition itself depends on facts about the broader context, it is not an intrinsic property of *S*.

Now that we have an overview of Wittgenstein's views about the individuation of mental content, we can see that there is no in principle reason why he does not give Twin Earth-style thought experiments which are addressed primarily to mental content or concepts instead of mental attitudes and activities. Wittgenstein thinks that what makes it the case that *S* understands that *p* (as opposed to that *q* or that *r*) are not facts about what the subject is doing or what is going on within the subject at the time (where these things are described in terms which do not take it for granted that *S* understands that *p*); they include facts about the broader context. Two individuals who have the same things come before their mind and who engage in the same behaviour (where these are relevantly described), but who happen to be situated in contexts which differ in relevant respects will, Wittgenstein thinks, differ with respect to the content of their mental states.

2. THE DRIVING INTUITION ABOUT SELF-KNOWLEDGE

2.1 Introducing the Intuition

The focus of the discussion thus far has been on what I am calling externalism's driving intuition and on thought experiments offered by both Burge and Wittgenstein which prima facie

support that intuition. In this subsection the focus of the discussion will shift, from externalism's driving intuition to what I shall call *the driving intuition about self-knowledge*.

At least since Descartes, the consensus view among philosophers has been that the knowledge we have about our own mental states is importantly different from the knowledge we have of the mental states of other people. One relatively uncontroversial way of characterising this difference is in terms of a difference between the way in which we know our own mental states and the way in which we know the mental states of other people. Judgments about the mental states of another person are, on most accounts, inferences which we make on the basis of observational evidence, namely, the things which that person says or does. Our warrant for such judgments depends on their being based on such evidence. In comparison, our warrants for judgments about our own mental states (hereafter *first-person judgments*) do not, generally speaking, depend on evidence in this way. In order for my self-ascriptions to be warranted, I generally do not need to observe my own behaviour, the things that I say and do. Nor, it seems, do I need to consult any sort of *inner* evidence.

We can summarise this difference between the way in which we know our own mental states and the way in which we know the mental states of other people in the following way. We know what other people are thinking on the basis of grounds, the things that they say and do. In comparison, we normally know our own mental states groundlessly. Hereafter, I shall refer to this intuition about first-person judgments—the thought that we normally know groundlessly what we are thinking—as *the driving intuition about self-knowledge*.

In doing so, I do not mean to suggest that this thought is the only guiding intuition that we have about self-knowledge. Nor do I mean to be claiming that it is the most fundamental. Intuitively, first-person judgments are generally speaking uniquely authoritative. If *S* says 'I hope it will stop raining soon' then in most cases, this gives us a stronger reason to believe that *S* really does hope that it will stop raining soon than if someone else were to say '*S* hopes it will stop raining soon'. It is not clear, however, that the unique authoritativeness of first-person judgments

is to be explained directly in terms of their groundlessness. Rather, these two thoughts about the nature of first-person judgments seem to constitute distinct intuitions.

Why, then, identify the apparent groundlessness of first-person judgments as the driving intuition about self-knowledge? Why not their unique authoritativeness? I am interested in whether there is a tension between externalism's driving intuition and our guiding intuitions about self-knowledge. If we want to understand why some people have thought that there *is* such a tension, then we do better to focus on the apparent groundlessness of first-person judgments rather than their unique authoritativeness. In Chapter 2, Section 1 I will consider a class of objections which purport to show that if externalism's driving intuition is true, then in order to know what we are thinking, we must first consult our context. If the objections succeed in showing this, then arguably they succeed in showing that if externalism's driving intuition is true, then our first-person judgments cannot normally be uniquely authoritative (for it is difficult to see how one could both endorse a particular theory about the individuation of mental content and claim that our first-person judgments carry any special authority, if it follows from the theory in question that in order to know what we are thinking, we must first investigate our context). But they do so only insofar as they succeed in showing that if externalism's driving intuition is true, then our first-person judgments must normally be grounded in empirical evidence. In other words, it is the relation between externalism's driving intuition and an intuition about the groundlessness of our first-person judgments which is really at issue in the objections.

Occasionally, I will make reference to *our intuitive picture of self-knowledge*. I understand this picture to include the two intuitions already mentioned, namely, the intuition that normally we know our mental states groundlessly and the intuition that our first-person judgments are, generally speaking, uniquely authoritative. But I understand it to include other intuitions about self-knowledge besides. For example, I take it to be part of our intuitive picture of self-knowledge that changes in one's context cannot by themselves bring about a change in the truth-values of one's past-tensed first-person judgments, judgments about one's remembered thoughts, for example. There is, however, an important difference between the status of the driving intuition about self-

knowledge and the status of those various other intuitions which belong to what I am calling our intuitive picture of self-knowledge. I take the thought that we normally know our mental states groundlessly to constitute a datum, information about self-knowledge against which candidate theories about the individuation of mental content ought to be assessed. If it should turn out that externalism is in tension with the claim that normally we know groundlessly what we are thinking, then that is automatically a reason to reject or revise externalism's driving intuition.

The various other intuitions which belong to our intuitive picture of self-knowledge do not have this status. A tension between externalism and one or more of these intuitions is not automatically a reason to reject or revise externalism. It may be a reason to reassess those intuitions. In recent history at least, internalist intuitions about mental content have enjoyed default standing in philosophical and non-philosophical thinking about the mind. Our intuitive picture of self-knowledge has informed, and in turn has been informed by, these intuitions. So it should come as no surprise if externalism should turn out to be in tension with *some* components of that picture. Indeed, it would be surprising if it did not. When such tensions do arise, we should take seriously the possibility that we have a reason to reject or revise the components of the intuitive picture, and not the relevant externalist theses about the individuation of mental content.

2.2 Accounting for the Groundlessness of Self-Knowledge

According to the driving intuition about self-knowledge, we normally know groundlessly what we are thinking. Generally speaking, a judgment's being groundless is a consideration which counts against its constituting knowledge. To constitute knowledge, a particular judgment must be warranted and groundless judgments are, generally speaking, unwarranted. So we might well ask why first-person judgments should be any different. What is it about first-person judgments which explains how they *could* normally be warranted and yet groundless?

Different accounts of self-knowledge answer this question in different ways.[25] We can classify such accounts on the basis of two considerations. The first consideration is whether the account is *epistemic* or *non-epistemic*. The second consideration is whether the account is *substantive* or *non-substantive*. Let us say that an account of self-knowledge is epistemic if it maintains that, in the normal case at least, one's warrant for first-person judgments has its basis in some epistemically privileged way of knowing. The proponent of a non-epistemic account denies this claim. On her view, it is some non-epistemic feature of first-person judgments which serves as the basis for our warrant for them. An account is substantive if it holds that first-person judgments about intentional states involve a cognitive achievement, where the judgment that one believes that *p* involves a cognitive achievement if and only if it involves the detection of a state of affairs which exists at least partly independently of one's judging (under ideal conditions) that one believes that *p*. If self-knowledge does not involve a cognitive achievement, then one's judgment (under ideal conditions) that one believes that *p* is wholly constitutive of one's believing that *p*.[26]

It is difficult to see how an account of self-knowledge could be both epistemic and non-substantive. An account is epistemic if it maintains that one's warrant for first-person judgments has its basis in an epistemically privileged way of knowing and an epistemically privileged way of knowing just is an epistemically privileged way of *detecting* a state of affairs. But non-substantive accounts reject the claim that knowing what one is thinking involves the detection of a state of affairs (for they deny that there is some state of affairs—one's believing that *p*—which exists independently of one's judging that it does). So it seems that there cannot be an account which is both epistemic and non-substantive. In effect, then, there are three possible categories into which a particular account of self-knowledge might fit: first, an account might be epistemic and

[25] I focus only on accounts of how we know our occurrent *intentional* mental states. Accounts of how we know, for example, our sensations I leave to one side.
[26] This conception of what it is for a first-person judgment to involve a cognitive achievement is informed by the discussion in Moran (2001: 17-19).

substantive; second, it might be non-epistemic and non-substantive; and third, it might be non-epistemic and substantive.[27]

Direct-observation accounts are examples of accounts which fall into the first category. According to such accounts, we each of us enjoy special epistemic access to the objects of our inner world. This access, which is conceived on the model of ordinarily perceptual observation, is special insofar as it is *immediate* in both a metaphysical and an epistemic respect. Unlike ordinary perceptual observation, nothing mediates between the object of my inner 'observation' and the state of my 'observing' that object. Moreover, judgments formed on the basis of this inner 'observation' are, unlike ordinary perceptual judgments, non-inferential. My warrant for first-person judgments is to be explained in terms of their being the results of this sort of special access. To this extent, accounts of this sort are epistemic. Typically, the objects which I have access to are held to be there whether or not I judge that they are. To this extent, accounts of this sort are substantive.

An example of an account which falls into the second category is the deflationary account which Crispin Wright attributes to Wittgenstein. According to Wright, Wittgenstein rejects outright the idea that self-knowledge is a matter of 'being in cognitive touch with' states of affairs which confer truth on our first-person judgments (Wright, 2001: 312). Rather, on Wright's interpretation it is Wittgenstein's view that:

> ... the authority standardly granted to a subject's own beliefs, or expressed avowals, about his intentional states is a *constitutive principle*: something that is not a by-product of the nature of those states, and an associated epistemologically privileged relation in which the subject stands to them, but enters primitively into the conditions of identification of what a subject believes, hopes, and intends. (Wright, 2001: 312)

On Wright's interpretation, it belongs to the grammar of intentional states that my sincere avowal that I am thinking that *p* has default standing; in the absence of any reason to reject it, my

[27] In setting up this classificatory scheme, I am of course assuming that there are some epistemic accounts which are consistent with thinking of self-knowledge as normally groundless. As I understand it, the epistemic/non-epistemic distinction is not extensionally equivalent to the non-groundless/groundless distinction.

avowal is to be counted as correct. On this view, the relation between one's first-person judgments and one's intentional mental states is not epistemic but *constitutive*. In the absence of evidence to the contrary, my sincere avowal (or the fact that I am disposed to sincerely avow) that I am thinking that *p makes it the case* that I am in fact thinking that *p*. Accordingly, the authority of first-person judgments is accounted for in terms of their constitutive role, and not in terms of 'an associated epistemologically privileged relation in which the subject stands to [those states]'. To this extent, the account is clearly both non-epistemic and non-substantive (I shall have more to say about Wright's interpretation of Wittgenstein in Chapter 5, Subsection 1.2).

Richard Moran's commitment account, set out *in Authority and Estrangement* and elsewhere, is an example of an account which falls into the third category. On Moran's view, purely epistemic accounts of self-knowledge too readily ignore its evaluative aspect, the fact that first-person judgments are often expressive of one's commitment to the attitudes one is ascribing to oneself. We do not just *report* our attitudes, as we might the attitudes of another person; we *avow* them. According to Moran, it is in terms of this feature of first-person judgments—the fact that they are expressive of our commitments—and not purely in terms of any epistemic feature, that our warrant for those judgments ought to be explained.[28] To this extent, Moran's account is non-epistemic. But it is also substantive. Moran readily acknowledges that first-person judgments can play a constitutive role of *some* sort with respect to one's mental states. For example, a conception of one's pride as sinful, 'suffices for [one's] pride to be of an essentially different nature from someone else's pride, or from his own pride before he came to see it that way' (Moran, 2001: 49). However, it is not part of Moran's view that my judging that I, say, feel pride is necessary for it to be true that I am in fact feeling pride. Moran accepts that self-knowledge can involve awareness of a state of affairs which exists independently of one's judging that it does.

[28] 'We may allow any manner of inner events of consciousness, any exclusivity and privacy, any degree or privilege and special reliability, and their combination would not add up to the ordinary capacity for self-knowledge. For the connection with the *avowal* of one's attitudes would not be established by the addition of any degree of such epistemic ingredients'. (Moran, 1997: 155)

CONCLUSION

Two aims have guided the discussion in this chapter. The first aim was to introduce externalism's driving intuition—the thought that at least some types of mental states are at least partly individuated by the relations which a subject bears to a context—and to discuss some of the thought experiments offered by both Burge and Wittgenstein in support of this intuition. In Subsection 1.2, I considered three thought experiments offered by Burge in support of externalism's driving intuition, namely, the arthritis, water and sofa cases. I discussed the principles on which these thought experiments respectively rely and drew attention to the way in which they broaden the scope of Burge's externalism. In Subsection 1.3, I considered two thought experiments drawn from Wittgenstein's later work which are clearly externalist in spirit, but which are addressed primarily to mental attitudes and activities rather than mental content. I discussed some of the salient points of similarity and difference between these thought experiments and the thought experiments outlined in Subsection 1.2. I went on to summarise Wittgenstein's positive view about the individuation of mental states. This summary showed that there is no principled reason why Wittgenstein does not put forward arguments structurally identical to the arguments he puts forward in the passages I cited, but which are addressed primarily to mental content instead of mental attitudes and activities.

The second guiding aim was to introduce the driving intuition about self-knowledge—the thought that we normally know groundlessly what we are thinking—I drew a distinction between what I am calling the driving intuition about self-knowledge and various other guiding intuitions which belong to what I called our intuitive picture of self-knowledge. I traced the distinction to a difference in the *status* of the respective intuitions. The thought that we normally know groundlessly what we are thinking has the status of a datum. If it should turn out that externalism is in tension with this thought, then that is automatically a reason to reject or revise externalism's driving intuition. The various other intuitions which belong to our intuitive picture of self-knowledge do not have this status. A tension between externalism and one or more of these

intuitions is not automatically a reason to reject or revise externalism's driving intuition. It may be a reason to reassess the intuitions which belong to our intuitive picture of self-knowledge. I went on to consider several different ways in which philosophers have sought to account for groundless self-knowledge.

Chapter 2

A TENSION BETWEEN THE TWO INTUITIONS

INTRODUCTION

The discussion in Chapter 1 centred on two intuitions, which I called externalism's driving intuition and the driving intuition about self-knowledge, respectively. My aims in Chapter 2 are essentially twofold: first, to introduce a familiar line of argument for the claim that there is an inconsistency between these two intuitions; and second, to consider Burge's response to this line of argument.

The chapter consists of three sections. In Section 1, I introduce slow-switching objections. Typically, these objections purport to show that if externalism's driving intuition is true, then

switching a subject slowly between two relevantly dissimilar contexts, in such a way that the subject cannot tell that she is being slowly switched, can bring it about that the subject cannot know groundlessly certain of her mental states.[1] I defend the view that slow-switching objections have initial plausibility as objections to the arthritis and water cases, but not as objections to the sofa case. In Section 2, I consider Burge's response to slow-switching objections. That response has two components. The first component involves defending a claim about the *accuracy* in slow-switching cases of one's judgments about one's mental states. The claim is that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. The second component of Burge's response involves defending a claim about the source of one's warrant for judgments about one's mental states. The claim is that this warrant takes the form of an entitlement and derives from one's identity as a critical reasoner. It is not grounded in one's ability to identify one's thoughts or discriminate them from relevant alternatives. In Section 3, I consider the second component of Burge's response in more detail. I ask what it means to say that one's warrant for judgments about one's mental states takes the form of an entitlement which *derives* from one's identity as a critical reasoner.

1. SLOW-SWITCHING OBJECTIONS

Why think that if externalism's driving intuition is true, then we cannot normally know our mental states groundlessly? Here is a first pass at an answer. If the thought that *p* is partly individuated by relational fact *R*, then *R*'s obtaining is a necessary condition for my thinking that *p*. If *R*'s obtaining is a necessary condition for my thinking that *p*, then I know that I am thinking that *p* only if I have confirmed that *R* does in fact obtain. But if I must confirm that *R* does in fact obtain before I can know that I am thinking that *p*, then I cannot know groundlessly that I am thinking that *p*. If I judge that I am thinking that *p* without having first confirmed that *R* does in

[1] A more careful formulation of what it is that the relevant class of objections purport to show is offered in fn. 6.

fact obtain, then my judgment that I am thinking that *p* is not warranted and, consequently, I do not know that I am thinking that *p*. Call this *the swift argument for inconsistency*.

The problem with the swift argument for inconsistency is that it relies on what is surely a problematic assumption, namely, that one knows that one is thinking that *p* only if one has confirmed that all the various enabling conditions for one's thinking that *p* obtain. As Burge himself is keen to point out, this assumption is too strong, as we can see when we consider our warrant to perceptual judgments:

> Our epistemic right to our perceptual judgments does not rest on some prior justified belief that certain enabling conditions are satisfied. In saying that a person knows, by looking, that there is food there, we are not required to assume that the person knows the causal conditions that make his perception possible. We certainly do not, in general, require that the person has first checked that the light coming from the food is not bent through mirrors, or that there is no counterfeit food in the vicinity. We also do not require that the person be able to recognise the difference between food and every imaginable counterfeit that could have been substituted. (Burge, 1988: 654-655)

Burge's point is certainly true 'in general'. But clearly there are circumstances in which we *would* require of someone that they conduct the kind of empirical investigations Burge has in mind before we describe them as knowing that there is food there. If the person is visiting a wax museum or a hall of mirrors, for example, then they must rule out the possibility that what they are seeing are actually carefully crafted pieces of wax or a reflection. Although we do not require someone to have ruled out *every* alternative to their actually seeing *x* before we describe them as knowing that they see *x*, we do require that they have ruled out *relevant* alternatives.[2] One's seeing *y* (thinking that *q*) is a relevant alternative to one's seeing *x* (thinking that *p*), let us say, just in case seeing *y* is incompatible with seeing *x* and seeing *y* is relevant given the context.[3] We should not expect a systematic account of what it is for a particular alternative to be relevant given the context. Generally speaking, however, an alternative can be relevant for one even if one is not in a

[2] See Paul Boghossian (1998: 158) on this point. The notion that knowledge requires a capacity to distinguish between the actual case and relevant alternatives can be traced back to Alvin Goldman (1976).
[3] This characterisation belongs to Falvey and Owens (1994: 116). For discussion of the sorts of contextual considerations which render an alternative relevant, see David Lewis (1996).

position to recognise that it is.[4] If I am visiting a wax museum but do not know that I am, then nevertheless, I must rule out the possibility that what I am seeing are carefully crafted pieces of wax before I can know that there is food there.

These considerations yield a more sophisticated argument for the claim that if externalism's driving intuition is true, then we cannot normally know our mental states groundlessly. Consider the following *slow-switching* case.[5] Suppose that Alf is being switched slowly back and forth between Earth and Twin Earth completely unawares. Suppose further that Alf spends a sufficient period of time in each location that he comes to acquire the concept appropriate to that location. Immediately prior to his being switched back to Earth from Twin Earth, Alf uses the term 'arthritis' to express the concept tharthritis. Immediately prior to being switched back to Twin Earth, he uses the term to express the concept arthritis, and so on. Now suppose that on one occasion immediately prior to his being switched from Twin Earth back to Earth, Alf reports to his doctor 'My arthritis has worsened'. Because he has been most recently passing time on Twin Earth, Alf uses the utterance to express the thought that his *tharthritis* has worsened. But can Alf know groundlessly that this is what he is thinking? If Alf needs to exclude relevant alternatives to his thinking that his tharthritis has worsened before he can know that this is what he is thinking, then it seems that he cannot. Because he is being slowly switched between Twin Earth and Earth, the possibility that he is actually thinking that his *arthritis* has worsened is now a relevant alternative. Consequently, he must rule out that possibility before he can know that he is thinking that his tharthritis has worsened. However, the only way he can do that is by determining through empirical investigation whether he is on Earth or Twin Earth. Thus, it seems that Alf cannot know groundlessly that he is thinking that his tharthritis has worsened.

[4] This is not to say that whether or not an alternative is relevant is entirely insensitive to what one (reasonably) believes about one's situation. If I wrongly (but reasonably) believe that I am visiting a wax museum, then I must rule out the possibility that what I am seeing are carefully crafted pieces of wax before I can know that there is food there.

[5] This slow-switching case is modelled on the case set out by Boghossian (1998: 158-160), which is itself modelled on the case set out by Burge (1988).

Hereafter, I shall refer to arguments for a tension between externalism's driving intuition and the driving intuition about self-knowledge which make use of slow-switching scenarios in this way as *slow-switching objections*.[6]

In setting out this objection, I assumed that if Alf were being slowly switched between Earth and Twin Earth, then he would use the term 'arthritis' to express the concept tharthritis on Twin Earth and the concept arthritis after having spent time back on Earth. Would Burge accept this assumption? When Burge discusses slow-switching objections, it is usually in connection with arguments for externalism which exploit a difference between the natural (and not, or not purely, the social) contexts on Earth and Twin Earth.[7] So it is not immediately clear what he would say about the way in which Alf's concepts might change if he were being slowly switched. We can, however, draw some conclusions about what Burge would say about such a case by considering what he does in fact say about slow-switching in connection with arguments which exploit a difference between the natural (and not, or not purely, the social) contexts on Earth and Twin Earth.

Suppose that a subject, *S*, is slowly switched between Earth and Twin Earth completely unawares. Suppose that on Earth, 'aluminium' means aluminium while on Twin Earth 'aluminium' means *twaluminium*, where twaluminium is a metal which has a different substructure to aluminium, but is otherwise indistinguishable (suppose that there is no aluminium on Twin

[6] Proponents sometimes present slow-switching objections as demonstrating an *inconsistency* between externalism's driving intuition and the driving intuition about self-knowledge, that is, as demonstrating that if externalism's driving intuition is true, then the driving intuition about self-knowledge *must* be false (see, for instance, Ludlow, 1995). But even if sound, slow-switching objections would not demonstrate an inconsistency between the two intuitions. To demonstrate an inconsistency, it would need to be shown that in *every* world in which externalism's driving intuition is true, the driving intuition about self-knowledge is false. But as Warfield (1997: 283-4) makes clear, slow-switching objections are not of the right form to show this. In order for slow-switching objections to show that in every world in which externalism's driving intuition is true, the driving intuition about self-knowledge is false, it would need to be true that there are instances of slow-switching in every world in which externalism's driving intuition is true. But presumably there are some worlds in which externalism's driving intuition is true in which there are no cases of slow-switching. Given these considerations, a more careful formulation of what it is that slow-switching objections purport to show is as follows: there are some possible worlds, including the actual world, in which we cannot normally know our mental states groundlessly, if the driving intuition about externalism is true. For ease of exposition, however, I will continue to describe slow-switching objections as purporting to demonstrate an inconsistency between externalism's driving intuition and the driving intuition about self-knowledge. No issues important to the discussion are affected by my doing so.

[7] See, for example, Burge (1988: 652-653, 1996: 95-96).

Earth and no twaluminium on Earth). Imagine that *S* is not an expert, that she knows nothing about the microstructure of metals. How, if at all, will the switch affect the concept which *S* uses the term 'aluminium' to express? The answer to this question, Burge thinks, depends on further particulars of the case. Burge distinguishes between *disjoint* and *amalgam type* slow-switching cases. In disjoint type cases, a subject who is being slowly switched comes to use the relevant term 'on different occasions [to express] two concepts whose extensions are disjoint' (Burge, 2013d: 89). In amalgam type cases, the subject's concept broadens, so that it applies, for example, to *both* aluminium and twaluminium. According to Burge:

> Whether in slow switching we have an Amalgam Type case depends partly, probably mainly, on how the individual is committed to the standards of the communities in the two environments. The relevant individual lacks knowledge of the metals' substructures. In the absence of commitments to communal standards and to communal understanding, and given extensive, prolonged new experiences with twaluminium, the relevant individual will lack the cognitive resources for the referent of the word-form "aluminium" to be fixed as a single natural kind. In the absence of other constraints, only normal experience with metals that are in fact instances of a single natural kind would so fix the referent. In the face of normal, constant perceptual application of the term to different natural kinds (where there is no countervailing pressure having, for example, to do with differential values or purposes to which the metals might be put), the extension will commonly broaden. (Burge, 2013d: 89-90)

If *S* has 'extensive, prolonged new experiences' with samples of twaluminium during her stay on Twin Earth, but she does not come to depend on the communal standards there, then it is likely that the concept she uses 'aluminium' to express will broaden. In comparison, if *S* does come to depend on these standards, then she will come to use the term 'aluminium' to express a concept the extension of which in no way overlaps with the extension of the concept <u>aluminium</u>.

We already have a sense of the way in which dependence on communal standards and extensive and prolonged experiences with samples of a particular kind can play individuating roles with respect to a subject's concepts and hence mental content. In Chapter 1, Subsection 1.2 I noted that the arthritis and water cases rely on *Deference*, the claim that if the concept <u>*C*</u> is standardly associated with the term '*C*' by the experts, then a subject who either misunderstands <u>*C*</u> or lacks expert knowledge about *C*'s can, by deferring to those experts, use '*C*' to express <u>*C*</u>. A subject who defers to the experts in her community depends to some greater or lesser extent on her

community's standards. So *Deference* implies that dependence on communal standards can play an individuating role with respect to a subject's mental content. In that same subsection, I noted that the sofa case relies on *Direct Causal Contact*, the claim that the direct causal relations at work in perception and in perceptually-backed demonstrative applications of an empirically applicable term, '*C*', which connect a subject with actual *C*'s, can bring it about that she uses '*C*' to express the concept *C*. I take it that a subject who makes regular perceptually-backed, demonstrative applications of the term '*C*' to *C*'s is a subject who has extensive and prolonged experiences with *C*'s. So *Direct Causal Contact* implies that extensive and prolonged experiences with a particular kind can play an individuating role with respect to a subject's mental states.

Neither *Deference* nor *Direct Causal Contact*, however, tell us why Burge thinks that in cases where slow-switching brings it about that a subject has extensive and prolonged new experiences with samples of a particular kind, whether she acquires a second, distinct concept will depend on whether she comes to depend on Twin Earth communal standards. *Deference* and *Direct Causal Contact* claim, respectively, that deferential and direct causal relations *can* bring it about that a subject has thoughts involving the relevant concept. Neither tells us anything about the conditions under which these relations *will* bring it about that the subject has thoughts involving that concept. So why does Burge think that in slow-switching cases of the relevant sort, whether the subject acquires a second, distinct concept will depend on whether she comes to depend on Twin Earth communal standards?

The slow-switching cases Burge has primarily in mind in the passage above are cases in which the subject lacks knowledge which could enable her to discriminate the samples with which she has extensive and prolonged new experiences on Twin Earth from the samples she has encountered on Earth.[8] *S*, for example, lacks knowledge which could enable her to discriminate samples of twaluminium from samples of aluminium; she is ignorant about the microstructure of

[8] I say '*could* enable' because the subject may not be in a position to discriminate the samples, even though she possesses knowledge of the appropriate kind. She may not, for example, know how to carry out the relevant tests.

metals.[9] To the extent that slow-switching brings it about that *S* has extensive and prolonged new experiences with samples of twaluminium, Burge thinks, *S* will come to use the term 'aluminium' to express a concept which includes twaluminium in its extension. But because *S* lacks the 'cognitive resources' to discriminate samples of twaluminium from samples of aluminium, her concept, Burge thinks, will also include aluminium in its extension. Burge thinks that because *S* lacks knowledge which could enable her to discriminate the samples she encounters on Twin Earth from samples she has encountered on Earth, slow-switching will result in a broadening of the concept which *S* uses the term 'aluminium' to express, rather than in her coming to acquire a second, distinct concept.

Suppose, in contrast, that *S* comes to depend on Twin Earth communal standards. Burge's thought is that in this case, *S* can rely on the cognitive resources of the experts on Twin Earth, specifically, on their ability to discriminate samples of twaluminium from samples of aluminium. To the extent that she does come to rely on these resources, Burge thinks, *S* will come to use the term 'aluminium' to express a concept distinct from the concept aluminium.

How does this discussion help us settle the question I raised several paragraphs back, namely, 'Would Burge agree that if Alf were being slowly switched between Earth and Twin Earth, he would come to use the term 'arthritis' to express the concept tharthritis on Twin Earth and the concept arthritis on Earth'? Given the nature of his misunderstanding—given that he believes that arthritis can afflict one's thigh—Alf lacks knowledge which could enable him to discriminate instances of arthritis from instances of tharthritis. So the arthritis case is the sort of case Burge has primarily in mind in the passage cited above. In Chapter 1, Subsection 1.2 I noted that the arthritis case relies on the assumption that Alf uses the term 'arthritis' with a deferential intention. If we accept Burge's views about the considerations that determine whether slow-switching cases involving subjects who lack discriminatory knowledge are amalgam or disjoint type, then the fact that Alf uses the term 'arthritis' with a deferential intention is a reason for us to

[9] I consider what Burge might say about slow-switching cases involving subjects who do possess discriminatory knowledge further on in this section.

think that were he to be slowly switched, Alf's would be a disjoint type case. There is no reason to think that if he is slowly switched, Alf will stop using the term 'arthritis' with a deferential intention. However, his being slowly switched will probably bring about a change in the content of that intention. Given a sufficient period of time spent on Twin Earth, we can expect that the experts whose use of the term 'arthritis' Alf intends his own use to accord with will be the experts on Twin Earth.[10] But if Alf does come to use the term with this intention, then he will have come to be dependent on the communal standards of Twin Earth. So we should say that, were he to be slowly switched, Alf's would be a disjoint type case, that is, that he would come to use the term 'arthritis' on Twin Earth to express a concept which is distinct from the concept arthritis.[11]

We can expect Burge to agree that given enough time on Twin Earth, Alf will come to use the term 'arthritis' to express the concept tharthritis. But my earlier assumption was not just an assumption about the concept which Alf uses the term 'arthritis' to express on Twin Earth. It was an assumption about the concept which he uses the term to express during his stays back on Earth. I assumed that given enough time back on Earth, Alf will once more come to use the term 'arthritis' to express the concept arthritis. Would Burge agree with this assumption? I think so. Burge is explicit on two points: first, a subject who acquires new concepts because of a change in their context does not normally lose their original concepts; and second, time spent in one's original context can reactivate those concepts:

> Moving to the other environment and acquiring new concepts will not normally obliterate old concepts or memories that derive from the first environment … The old abilities will normally still be there; and there are situations, such as invocation of memory, or reasoning based on memory, or return to the first environment with acts of deference to its communal norms, that can bring these abilities into play. (Burge, 2013d: 92)

Given that he uses the term 'arthritis' with a deferential intention, we can expect that after having spent a sufficient period of time back on Earth, Alf will come to depend on the communal

[10] At least for the duration of his stay on Twin Earth. We can expect that his intention would change back given enough time spent back on Earth.
[11] Of course, in this case it is not true to say that Alf acquires a concept the extension of which in no way overlaps with the extension of the concept arthritis. It is a feature of the case that anything which falls under the extension of the Earth concept, falls under the extension of the Twin Earth concept.

standards of his original community. So we can expect Burge to agree that after returning to Earth, Alf will come to use the term 'arthritis' to express the concept arthritis.

In summary, then, the assumption which I made when setting out the slow-switching objection in connection with the arthritis case would appear justified. We can expect that Burge would agree that if Alf were being slowly switched, he would come to use the term 'arthritis' to express the concept tharthritis when on Twin Earth and the concept arthritis during his stays on Earth.

Our discussion of slow-switching thus far has focused on Burge's arthritis case. What effects might slow-switching have on the protagonists in Burge's water and sofa cases, Carl and Tom? Like Alf, Carl lacks relevant discriminatory knowledge; Carl is ignorant of the molecular structure of water. Moreover, he uses the relevant term (in Carl's case, the term 'water') with a deferential intention. As in Alf's case, we can expect that given enough time, the experts whose use of the term 'water' Carl defers will be the experts on Twin Earth. But if Carl does come to defer to the experts on Twin Earth, then he will have come to be dependent on the communal standards of Twin Earth. So if we accept Burge's views about the considerations that determine whether slow-switching cases involving subjects who lack discriminatory knowledge are amalgam or disjoint type, then the fact that Carl uses the term 'water' with a deferential intention is a reason for us to think that, were he to be slowly switched, his would be a disjoint type case.[12]

If this is correct, then the slow-switching objection which I outlined earlier in connection with Burge's arthritis case has initial plausibility as an objection to the water case. Suppose that immediately prior to his being switched back to Earth, Carl says 'I'd like a glass of water'. Because Carl has most recently been passing time on Twin Earth, he uses these words to express

[12] I have suggested that Burge has a definite view about the way in which slow-switching will affect Alf and Carl. I should point out, however, that Burge does not think that there will be determinate answers in every case to questions about how slow-switching will affect the subjects involved: 'Of course, there can arise difficult questions about whether one is still employing thoughts from the departed situation or taking over the thoughts appropriate to the new situation. I think that general principles govern such transitions, but such principles need not sharply settle all borderline cases' (Burge, 1988: 652, fn. 5).

the thought that he would like a glass of *twater*. But Carl is not it seems in a position to know groundlessly that this is what he is thinking. Before he can know that he is thinking that he would like a glass of twater, Carl must rule out the possibility that he is thinking that he would like a glass of water, for the fact that he is being slowly switched renders this possibility a relevant alternative. The only way he can do that, however, is by determining through empirical investigation whether he is on Earth or Twin Earth. Thus, it seems that Carl cannot know groundlessly that he is thinking that he would like a glass of twater.

How, if at all, might slow-switching affect Tom, the protagonist in Burge's sofa case? We know that Tom, unlike Alf and Carl, is an expert. Generally speaking, experts will possess relevant discriminatory knowledge. Generally speaking, a sofa expert will possess knowledge which could enable her to discriminate between sofas and safos. Insofar as he has a non-standard theory about sofas, however, Tom does *not* possess relevant discriminatory knowledge. Tom believes that sofas are not items of furniture meant primarily for sitting, but rather religious artefacts. To the extent that he believes this, Tom lacks knowledge which could enable him to discriminate sofas from safos, objects which are superficially indiscernible from sofas but are in fact religious artefacts.

Although he is an expert, Tom, like Alf and Carl, lacks relevant discriminatory knowledge. Unlike Alf and Carl, however, Tom does not use the relevant term (in Tom's case, the term 'sofa') with a deferential intention. Tom is a fully competent speaker who has come to question the meaning-giving characterisations associated with the term 'sofa'. Because Tom does not use the term 'sofa' with a deferential intention, slow-switching is unlikely to result in his coming to depend on Twin Earth communal standards. Slow-switching may, however, result in Tom's having extensive, direct causal contact with actual safos. Suppose that it does. Will such contact bring about a change in the concept which Tom uses the term 'sofa' to express?

We can expect Burge to answer 'Yes' to this question. Specifically, we can expect Burge to say that such contact will generate an amalgam type case. Insofar as he does not possess

relevant discriminatory knowledge, and insofar as he is unlikely to come to rely on Twin Earth communal standards, Tom lacks the cognitive resources for the referent of the word 'sofa' to be fixed as a single kind. Given that he lacks these resources, we can expect extensive, direct causal contact with actual sofas to result in a broadening of Tom's existing 'sofa' concept.

Do slow-switching objections have initial plausibility as objections to the sofa case, given that slow-switching will result in a broadening of Tom's existing concept? They do not. Slow-switching objections rely on there being regular change in the subject's concepts, that is, change each time the subject is switched between Twin Earth and Earth. It is only because there is regular change that alternative thoughts are rendered relevant. For example, what makes it the case that when on Twin Earth Alf says 'My arthritis has worsened', his expressing a thought about arthritis is a *relevant* alternative to his expressing a thought about tharthritis is the fact that: first, it could quite easily be the case that Alf is on Earth; and second, if he *were* on Earth he would be expressing a thought about arthritis. If subsequent switches between Twin Earth and Earth did not bring about some change in Alf's concepts—if after his initial switch, Alf used the term 'arthritis' to express the same concept on Earth as he used it to express on Twin Earth—then there would be no alternative thought rendered relevant by the fact that it could quite easily be the case that Alf is on Earth. But if there is no alternative thought which is relevant, then there is no alternative which Alf needs to rule out before he knows that he is thinking that his tharthritis has worsened.

If a slow-switching case involving Tom will be amalgam type, then on Twin Earth Tom uses the term 'sofa' to express a concept which applies to both sofas and safos (the concept <u>sofa</u>*, say). Burge explicitly leaves open the question of whether subjects in amalgam type cases retain their original concept (Burge, 2013d: 90). Suppose that Tom does retain his original concept, that in addition to the concept <u>sofa</u>*, Tom possesses the concept <u>sofa</u>. It is unlikely that during his stays back on Earth, Tom's original concept will be reactivated, as we might expect it to be in disjoint type cases. During his time back on Earth Tom may, of course, have extensive, direct causal contact with sofas. But if extensive, direct causal contact with safos on Twin Earth is insufficient to bring it about that Tom acquires a second, distinct 'sofa' concept, then there is no

reason to think that extensive, direct causal contact with sofas will reactivate Tom's original 'sofa' concept or cause his broadened concept to narrow. We can expect that having brought about an initial broadening of Tom's concept, further switching will bring about no further change, that Tom will use the term 'sofa' to express the same concept during his stays on Earth as he uses it to express during his stays on Twin Earth, namely, the concept sofa*.

If this is right, then slow-switching objections do not have initial plausibility as objections to the sofa case. In order for such objections to have initial plausibility, continued slow-switching must result in regular change in the subject's concepts. But if a slow-switching case involving Tom will be amalgam type, continued slow-switching is not likely to result in regular change in the concept which Tom uses the term 'sofa' to express.[13]

What might Burge say about slow-switching cases involving subjects who *do* possess relevant discriminatory knowledge and have extensive, direct causal contact with Twin Earth samples (but do not come to rely on Twin Earth communal standards)? It is not obvious what Burge might say about such cases, or indeed what he ought to say, given other aspects of his view. It is clear from the passage in which he discusses amalgam and disjoint type cases that Burge thinks that in slow-switching cases where the subject lacks relevant discriminatory knowledge, extensive, direct causal contact with Twin Earth samples will be sufficient to bring about a change in the subject's existing concept. But that passage—and the various others in which Burge discusses slow-switching scenarios—leave it unclear whether he thinks that extensive, direct causal contact with Twin Earth samples will be sufficient to bring about a change in the concept of a subject who *does* possess relevant discriminatory knowledge.

[13] There is, of course, a more general lesson here about the applicability of slow-switching objections to amalgam type cases. It seems that generally speaking, such objections will not have initial plausibility as objections to such cases. In amalgam type cases, the subject's circumstances are such that they do not acquire a second, distinct concept during their time on Twin Earth. But if a subject's circumstances are such that they do not acquire a second, distinct concept on Twin Earth, then they are not likely to be such that the subject's original concept is reactivated during their stays back on Earth (or such that their broadened concept narrows). If their original concept is not reactivated during their stays back on Earth (and their broadened concept does not narrow), then it is difficult to see how further switching could result in the sort of change in the subject's concepts which is necessary for slow-switching objections to have initial plausibility.

In the passage in which he discusses the sorts of circumstances which typically give rise to amalgam and disjoint type cases, Burge writes 'The relevant individual lacks knowledge of the metals' substructures. In the absence of commitments to communal standards and to communal understanding, and given extensive, prolonged new experiences with twaluminium, the relevant individual will lack the cognitive resources for the referent of his word-form "aluminium" to be fixed as a single natural kind'. Presumably, slow-switching cases in which the subject possesses relevant discriminatory knowledge are cases in which the subject does *not* lack the cognitive resources for the referent of the relevant word to be fixed as a single natural kind. So presumably, it is Burge's view that slow-switching cases involving subjects who possess relevant discriminatory knowledge will not be amalgam type. But it is unclear whether Burge thinks that in such cases, the kind picked out exclusively by the word-form will be the kind the subject has encountered on Earth or the kind she has encountered on Twin Earth. In other words, it is unclear whether on Burge's view, slow-switching cases where the subject possesses relevant discriminatory knowledge will be cases in which the subject's concept remains constant or whether he thinks that in such cases, the subject will acquire a second, distinct concept.

I am going to leave this question open. To this extent, I also leave open the question whether slow-switching objections have initial plausibility as objections to cases involving subjects who do possess relevant discriminatory knowledge (and have extensive, causal contact with Twin Earth samples but do not come to rely on Twin Earth communal standards). Clearly, if it is Burge's view that in such cases the subject's concept will remain constant, then slow-switching objections do not have initial plausibility as objections to such cases. If slow-switching will not affect a subject's concepts, then slow-switching objections simply cannot get going. Suppose, however, that it is Burge's considered view that extensive, direct causal contact with, say, samples of twaluminium on Twin Earth will be sufficient to bring it about that a subject who possesses relevant discriminatory knowledge acquires a second, distinct 'aluminium' concept. In this case, slow-switching objections *do* have initial plausibility as objections to cases of the relevant sort (for

the same sorts of reasons that they have initial plausibility as objections to slow-switching cases involving Alf or Carl).

We have been considering what Burge might say about slow-switching cases involving subjects who possess discriminatory knowledge but do not come to rely on Twin Earth communal standards. What might Burge say about the way in which slow-switching will affect the mental states of subjects who possess relevant discriminatory knowledge and *do* come to rely on Twin Earth communal standards? Such cases are surely possible. Suppose, for example, that Janice's knowledge of aluminium is limited and that she defers to the experts when she uses the term 'aluminium'. Suppose, however, that Janice possesses *some* knowledge about the microstructure of metals, knowledge which could enable her to discriminate samples of aluminium from samples of twaluminium. How, if at all, might slow-switching affect Janice's 'aluminium' concept, on Burge's view?

I think that Burge will say that slow-switching will bring it about that Janice acquires a second, distinct 'aluminium' concept. Janice uses the term 'aluminium' with the intention of expressing the concept which the experts standardly use the term to express. We can expect that given enough time on Twin Earth, the experts to whose use Janice defers will come to be the experts on Twin Earth. Given this consideration, and given Burge's views about the role which deferential relations can play in individuating a subject's mental content, we should expect that Janice will come to use the term 'aluminium' on Twin Earth to express the concept twaluminium. We should expect that Janice will come to use the term 'aluminium' to express a second, distinct concept as opposed to a broadened concept because, insofar as she comes to depend on Twin Earth communal standards, the cognitive resources of the experts will serve to fix the referent of the term 'aluminium' on Janice's lips as a single kind. If slow-switching will result in her acquiring a second, distinct 'aluminium' concept, then slow-switching objections have initial plausibility against cases involving subjects like Janice (for the same sorts of reasons that they have initial plausibility as objections to slow-switching cases involving Alf or Carl).

To summarise the main points of this section, I think that slow-switching objections have initial plausibility as objections to the arthritis and water cases. Insofar as a slow-switching case involving Tom is likely to be amalgam type, however, slow-switching objections do not have initial plausibility as objections to the sofa case.

2. BURGE'S RESPONSE TO SLOW-SWITCHING OBJECTIONS

In this section, I am going to consider Burge's response to slow-switching objections. That response has two components. The first component involves defending a claim about the *accuracy* in slow-switching cases of one's judgments about one's mental states. The claim is that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error.[14] This claim is not, however, sufficient to block slow-switching objections. The proponents of such objections can accept this claim and still insist that slow-switching undermines a subject's ability to *know* her mental states, for they can claim that in slow-switching scenarios, subjects lack the sort of discriminatory abilities which would *warrant* their judgments about their mental states. The second component of Burge's response involves defending a claim about the source of one's warrant for judgments about one's mental states. The claim is that this warrant takes the form of an entitlement and derives from one's identity as a critical reasoner. It is not grounded in one's ability to identify one's thoughts or discriminate them from relevant alternatives. I will discuss each of these components in turn.[15]

---

[14] Passages like the following are evidence that Burge wishes to defend this claim and not some weaker one (for example, the claim that slow-switching cannot bring it about that one's judgments about one's mental states are significantly more vulnerable to error): 'The self-ascription in the that-clause way cannot involve a mistake about the intentional content. So the possibility of a switch does not threaten a mistake. I think therefore that such possibilities pose no relevant alternative threat to one's entitlement to one's judgment about the that-clause content of one's thoughts. I believe that the relevant minimal understanding suffices for knowledge in *cogito*-like judgments. Even in non-*cogito*-like judgments, switches in content cannot, for the same reason, undermine knowledgeability of the content of self-ascriptions'(Burge, 1996: 97, fn. 2). It is true that slow-switching cannot undermine knowledgeability of the content of self-ascriptions only if it is true that slow-switching cannot bring it about that one's first-person judgments are subject to error.

[15] I do not have the space to discuss what is in my view a promising alternative response to slow-switching objections. This response involves defending the claim that alternatives which are relevant in slow-switching cases are not *normally* relevant. They are not normally relevant, one could argue, because we are

Burge's defence of the first component is developed over a series of articles. At the core of that defence is Burge's discussion, in 'Individualism and Self-Knowledge' and elsewhere, of a class of self-ascriptions which he refers to as *cogito-like*, judgments like 'I am, with this very thought, thinking that my arthritis has worsened'.[16] Cogito-like judgments are self-verifying; they cannot be false even if the subject who makes the judgment is being slowly switched. Of course, if the subject *is* being slowly switched, then the content of the first-order thought may well be different from what it would have been if she was not being slowly switched. But in cases where slow-switching brings it about that the content of the relevant first-order thought is different, that difference will necessarily be reflected at the second-order level. If the subject's first-order thought is about arthritis, then the second-order judgment necessarily ascribes a thought about arthritis. If the subject's first-order thought is about tharthritis, then the second-order judgment necessarily ascribes a thought about tharthritis, and so on. Because thinking the second-order judgment involves thinking the relevant first-order thought, the judgment cannot but be true.[17]

Of course, a great many of one's judgments about one's mental states are not *cogito-like* and consequently not self-verifying in this way. Nevertheless, some such judgments have something important in common with those which are cogito-like. Consider non cogito-like first-person judgments which ascribe occurrent mental states, judgments like 'I'd like a glass of water' or 'I hope it will stop raining soon'. Unlike cogito-like judgments, these judgments can of course be false; one may simply lack the first-order thought being ascribed. But they cannot be false because of a mismatch in content between the self-ascription and the relevant first-order thought. If one's context is changed around so that one actually desires *twater*, say, instead of water, then

---

not normally being slowly switched between relevantly dissimilar contexts. Because alternatives which are relevant in cases where one is being slowly switched are not normally relevant, one could go on, slow-switching objections do not show that if externalism's driving intuition is true, subjects cannot normally know groundlessly what they are thinking. Consequently, such cases do not show that there is a tension between externalism's driving intuition and the driving intuition about self-knowledge. For examples of this response, see Sawyer (1999) and Brown (2004). For replies to the response, see Ludlow (1995) and Butler (1997).

[16] Burge is not alone in drawing attention to such judgments within the context of the debate about the compatibility of content-externalism and our intuitions about self-knowledge. See, for instance, Heil (1988).

[17] We might wonder about cases in which the switching is such that one is not thinking anything determinate at all. Isn't the ascription 'I am, with this very thought, thinking that *p*' false in such a case? I think Burge's reply would be that in such cases, you do not really succeed in self-ascribing a thought to begin with.

this difference will necessarily be reflected at the second-order level.[18] Ignorance of one's surroundings will not affect the truth-value of second-order first-person judgments which ascribe occurrent mental states.

What about self-ascriptions of remembered states or long-term standing states? It might seem that it is possible for slow-switching to bring about an error in the case of these judgments. Take the case of remembered states. Suppose that five years ago, before his slow-switching ordeal began, Alf believed that his arthritis would only get worse. Now suppose that after having spent time on Twin Earth, Alf remembers this belief and remarks 'Five years ago, I believed that my arthritis would only get worse'. Because Alf has most recently been passing time on Twin Earth, one might think that Burge is committed to saying that when Alf utters these words, he ascribes to himself a past belief involving the concept tharthritis. But if this is the content of the belief Alf is ascribing to himself, then the self-ascription is false. Alf's earlier belief involved the concept arthritis, not the concept tharthritis. So it might look as though this is a case where one's self-ascription is in error because of a mismatch in content brought about by slow-switching between the relevant first-order state and the self-ascription.[19]

According to Burge, this sequence of reasoning rests on a mistaken assumption about what happens when one remembers a past state. On Burge's view, memory can function 'to preserve past thoughts for current use, without introducing new contents or attitudes (for example, as premises), with their own warrants and subject matter, into current cognition' (Burge, 2013c: 12).[20] If memory is functioning in its purely preservative capacity:

> … and if the causal-memory chains are intact, the individual's self-attribution is a reactivation of the content of the past one, held in place by a causal memory chain linking present to past attributions. (Burge, 2013d: 94)

[18] A similar point is made by Davidson (2001c: 30).
[19] For an objection along these lines see Boghossian (1998: 171-172).
[20] For more on purely preservative memory see Burge (1993).

When Alf remembers his earlier thought, his memory functions by preserving the content of that thought. Consequently, the content of his past-tense self-ascription is that he used to believe that his *arthritis* would only get worse. So the worry is misplaced. In the case of remembered states at least, there is no reason to think that slow-switching will bring about a mismatch in content between the original thought and the self-ascription and hence no reason to think that it will be directly responsible for errors in one's self-ascriptions.

What about self-ascriptions of long-term standing states? Suppose that for some time prior to his being slowly switched—six years, say—Alf believed that his arthritis will only get worse. Suppose further that having spent time (but less than six years) on Twin Earth, Alf remarks 'I have, for the past six years, believed that my arthritis will only get worse'. Again, because Alf has most recently been passing time on Twin Earth, one might think that there is pressure on Burge to say that when Alf utters these words, he is ascribing to himself the belief that he has, for the past six years, believed that his *tharthritis* will only get worse. But if this is the content of the belief which Alf is ascribing to himself, then the self-ascription is false. Alf has not, for the past six years, believed that his tharthritis will only get worse. For at least some of that time, Alf has believed that his *arthritis* will only get worse. In this case, it seems that slow-switching is directly responsible for an error in Alf's self-ascription, not by bringing about a mismatch in the content of the relevant first-order state and the self-ascription, but by bringing it about that Alf has a series of different states over time.

Burge's response to worries of this sort is less clear than his response to worries about remembered states.[21] He could deny that the content of Alf's standing state changes when he is switched to Twin Earth. But it is difficult to see what the rationale for this denial might be, given that Alf will likely come to depend on Twin Earth communal standards after the switch. Alternatively, Burge could insist that there is no good reason to begin with to think that we are reliable at judging what thoughts we have been thinking for the past, say, six years. So even if the

[21] Burge discusses standing states in connection with slow-switching cases briefly in his 2003b (431-432). His remarks do not, however, address the objection which I raise here.

objection shows that slow-switching is directly responsible for the error in Alf's self-ascription, it does not show that the first-person judgments of a subject who is being slowly switched are any more vulnerable to error than they would be otherwise. The proponent of slow-switching objections could concede this point. They might insist, however, that the claim that changes in one's environment might be directly responsible for errors in one's judgments about one's long-term standing states is in tension with our intuitive picture of self-knowledge. Burge could agree that this claim is in tension with our intuitive picture. But he could deny that this is a reason to reject the claim, as opposed to the relevant component of the intuitive picture. If he were to argue in such a way, he would be deploying the sort of strategy which I gestured at towards the close of Chapter 1, Subsection 2.1.

In any case, I do not want to pursue this issue further here. Let us assume, for the time being at least, that if Burge's views about the individuation of mental content are correct, then slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. Note, however, that this claim about the accuracy of one's judgments about one's mental states is not sufficient to block slow-switching objections. The proponents of such objections can accept this claim and yet still claim that slow-switching undermines the *knowledgeability* of a subject's mental states. In order for *S*'s judgment that she is thinking that *p* to constitute knowledge, that judgment must be warranted. *S*'s judgment that she is thinking that *p* is not warranted, the proponent might continue, unless she possesses certain discriminative abilities.[22] More specifically, the proponent might insist that the notion of warrant is bound up with the notion of discriminative abilities such that the following claim is true:

> ***Discrimination***: *S* is warranted in judging, on the basis of reflection alone, that she is thinking that *p* only if she is able to discriminate on the basis of reflection alone between instances in which she is thinking that *p* and instances in

[22] Both Vahid (2003) and Brown (2004) press this line of argument.

which she is thinking that *q*, where *S*'s thinking that *q* is a relevant alternative to her thinking that *p*.

Why think that *Discrimination* is true? In answering this question, the proponent might defer to the apparent connection between discriminative abilities and knowledge generally.[23] Consider, for example, the case of perceptual knowledge. Suppose that Sally judges correctly that Jane drives a Mercedes. Moreover, suppose that Sally's judging that Jane drives a Mercedes is caused by Jane's driving a Mercedes. Perhaps Sally sees Jane one day while Jane is driving a Mercedes and forms her judgment on that basis. Whether or not we should say that Sally *knows* that Jane drives a Mercedes seems to be sensitive to facts about Sally's discriminative abilities. Specifically, if Sally cannot reliably discriminate between Mercedes and relevant alternatives, BMWs, say, then it seems that we should not say that Sally knows that Jane drives a Mercedes. The reason is that Sally's driving a BMW is a relevant alternative to her driving a Mercedes and if Sally cannot discriminate between BMWs and Mercedes, then for all Sally knows, Jane *does* drive a BMW.

In cases in which he is being slow-switched, the proponent might argue, Alf lacks the relevant discriminative ability. In such cases, Alf's thinking a thought involving the concept arthritis is a relevant alternative to his thinking a thought involving the concept tharthritis. But it seems that Alf is not able to discriminate on the basis of reflection alone between instances in which he is thinking a thought involving the concept arthritis and instances in which he is thinking a thought involving the concept tharthritis. Because he cannot discriminate on the basis of reflection alone, Alf is not warranted in judging on the basis of reflection alone that he is thinking a thought involving the concept tharthritis. Consequently, in cases where he is being slowly switched, Alf cannot know on the basis of reflection alone that he is thinking a thought involving the concept tharthritis.

[23] See, for example, Brown (2004: 41-42). The claim that there is connection between discriminative abilities and knowledge can be traced back at least as far as Goldman (1976).

Clearly, the force of this objection depends on how the notion of an ability to discriminate on the basis of reflection alone is understood. For the objection to worry Burge, an ability to discriminate on the basis of reflection alone between instances in which one is thinking that *p* and instances in which one is thinking that *q* must involve more than an ability to *think* that *p* or that *q*. If this was sufficient, then obviously Alf would possess the relevant ability. I take it that *S* is able to discriminate on the basis of reflection alone, in the sense relevant to *Discrimination*, between instances in which she is thinking that *p* and instances in which she is thinking some relevant alternative thought, that *q*, only if she is able on the basis of reflection alone, first, to reliably judge whether her thoughts have recently changed from thoughts that *p* to thoughts that *q*, or vice versa, and second, in cases where they have recently changed, to reliably judge when the change occurred. Clearly, *S* may be able to think that *p* and that *q* and yet not be in possession of either of these two further abilities.

It would seem that in cases where he is being slowly switched, Alf does not possess either of these two further abilities. He is not able to reliably judge whether his thoughts have recently changed from thoughts involving the concept arthritis to thoughts involving the concept tharthritis (or vice versa), and, if they have recently changed, when the change occurred. Consequently, it seems that in cases where he is being slowly switched, Alf is not able to discriminate on the basis of reflection alone between instances in which he is thinking a thought involving the concept arthritis and instances in which he is thinking a thought involving the concept tharthritis. So it seems that in such cases, he cannot know on the basis of reflection alone that he is thinking a thought involving the concept tharthritis. The analogous line of reasoning supports a similar conclusion about Carl, namely, that in cases in which he is being slowly switched, he cannot know on the basis of reflection alone that he is thinking a thought involving the concept twater.

The second component in Burge's response to slow-switching objections involves defending a claim about the source of our warrant for first-person judgments. This claim is intended inter alia to block this way of pressing slow-switching objections. The claim is that one's warrant for first-person judgments takes the form of an entitlement and derives from one's identity

as a critical reasoner—as a reasoner that at least sometimes engages in critical reasoning. It is not a justification which is grounded in one's ability to identify one's thoughts or to discriminate them from relevant alternatives.[24] If this claim is true, then *Discrimination* is false. And indeed, Burge explicitly rejects this principle:

> One knows one's thought to be what it is simply by thinking it while exercising second-order, self-ascriptive powers. One has no "criterion", or test, or procedure for identifying the thought, and one need not exercise comparisons between it and other thoughts in order to know it as the thought one is thinking. Getting the "right" one is simply a matter of thinking the thought in the relevant reflexive way. The fact that we cannot use phenomenological signs or empirical investigation to discriminate our thoughts from other thoughts that we might have been thinking if we had been in a different environment in no way undermines our ability to know what our thoughts are. We "individuate" our thoughts, or discriminate them from others, by thinking those and not the others, self-ascriptively. Crudely put, our knowledge of our own thoughts is immediate, not discursive. Our epistemic right rests on this immediacy, as does our epistemic right to perceptual beliefs. For its justification, basic self-knowledge in no way needs supplementation from discursive investigations or comparisons. (Burge, 1988: 61)

In this passage, Burge denies that in order to know what one is thinking, one must be able to discriminate one's thoughts from relevant alternatives in the sense relevant to *Discrimination* (although he agrees that there is a sense in which we are able to discriminate our thoughts, namely, by thinking those thoughts self-ascriptively). As the last line makes clear, Burge has explicitly in mind instances of *basic-self-knowledge*, that is, cogito-like judgments. But he holds the basic thought to be true of present tense judgments about one's mental states which are non-cogito-like.[25] On Burge's view, in order for Alf's judgment that he is thinking a thought involving the concept tharthritis to be warranted on the basis of reflection alone, Alf does not need to be able to discriminate on the basis of reflection alone between instances in which he is thinking a thought involving the concept arthritis and instances in which he is thinking a thought involving the

[24] To be clear, entitlements and justifications are two species of epistemic warrant, on Burge's view. The difference is that 'A *justification* is a warrant that consists partly in the operation or possession of a reason…An *entitlement* is a warrant whose force does not consist, even partly, in the individual's using or having a reason'(Burge, 2013c: 3-4).

[25] 'Much of the literature on this subject deals with problems that arise from the assumption that we need to *identify* the content of our thoughts in such a way as to be able to rule out relevant alternatives to what the content might be. Boghossian, unlike many of those who write on this subject, seems to recognise that this assumption is not acceptable on my view' (Burge, 2013d: 91).

concept tharthritis.[26] So the fact that in cases where Alf is being slowly switched, he is unable to do so in no way undermines the claim that in such cases, Alf can know groundlessly that he is thinking a thought involving the concept tharthritis.

On Burge's view, slow-switching cannot bring it about that one's first-person judgments are subject to error. But Burge endorses an even stronger claim. Burge thinks that slow-switching cannot undermine the *knowledgeability* of one's judgments about one's mental states. On Burge's view, slow-switching cannot bring it about, either that one's judgments about one's mental states are subject to error, or that those judgments, although true, do not constitute knowledge.

I think that Burge is absolutely right to reject *Discrimination*. I do not think that it is true, generally speaking, that in order to know on the basis of reflection alone what one is thinking, one must be able to discriminate, on the basis of reflection alone, one's thoughts from relevant alternatives. Of course, one can reject *Discrimination* without endorsing Burge's positive proposal—the claim that our warrant for first-person judgments takes the form of an entitlement which derives from one's identity as a critical reasoner. In Section 3, I am going to elucidate this claim and the general picture of self-knowledge which informs it.

3. BURGE'S ACCOUNT OF OUR WARRANT FOR FIRST-PERSON JUDGMENTS

What does it mean to say that our warrant for first-person judgments takes the form of an entitlement which *derives* from our identity as critical reasoners? To answer this question, we need to understand what critical reasoning involves. On Burge's view:

> Critical reasoning is reasoning that involves an ability to recognise and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. (Burge, 1996: 98)

[26] Burge is not alone in rejecting *Discrimination*. Other philosophers who have rejected the principle include Goldberg (2006), Falvey and Owens (1994) and Bar-On (2004).

All reasoning requires a capacity to be responsive to reasons. What is distinctive about critical reasoning, Burge thinks, is that it requires a capacity to *reflect* on one's reasons and reasoning. To count as a critical reasoner a creature must be able to think of reasons *as such*, to evaluate reasons thus conceptualised and assess one's reasoning in the light of this evaluation.

On Burge's view, only mental states can be reasons: '… reasons are necessarily propositional contents taken with their modes' (Burge, 2013e: 193). So as Burge is thinking of it, a capacity to reflect on one's reasons just is a capacity to reflect on the rational, or reason-supporting, relations between mental states—in the case of one's *epistemic reasons*, the reasons one has for believing particular propositions—and between mental states and actions—in the case of one's *practical reasons*, the reasons one has for performing particular actions. Insofar as critical reasoning requires a capacity to recognise and evaluate one's reasons, it requires the capacity to recognise and evaluate the rational relations between mental states. Burge is explicit on this point:

> As a critical reasoner, one not only reasons. One recognises reasons as reasons. One evaluates, checks, weighs, criticises, supplements one's reasons and reasoning. Clearly, this requires a second-order ability to think about thought contents or propositions, and rational relations among them. (Burge, 1996: 98)

According to Burge, our identity as critical reasoners entails an ability to think reflectively about our mental states, our reasons and our reasoning. Indeed, on Burge's view it entails something stronger, namely, 'that that thinking be normally knowledgeable' (Burge, 1996: 100). In other words, Burge endorses the following claim:

> ***Knowledgeable Reviewability***: In order for one to be a critical reasoner, one's mental states, reasons and reasoning must normally be knowledgeably reviewable.

Knowledgeable Reviewability is really the conjunction of two separate claims, which I am going to call *Entitlement* and *Correctness*, respectively:

Entitlement: In order for one to be a critical reasoner, one must have an entitlement to judgments about one's mental states, reasons and reasoning.

Correctness: In order for one to be a critical reasoner, judgments about one's mental states, reasons and reasoning must normally be correct.

The truth of both *Entitlement* and *Correctness* is sufficient for the truth of *Knowledgeable Reviewability*.[27] If it is true that in order for one to be a critical reasoner, first, one must be entitled to judgments about one's mental states, reasons and reasoning, and second, those judgments must normally be correct, then it is true that in order to be a critical reasoner, one's mental states, reasons and reasoning must normally be knowledgeably reviewable. But why think that *Entitlement* and *Correctness are* true? In answering this question, Burge makes appeal to a third claim:

Reflection: Reflection on our mental states, reasons and reasoning adds a rational element to reasoning by giving one rational control over one's reasoning.[28]

To improve one's reasoning on the basis of careful reflection on one's mental states, reasons and reasoning just is to reason critically, on Burge's view. So *Reflection* just is the claim that critical reasoning is possible.

Burge's strategy is to argue from *Reflection*—the truth of which he apparently regards as beyond dispute—to *Entitlement* and *Correctness*, respectively. Consider first *Entitlement*.

---

[27] But it is not necessary. As I pointed out in fn. 24, entitlements and justifications are two species of epistemic warrant, on Burge's view. Instead of taking the form of an entitlement, our warrant for first-person judgments could conceivably take the form of a justification. In this case, *Knowledgeable Reviewability* would be true but *Entitlement* would be false.

[28] 'Adds a rational element to reasoning' implies that we could still be reasoners, even if we could not reflect on our reasons as reasons. And this is indeed Burge's view: 'A non-critical reasoner reasons blind, without appreciating reasons as reasons. Animals and small children reason in this way' (Burge, 1996: 99).

Suppose that we lacked an entitlement to first-person judgments. In this case, Burge claims, reflection could not add a rational element to reasoning. In order for reflection to add a rational element to reasoning, one's judgments which one forms on the basis of that reflection about one's mental states, reasons and reasoning must be reasonable: 'To be reasonable in the whole enterprise [of critical reasoning], one must be reasonable in that essential aspect of it' (Burge, 1996: 101). But one's judgments about one's mental states, reasons and reasoning are only reasonable if one is entitled to those judgments. So, if we lacked an entitlement to first person judgments, then *Reflection* would be false—reflection on one's mental states, reasons and reasoning could not add a rational element to reasoning. But reflection *does* add a rational element to reasoning, so we must have an entitlement to judgments about our mental states, reasons and reasoning. Since *Reflection* just is the claim that critical reasoning is possible, we can conclude that one must have an entitlement to judgments about one's mental states, reasons and reasoning in order to be a critical reasoner, that is, that *Entitlement* is true.

Burge makes a similar move in defence of *Correctness*. If our self-ascriptions were not normally correct, he argues, then reflection could not improve the reasonability of reasoning, for reflection 'could not rationally control and guide the attitudes being reflected upon' (Burge, 1996: 102). But reflection *can* improve the reasonability of reasoning, so *Correctness* must be true. Again, since the claim that reflection adds a rational element just is the claim that critical reasoning is possible, we can conclude that in order for one to be a critical reasoner, one's judgments about one's mental states, reasons and reasoning must normally be correct, that is, that *Correctness* is true.

It is not my present concern to assess Burge's defence of either *Entitlement* or *Correctness*. My aim is simply to draw the reader's attention to the form of that defence, which is the same in each case. In each case Burge appeals to a third claim, *Reflection*—the claim that reflection on our mental states, reasons and reasoning adds a rational element to reasoning by giving one rational control over one's reasoning—which he regards as obviously true. He then argues that the truth of

Entitlement and the truth of *Correctness* are necessary conditions for the truth of *Reflection*, so that the truth of *Reflection* entails the truth of both *Entitlement* and *Correctness*.

Of course, we still want an answer to what Burge means when he says that our entitlement to first-person judgments derives from our identity as critical reasoners. We have established that on Burge's view, being entitled to judgments about one's mental states, reasons and reasoning is a necessary condition for being a critical reasoner. But in what sense does the entitlement itself *derive* from one's identity as a critical reasoner? On Burge's view, one's entitlement to judgments about one's mental states, reasons and reasoning derives from one's identity as a critical reasoner in the sense that it is an entitlement that 'consists in a status of operating in an appropriate way in accord with norms of [critical] reason…' (Burge, 1996: 93). To operate in an appropriate way in accord with norms of critical reason is, I take it, to operate in a way which is *guided by* the norms of critical reason. One is entitled to the relevant judgment, on Burge's view, only insofar as one is operating in a way which is guided by the norms of critical reason. If one is operating in a way which is not guided by the norms of critical reason, then one's entitlement to the judgment lapses.[29]

In effect, then, Burge's account of our entitlement to self-knowledge is at base circular. Burge thinks that our having an entitlement to first-person judgments is a necessary condition for our being critical reasoners. If we were not entitled to judgments about our mental states, reasons and reasoning, then those judgments would not be reasonable. But if our judgments about our mental states, reasons and reasoning were not reasonable, then we would not be critical reasoners. But at the same time, Burge wants to say, our being critical reasoners—our operating in a way

[29] What are the norms of critical reason? In his 2013e, Burge writes, 'Norms of critical reason include norms of first-order rationality—those that apply to any reasoner—together with those rational norms that are specific to critical reasoners. Norms specific to critical reason apply only to propositional states and events that are in principle accessible to an individual's rational powers, immediate self-understanding, and rational evaluation, using the concept reason or some variant' (173). Explicit examples of norms which satisfy this description are, however, few are far between. One such example is given on p. 216. Burge tells us that the following is a norm of critical reasoning: (CR): If, in critical reasoning, one correctly and with warrant judges that a lower-level state is (or is not) reasonable, then it rationally follows directly that one has reason to sustain (or change) the lower-level state.

which is guided by the norms of critical reason—is a necessary condition for our being entitled to those judgments about our mental states, reasons and reasoning.

In Chapter 1, Subsection 2.2 I set out a schema for classifying accounts of self-knowledge on the basis of two considerations. The first consideration is whether the account is *epistemic* or *non-epistemic*. The second consideration is whether the account is *substantive* or *non-substantive*. An account is epistemic, I suggested, if it maintains that, in the normal case at least, one's warrant for first-person judgments has its basis in some epistemically privileged way of knowing. I suggested that an account is substantive if it holds that first-person judgments involve a cognitive achievement, where the judgment that one, say, believes that *p* involves a cognitive achievement if and only if it involves the detection of a state of affairs which exists at least partly independently of one's judging (under ideal conditions) that one believes that *p*.

How does Burge's account fit into this scheme? Burge's account is most fairly described as epistemic and substantive. The account is epistemic because it identifies some epistemically privileged way of knowing—one's operating in a way which is guided by the norms of critical reason—as grounding one's warrant for those judgments. It is substantive because it maintains that first-person judgments involve a cognitive achievement. If the account is correct, then judgments about one's mental states, reasons and reasoning at least sometimes involve the detection of states of affairs which exist at least partly independently of one's judging (under ideal conditions) that they do.[30]

CONCLUSION

[30] Cogito-like judgments, which I discussed in Chapter 2, Section 2, are an example of first-person judgments which on Burge's view *do not* involve a cognitive achievement in this sense. My judging that 'I am, with this very thought, thinking that my arthritis has worsened' makes it true that I am indeed thinking that my arthritis has worsened.

The discussion in this chapter has been structured around two aims. The first aim was to introduce a familiar line of argument for the claim that externalism's driving intuition and the driving intuition about self-knowledge are inconsistent. The line of argument in question begins with the assumption that externalism is true. There is then a move from this assumption to the claim that we cannot normally know our mental states groundlessly. The second aim was to consider Burge's response to this line of argument.

The first aim was pursued in Section 1. I introduced slow-switching objections and defended the view that these objections have initial plausibility as objections to the arthritis and water cases, but not as objections to the sofa case. The second aim was pursued in Sections 2 and 3. In Section 2, I considered Burge's response to slow-switching objections. That response has two components. The first component involves defending a claim about the *accuracy* in slow-switching cases of one's judgments about one's mental states. The claim is that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. The second component of Burge's response involves defending a claim about the source of one's warrant for judgments about one's mental states. The claim is that this warrant takes the form of an entitlement and derives from one's identity as a critical reasoner. It is not grounded in one's ability to identify one's thoughts or discriminate them from relevant alternatives. Consequently, the fact that in slow-switching scenarios, subjects cannot discriminate their mental states from relevant alternatives in the relevant sense does not undermine their warrant for judgments about their mental states. Together, these two components support the claim that slow-switching cannot undermine the *knowledgeability* of one's judgments about one's mental states, a claim which Burge explicitly endorses. On Burge's view, slow-switching cannot bring it about, either that one's judgments about one's mental states are subject to error, or that those judgments, although true, do not constitute knowledge. In Section 3, I considered what it means to say that one's warrant for first-person judgments takes the form of an entitlement which *derives* from one's identity as a critical reasoner. I concluded that one's warrant for judgments about one's mental states derives from one's identity as a critical reasoner, on Burge's view, in the sense that it is an

entitlement that ‘consists in a status of operating in an appropriate way in accord with norms of [critical] reason…’

Chapter 3

SLOW-SWITCHING, EPISTEMIC REASONS AND BRUTE SUCCESS

INTRODUCTION

In Chapter 2, Section 1 I discussed a class of objections which purport to show that given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of a subject's mental states. In this chapter, I develop a different sort of slow-switching objection to Burge's views. I argue that given Burge's views about the way in which mental content is individuated, slow-switching can undermine knowledgeability of a subject's epistemic reasons. This is a prima facie troubling result, I claim, given Burge's views about the sorts of things that epistemic reasons are. As the discussion in Chapter 2, Section 3 made clear, on Burge's view epistemic reasons are rational relations between mental states. If, as I argue, slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of these relations. This result is prima facie troubling, for it seems to be part of our intuitive picture of self-knowledge that the knowledgeability of the rational relations between one's mental states is not sensitive to changes in one's context in this way. It is part of our intuitive picture that the knowledgeability of the rational relations between mental states is sensitive to certain considerations, for example, facts about the subject's reasoning competency, level of attention, and so on. But considerations having to do with one's context do not seem to be among them.

The chapter consists of three sections. In Section 1, I introduce Burge's distinction between *brute* and *non-brute* error. In Section 2, I introduce the *objection from brute error*, which

on a first pass seems to demonstrate that if Burge's views about the individuation of mental content are true, then slow-switching can undermine knowledgeability of one's epistemic reasons. I go on to show that the objection from brute error rests on an assumption which Burge would reject, given his views about having reasons. In Section 3, I set out a second objection, *the objection from brute success*, which shows that, given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of a subject's epistemic reasons even if we grant his views about having reasons.

1. BRUTE ERROR AND BRUTE SUCCESS

The discussion in Chapter 2, Section 3 established that Burge endorses *Knowledgeable Reviewability*, the claim that in order for one to be a critical reasoner, one's mental states, reasons and reasoning must normally be knowledgeably reviewable. *Knowledgeable Reviewability* is really the conjunction of two separate claims, which I called *Entitlement* and *Correctness*, respectively. According to *Entitlement*, in order for one to be a critical reasoner, one must have an entitlement to judgments about one's mental states, reasons and reasoning. According to *Correctness*, in order for one to be a critical reasoner, judgments about one's mental states, reasons and reasoning must normally be correct.

It is consistent with *Correctness* that one might be a critical reasoner even though one's judgments about one's mental states, reasons and reasoning are sometimes incorrect. Indeed, it is consistent with that claim that one might be a critical reasoner even though those judgments are *often* incorrect. Their often being incorrect is, after all, compatible with their *normally* being correct. Burge distinguishes between two kinds of error to which judgments in general might be subject: *brute* and *non-brute*.[1] Non-brute errors arise from some malfunction or irrationality within

[1] The distinction is first drawn in Burge (1988: 657).

the subject or from carelessness on the part of the subject. Brute errors, in comparison, are due entirely to other factors, such as abnormal conditions or misleading evidence (Burge, 1988: 657).

How are the notions of irrationality, malfunction and carelessness to be understood? A rational failure is some error in reasoning, as when, for example, one reasons from if *p* then *q* and not *p* to not *q*. A malfunction is typically a biological failure. For example, my perceptual system has malfunctioned if damage to that system causes me to have a non-veridical visual impression as of, say, a moving object. If on the basis of my visual impression I judge that there is a moving object, my judgment is subject to non-brute error. The rational failure or malfunction in question may be due to carelessness on the part of the subject, but it need not be. In reasoning from if *p* then *q* and not *p* to not *q* I may be reasoning attentively, to the best of my ability, and so on. Nevertheless, a rational failure has occurred. Similarly, I may not have any reason to assume that my perceptual system is damaged. But if it is in fact damaged and the damage causes me to have a non-veridical visual impression, then my perceptual system has malfunctioned.

There are cases where a subject judges incorrectly, not because of any malfunction or irrationality, but because they are careless. For example, suppose that a subject has a non-veridical visual impression as of a sheep in the field and on the basis of this impression judges that there is a sheep in the field. Suppose further that in making the judgment, the subject is ignoring evidence that the situation is abnormal (perhaps she had earlier been told by the landowner that the actual sheep have been replaced by carefully crafted representations of sheep). Assuming that the subject's visual impression is caused by one of the carefully crafted representations of sheep, her judgment that there is a sheep in the field is not due to any malfunction in her visual system or to any irrationality. Rather, it is due entirely to her carelessness.[2] The subject's error in such cases is, as it is in cases involving irrationality or malfunction, non-brute.

[2] One might wonder whether there is as sharp a distinction between carelessness and irrationality as this claim seems to imply. After all, isn't there a sense in which the subject is being irrational in ignoring what she has been told by the landowner? Perhaps, but I do not think this is the sense which Burge has in mind. Failures of rationality are errors in reasoning. One is being irrational, in Burge's sense, if one either fails to infer what one has reason to infer or infers what one does not have reason to infer. In comparison, one is

Brute errors do not arise from rational failure, malfunction or carelessness in, or on the part of, the subject. In the case of perceptual judgments, there are several different ways in which brute errors might come about. For example, to adapt a case first imagined by Goldman, suppose that unbeknownst to me, I am travelling through an area of countryside in which there is a high proportion of facsimile barns.[3] Although fake, these facsimile barns look just like real barns to the average passer-by. Suppose that I have a visual impression caused by one of these facsimile barns and, on the basis of that visual impression, non-inferentially form the judgment that the structure in front of me is a barn. The structure in front of me is not in fact a barn but a facsimile, so my judgment is false. But it is not, or at least need not be, false because of any malfunction in my visual system. Let us assume that my visual system is functioning perfectly. Moreover, let us assume that in judging as I do, I am not ignoring any evidence that the area of countryside through which I am travelling is one in which there is a high proportion of facsimile barns. In this case, my error is brute, for it is due entirely to abnormal circumstances.[4]

Why does Burge draw a distinction between, on the one hand, errors which are the result of some malfunction, irrationality or carelessness in, or on the part of, the subject and on the other, errors which are the result of environmental abnormalities or misleading evidence? The distinction between brute and non-brute error is meant to map onto the distinction between judgments that are warranted and false and judgments which are unwarranted and false. Non-brute errors undermine one's warrant for the relevant judgment. Brute errors do not. If a judgment is false but warranted, then the error is brute. Burge is clear on this point: 'Brute error is the type of error that is compatible with being warranted in one's belief…' (Burge, 2013c: 10-11). Being warranted in one's judgments is, on Burge's view, a matter of according with relevant norms. Instances of

being careless if one either fails to include in one's premises what one has reason to include or fails to exclude what one has reason to exclude. The subject who ignores what she has been told by the landowner is guilty of an error of this latter sort.

[3] This scenario in question is introduced in Goldman (1976: 772).

[4] There is, of course, an assumption at work here about the level of checking required by normal standards of carefulness. It is being assumed that I am not being careless in failing to check whether the structure is a facsimile barn.

malfunction, irrationality and carelessness are instances in which the relevant norms—the norms associated with perception or with critical reasoning—are not accorded with.[5]

Of course, this reply simply moves our question back a step. Why does Burge want a distinction between kinds of error which is extensionally equivalent to the distinction between judgments which are false and warranted and judgments which are false and unwarranted? Burge's concern is with giving accounts of our epistemic warrant for, respectively, perceptual and first-person judgments. Ideally, such accounts should tell us something about the circumstances under which the relevant warrant lapses. Having in hand a distinction between, on the one hand, errors which are such that they undermine one's warrant for perceptual or first-person judgments and, on the other, errors which are such that they do not, is important in satisfying this desideratum.

I said earlier that it is consistent with *Correctness* that one might be a critical reasoner even though one's judgments about one's mental states, reasons and reasoning are often incorrect. However, according to Burge, there are 'severe limits' on the vulnerability of one's judgments about one's 'present ordinary, accessible' mental states[6] (hereafter simply *accessible mental states*) to brute error (Burge, 1996: 103).[7] A severe limit is not, however, an absolute limit. On

[5] With regards to perception, one might wonder whether there are not *some* cases of malfunction which do not undermine one's warrant for the relevant perceptual judgment. Suppose that my visual system malfunctions on a particular occasion but that I have no evidence that it has malfunctioned. Suppose further that the malfunction is not such that it undermines the general reliability of my visual system. Given the nature of the malfunction, we might want to say that judgments which I make on the basis of my visual impression are warranted. Burge agrees. He denies, however, that the warrant is such that, in cases where the perceptual state just happens to be veridical in spite of the malfunction, the relevant judgment could constitute knowledge: 'Token events that are momentary malfunctions in a reliable perceptual system and that produce a reliable perceptual state undermine knowledge but (I conjecture) not entitlement' (Burge, 2003c: 537, fn. 24). Instances in which the malfunction undermines the veridicality of the perception (but not the general reliability of the perceptual system) are akin to instances of brute error. Instances in which the perceptual state just happens to be veridical in spite of the malfunction are instances of what I go on to describe as *brute success*—the relevant judgment is both warranted and true but does not constitute knowledge.

[6] By 'accessible' here, I take it that Burge means accessible in the normal way. A mental state is accessible in the normal way if it is accessible on the basis of careful reflection alone. This clarification is important because there is a sense in which unconscious mental states are accessible—one may access one's unconscious beliefs by attending to the things one says and does—and yet Burge does not think that there is a severe limit on the vulnerability of judgments about *these* mental states to brute error.

[7] Elsewhere, Burge endorses a stronger constraint: 'I think that, in all cases of authoritative knowledge, brute mistakes are impossible. All errors in matters where people have special authority about themselves are errors which indicate something wrong with the thinker' (Burge, 1988: 658); 'A key *explanandum* that I

Burge's view, being a critical reasoner is consistent with one's judgments about one's accessible mental states *sometimes* being subject to brute error.

It is clear, however, that Burge is committed to denying that one's judgments about one's mental states can be subject to brute errors brought about by slow-switching. Recall that Burge's response to slow-switching objections culminates in the claim that slow-switching cannot undermine the knowledgeability of one's mental states. This claim is true only if it is true that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. But an error in one's judgments about one's mental states which is brought about by slow-switching is a brute error. If it is brought about by slow-switching, then it is not an error which arises from any malfunction or irrationality within the subject or from carelessness on the part of the subject. So insofar as Burge is committed to denying that slow-switching can undermine knowledgeability of a subject's mental states, he is committed to denying that slow-switching can bring it about that one's judgments about one's mental states are subject to brute error. Being a critical reasoner may be consistent with one's judgments about one's accessible mental states sometimes being subject to brute error. But Burge's response to slow-switching objections entails that the constraint on brute errors brought about by slow-switching is absolute.

For the same reason, Burge is committed to denying that slow-switching can bring it about that one's judgments about one's mental states are subject to what I am going to call *brute success*.[8] Cases of brute success are cases in which a judgment is both true and warranted but falls short of constituting knowledge. Earlier I imagined a scenario in which whilst travelling through an area of countryside in which there is a high proportion of facsimile barns, I judge falsely on the basis of a visual impression that this object in front of me is a barn. Suppose that we modify the scenario slightly. Suppose that I am driving through an area in which there is a high proportion of

proposed for any account of authoritative self-knowledge is that authoritative self-attributions do not appear to be subject to *brute error*' (Burge, 1999: 34). Whichever happens to be Burge's considered view, he certainly does think that certain *types* of self-ascriptions of accessible mental states are immune to brute error. See, for example, his 2013e (210-220).

[8] What I call brute success, Burge calls *brute truth*. See, for example, Burge (2003c: 507).

facsimile barns but that my judgment that this object in front of me is a barn happens to be true. I am not being careless in judging that the object in front of me is a barn. Nor is my judgment indicative of any malfunction or irrationality. So it seems right to say that my judgment is warranted. However, even though my judgment is both true and warranted, it does not seem right to say that it constitutes knowledge, that I *know* that the object in front of me is a barn. After all, given my circumstances, its being a facsimile barn is a relevant alternative and for all I know, it *is* a facsimile barn.[9] In this case, my judgment that this object in front of me is a barn is, it seems, subject to brute success.

If slow-switching can bring it about that certain of one's judgments about one's mental states are subject to brute success, then it can undermine knowledgeability of one's mental states. Just as Burge is committed to denying that slow-switching can bring it about that one's judgments about one's mental states are subject to brute error, so too is he committed to denying that slow-switching can bring it about that one's judgments about one's mental states are subject to brute success.

To summarise, Burge's response to slow-switching objections—his claim that slow-switching cannot undermine knowledgeability of a subject's mental states—commits him to two further claims: first, the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to brute error; and second, the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to brute success.

2. SLOW-SWITCHING AND KNOWLEDGEABILITY OF EPISTEMIC REASONS

2.1 Knowledgeable Reviewability and Epistemic Reasons

[9] Indeed, this is how the scenario is imagined by Goldman.

On Burge's view, the requirements for knowledgeability which follow from one's identity as a critical reasoner concern not just one's mental states, but also one's reasons and reasoning. It is not just one's mental states, but also one's reasons and reasoning which, according to *Knowledgeable Reviewability,* must normally be knowledgeably reviewable in order for one to be a critical reasoner. Burge denies that slow-switching can undermine knowledgeability of one's mental states. Reflection on *Knowledgeable Reviewability* may prompt us to ask whether Burge ought also to deny that slow-switching can undermine knowledgeability of one's reasons and reasoning (that is, whether Burge ought to deny that slow-switching can bring it about either that one's judgments about one's reasons or reasoning are subject to error or that those judgments, although true, do not constitute knowledge). In the discussion that follows, I am going to focus exclusively on the question whether Burge ought to deny that slow-switching can undermine knowledgeability of one's epistemic reasons. I am going to leave questions concerning the knowledgeability of one's non-epistemic reasons and one's reasoning to one side.

Ought Burge to deny that slow-switching can undermine knowledgeability of one's epistemic reasons? It seems clear that if the answer to this question is 'Yes', it cannot be for the reasons that Burge denies that slow-switching can undermine knowledgeability of one's mental states. As we noted in Chapter 2, Section 2 Burge thinks that slow-switching cannot undermine knowledgeability of one's mental states in part because he thinks that slow-switching cannot bring it about that one's judgments about the content of one's mental states are subject to error. He thinks this because he thinks that slow-switching cannot bring about a mismatch in content between the relevant self-ascription and the relevant first-order thought. In the case of cogito- and non cogito-like ascriptions of occurrent mental states, changes in the content of one's first-order thoughts are necessarily reflected at the second-order level. In the case of self-ascriptions of remembered states, purely preservative memory ensures that the content of the ascription is the content of the remembered state.

Let us focus on one's judgments about one's occurrent epistemic reasons. In what follows, I am going to leave questions concerning self-ascriptions of remembered epistemic reasons to one

side. It is true that one's judgments about one's occurrent epistemic reasons will necessarily reflect changes in the content of one's mental states which are the result of slow-switching. For example, suppose that before I am slowly switched, I judge correctly that the fact that I believe that I have arthritis in my ankle is a reason for me to believe that *p*. Now suppose that slow-switching brings it about that instead of believing that I have arthritis in my ankle, I believe that I have *tharthritis* in my ankle. This change will necessarily be reflected in my judgments about my occurrent epistemic reasons. When I reflect on my epistemic reasons on Twin Earth, instead of judging that the fact that I believe that I have arthritis in my ankle is a reason for me to believe that *p*, I will judge that the fact that I believe that I have *tharthritis* in my ankle is a reason for me to believe that *p*.

But it is clear that this fact—the fact that changes in the content of one's mental states brought about by slow-switching are necessarily reflected in one's judgments about one's occurrent epistemic reasons—does not suffice to show that slow-switching cannot bring it about that one's judgments about one's occurrent epistemic reasons are subject to error (as the analogous consideration does in the case of one's judgments about one's occurrent mental states). One may simply lack the epistemic reason which, as a consequence of changes in content brought about by slow-switching, one has come to ascribe to oneself. Suppose that my believing that I have tharthritis in my ankle is *not* a reason for me to believe that *p*. In this case, even though changes in content brought about by slow-switching are reflected in my judgments about my occurrent epistemic reasons, slow-switching has brought it about that at least one of those judgments is subject to error.

It would be a prima facie troubling result for Burge if it should turn out that slow-switching can undermine knowledgeability of one's epistemic reasons. But the reason it would be a prima facie troubling result does not have to do with Burge's reasons for denying that slow-switching can undermine knowledgeability of one's mental states. Rather, it has to do with Burge's views about the sorts of things that epistemic reasons *are*. On Burge's view, epistemic reasons are rational relations between mental states. Given this view, if slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of

the rational relations between one's mental states. But this would be prima facie troubling, for it seems to be part of our intuitive picture of self-knowledge that the knowledgeability of the rational relations between one's mental states is not sensitive in this way to changes in one's context. It is part of our intuitive picture that the knowledgeability of the rational relations between one's mental states is sensitive to certain considerations, for example, facts about the subject's reasoning competency, level of attention, and so on. But considerations having to do with one's context do not seem to be among them.

I said that it would be prima facie troubling if it should turn out that slow-switching could undermine knowledgeability of the rational relations between one's mental states. Why qualify my statement in this way? In Chapter 1, Subsection 2.1 I drew a distinction between what I am calling the driving intuition about self-knowledge—the thought that we normally know groundlessly what we are thinking—and the various other guiding intuitions about self-knowledge which belong to what I am calling our intuitive picture of self-knowledge. I explained the difference in this way. I understand the driving intuition about self-knowledge to constitute a datum, information about self-knowledge against which candidate theories about the individuation of mental content ought to be assessed. If it should turn out that externalism's driving intuition is in tension with the claim that we normally know groundlessly what we are thinking, then this is automatically a reason to reject or revise externalism's driving intuition. The various other intuitions which belong to our intuitive picture of self-knowledge do not have this status. A tension between externalism's driving intuition and one or more of these intuitions is not automatically a reason to reject or revise externalism's driving intuition. When such tensions do arise, we should take seriously the possibility that this is a reason for us to reject or revise the components of the intuitive picture, rather than externalist intuitions about the individuation of mental content.

The thought that slow-switching cannot undermine the knowledgeability of the rational relations between one's mental states has strong intuitive appeal. But I do not take it to have the status of a datum. If it should turn out that, given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of the rational relations between one's

mental states, this result would not automatically constitute a reason to revise those views. Rather, we should take seriously the possibility that it is a reason to reconsider the thought that slow-switching cannot undermine knowledgeability of the rational relations between one's mental states.

2.2 The Objection from Brute Error

I have argued that it would be a prima facie troubling result for Burge, given his views about the sorts of things that epistemic reasons are, if it should turn out that slow-switching can undermine knowledgeability of one's epistemic reasons. In this subsection I am going to do two things. First, I am going to present a slow-switching objection which on a first pass seems to show that, given Burge's views about the individuation of mental content, slow-switching *can* undermine knowledgeability of one's epistemic reasons.[10] I call this *the objection from brute error*. Second, I am going to differentiate the objection from brute error from a similar objection raised by Paul Boghossian.

Imagine that Jack, a competent member of an English speaking socio-linguistic community on Earth, has various thoughts involving the concept arthritis. For example, Jack believes that he has arthritis in his ankles, that the arthritis in his ankles is worse than the arthritis in his wrists, and so on. Let us suppose that Jack grasps the meaning-giving characterisations associated with the term 'arthritis' (unlike Alf, Jack understands that arthritis is an inflammation of the joints, that it is an ailment which cannot, for instance, afflict one's thigh). But let us suppose that Jack's knowledge about arthritis is limited and that he defers to the experts when he uses the term 'arthritis'.

Jack believes:

[10] Although developed independently, this case bears similarities to cases discussed by Brown (2000). But there are also significant differences. For example, Brown's cases do not involve slow-switching. They do not purport to show that slow-switching can undermine knowledgeability of one's epistemic reasons.

(1) My ankles are afflicted with—and only with—arthritis

and:

(2) Arthritis is not an ailment that can afflict my thigh

Were Jack to reflect on (1) in the light of (2), he might well infer that:

(3) My ankles are not afflicted with an ailment that can afflict my thigh

Suppose that we were to put it to Jack that (1) and (3) are inconsistent beliefs and that, because they are inconsistent, Jack has a reason to give up either (1) or (3). Jack would likely reject our contention. Reflecting on the fact that (3) is validly inferred from (1) and (2), Jack would likely respond that (1) and (3) are consistent beliefs and that therefore he does not have a reason to give up either (1) or (3). In responding in this way Jack would, of course, be correct. Insofar as (1) and (3) are thoughts about arthritis, they *are* consistent beliefs.

Now suppose that Jack is being slowly switched between Earth and Twin Earth, where speakers use the term 'arthritis' to express the concept tharthritis. Suppose that the switching is done in such a way that in the absence of empirical investigation, Jack cannot tell that he is being switched. We can expect that, given enough time, the experts to whom Jack defers will come to be the experts on Twin Earth. Jack understands that arthritis is an ailment of the joints, so he possesses knowledge which could enable him to discriminate between instances of arthritis and instances of tharthritis. In Chapter 2, Section 1 I reached a conclusion about what Burge would say regarding slow-switching cases involving subjects, like Jack, who possess relevant discriminatory knowledge but come to rely on Twin Earth communal standards. I suggested that on Burge's view, slow-switching will result in their coming to acquire a second, distinct concept. Consequently, it is in keeping with that view to think that Jack will come to use the term 'arthritis' on Twin Earth to express, not the concept arthritis, but the concept tharthritis.

On Twin Earth, Jack believes:

(4) My ankles are afflicted with—and only with—*tharthritis*

and:

(5) *Tharthritis* is not an ailment that can afflict my thigh

Were Jack to reflect on (4) in the light of (5), he might well infer:

(6) My ankles are not afflicted with an ailment that can afflict my thigh

Now suppose that we were to put it to Jack that (4) and (6) are inconsistent beliefs and that, because they are inconsistent, Jack has a reason to give up either (4) or (6). As before, Jack would likely reject our contention. Reflecting on the fact that (6) is validly inferred from (4) and (5), Jack would likely respond that (4) and (6) are consistent beliefs and that therefore he does not have a reason to give up either (4) or (6). But in giving this response, Jack would be incorrect. Insofar as (4) is a belief about *tharthritis*, (4) and (6) *are* inconsistent. We might think that because they are inconsistent, Jack has a reason to judge that they are inconsistent. Because he has a reason to judge that they are inconsistent, Jack has a reason to give up either (4) or (6). It seems that slow-switching has brought it about that Jack's judgments about his epistemic reasons are subject to brute error; Jack judges that he does not have a reason to give up either (4) or (6) when he in fact does. To this extent, it seems that slow-switching has undermined the knowledgeability of Jack's epistemic reasons.

The objection I have just raised purports to show that, given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of one's epistemic reasons by bringing it about that one's judgments about one's epistemic reasons are subject to brute error. For ease of discussion I will refer to it hereafter as *the objection from brute error*.

The objection from brute error as I have just formulated it relies on *Misunderstanding*, the claim that a subject can have thoughts involving concepts which she misunderstands (recall that this claim was first introduced in Chapter 1, Subsection 1.2 in connection with Burge's arthritis case). Judged according to the criteria for misunderstanding a concept as outlined in Chapter 1, Subsection 1.2, Jack misunderstands the concept tharthritis. Jack believes that necessarily, tharthritis cannot afflict one's thigh. But necessarily, tharthritis *can* afflict one's thigh. Moreover, Jack believes that the experts to whom he defers think as he does about tharthritis. But the experts to whom Jack defers *do not* think as he does about tharthritis. The experts believe that necessarily, tharthritis can afflict one's thighs. In the case as described, it is only because Jack misunderstands the concept tharthritis that he is wrong about his epistemic reasons. If he did not misunderstand the concept, then he would not hold (5) to be true. And if he did not hold (5) to be true, then he would not judge (6) to be true. But it is only because Jack judges (6) to be true that he has inconsistent beliefs. If he did not believe (6), then Jack would not be in error when he judges that his beliefs are consistent and that therefore he does not have a reason to give up one or another of his beliefs.[11]

The objection from brute error bears similarities to an objection raised by Paul Boghossian.

I think that Burge's response to Boghossian's objection is effective. But it will not suffice to block the objection from brute error. This is because the objection from brute error, unlike Boghossian's objection, does not rely on mistaken assumptions about the equivalence of concepts

[11] In Jack's case, conceptual misunderstanding brought about by slow-switching leads him to judge that two beliefs are consistent when they are in fact inconsistent, which in turn leads him to judge that he does not have a reason when he in fact does. Might there be cases of the opposite kind, cases in which conceptual misunderstanding brought about by slow-switching leads a subject to judge that two beliefs are inconsistent when they are in fact consistent, which in turn leads him to judge that he does have a reason when he in fact does not? I do not want to rule out the possibility that there are such cases. But if there are, then they are more complex in construction than Jack's case. Certainly, slow-switching can bring it about that a subject who has inconsistent beliefs on Earth, has consistent beliefs on Twin Earth. Suppose, for example, that on Earth, Jack believes: (*i*) My ankles are afflicted with—and only with—arthritis; (*ii*) Arthritis is an ailment that can afflict my thighs; and (*iii*) Therefore, my ankles are afflicted with an ailment that can afflict my thighs. (*i*) and (*iii*) are inconsistent beliefs. On Twin Earth, Jack believes: (*iv*) My ankles are afflicted with—and only with—tharthritis; (*v*) Tharthritis is an ailment that can afflict my thighs; and (*vi*) Therefore, my ankles are afflicted with an ailment that can afflict my thighs. (*iv*), (*v*) and (*vi*) form a consistent set. But it is unclear how conceptual misunderstanding might lead Jack to judge that they form an *in*consistent set.

across premises. Jack no doubt assumes that he is using the term 'arthritis' to express the same concept—namely, the concept tharthritis—when he expresses (4) and (5). But it is not part of the objection from brute error to show that this assumption is false.

2.3 Two Responses Considered

In this subsection, I want to consider two possible responses to the objection from brute error. I am going to argue that neither response succeeds in blocking the objection. The first response involves challenging my earlier characterisation of Jack's beliefs on Twin Earth. According to my earlier characterisation, when on Twin Earth Jack says 'My ankles are afflicted with—and only with—arthritis' he expresses a belief involving the concept tharthritis. Someone might challenge this characterisation. They might claim that the belief which Jack uses this utterance to express will be a belief involving the concept arthritis, not the concept tharthritis—specifically, (1). Of course, if this claim is right, then the objection loses its force. If Jack does use the utterance 'My ankles are afflicted with—and only with—arthritis' on Twin Earth to express (1), then his beliefs are not inconsistent, for (1) and (6) are not inconsistent. So Jack is right when he judges that his beliefs are consistent and that he does not have a reason to give up either (1) or (6).

The second response I want to consider involves accepting my earlier characterisation of Jack's beliefs on Twin Earth. The proponent of this response agrees that Jack's beliefs are inconsistent and that Jack is in error when he judges that they are consistent. She denies, however, that the error in question is brute. Because Jack's error in judging that (4) and (6) are consistent is non-brute, so too, the response goes, is Jack's error in judging that he does not have a reason to give up either (4) or (6). According to the second response, then, Jack's case does not show that slow-switching can bring it about that a subject's judgments about his epistemic reasons are subject to brute error. So it does not show that slow-switching can undermine knowledgeability of a subject's epistemic reasons. I want to consider each of these responses in turn.

There are at least three ways in which someone might defend the view that when on Twin Earth Jack says 'My ankles are afflicted with—and only with—arthritis' he expresses (1). Judged according to the criteria for misunderstanding a concept as outlined in Chapter 1, Subsection 1.2, Jack misunderstands the concept tharthritis. As already made clear, in the case as described it is only because Jack misunderstands the concept tharthritis that he is wrong about his epistemic reasons. One way in which to defend the view that on Twin Earth Jack will express (1) instead of (4) is by rejecting *Misunderstanding*, the claim that a subject can have thoughts involving concepts which she misunderstands. In Chapter 3, Subsection 3.3 I will consider the merits of rejecting *Misunderstanding*. But note that whatever those happen to be, this is not a response which Burge can make without adjustments to his view. We know from our discussion in Chapter 1, Subsection 1.2 that Burge endorses *Misunderstanding*. Indeed, the arthritis case—one of the thought experiments which he offers in support of externalism's driving intuition—depends on it. If Burge rejects *Misunderstanding*, then he gives up the arthritis case. But if he gives up the arthritis case, then the scope of his externalism will no longer include subjects who both misunderstand a concept and defer to the relevant class of expert speakers.

Someone might accept *Misunderstanding* and yet deny that Jack will *acquire* the concept tharthritis during his stay on Twin Earth. This is a second way in which someone might defend the view that on Twin Earth, Jack expresses (1) in place of (4). The problem with this way of defending the view is that it is difficult to see what the rationale for denying that Jack will acquire the concept tharthritis might be. It is part of the thought experiment that Jack defers to the experts when he uses the term 'arthritis'. It is reasonable to assume that given enough time on Twin Earth, the experts to whose use Jack defers will come to be the experts on Twin Earth. If Jack does come to defer to the experts on Twin Earth, then he will have come to be dependent on Twin Earth communal standards. Given these considerations, and given Burge's own views about the considerations which determine how, if at all, slow-switching will affect a subject's mental states

(as outlined in Chapter 2, Section 1), we should expect that Jack *will* acquire the concept tharthritis during his stay on Twin Earth.[12]

There is a third way of defending the view that on Twin Earth, Jack will express (1) in place of (4). One might agree that Jack will acquire the concept tharthritis during his stay on Twin Earth, but deny that he will come to use the term 'arthritis' on Twin Earth to express this concept. But again, it is difficult to see what the rationale for this claim might be. If the utterance 'My ankles are afflicted with—and only with—arthritis' expressed a remembered state—if it involved the functioning of preservative memory—then the claim would have credence. But it does not. Whatever its exact content, the belief expressed by this utterance is an occurrent mental state. It is not clear what further reason one could have for thinking that Jack will use the term 'arthritis' on Twin Earth to express the concept arthritis, given that he acquires the concept tharthritis during his stay.

I conclude that my earlier characterisation of Jack's beliefs on Twin Earth is one which Burge should accept, given his views as they stand.

The second response I want to consider involves accepting that on Twin Earth, Alf uses the sentence 'My ankles are afflicted with—and only with—arthritis' to express (4) and that therefore Jack has inconsistent beliefs. But it involves denying that Jack's error in judging that he does not have inconsistent beliefs is brute. This is a response which we can expect Burge to make, for he writes:

[12] Jack's situation is essentially the reverse of Alf's situation in the arthritis case. In the arthritis case, we imagine Alf being switched from a context in which he uses the term 'arthritis' to express the concept arthritis, which he misunderstands, to a context in which (if Burge is right) he comes to use the term to express the concept tharthritis, which he does not misunderstand. Jack, in comparison, moves from a context in which he uses the term 'arthritis' to express the concept arthritis, which he understands, to a context in which he comes to use the term to express the concept tharthritis, which he misunderstands. Could Burge appeal to this point of difference to motivate the claim that Jack would not acquire the concept tharthritis after having spent time on Twin Earth? I do not think so. Burge acknowledges that some thought experiments in support of externalism's driving intuition may not be, as he puts it, 'symmetrical'. But it is clear that he does not think that the arthritis case is among them: 'It is worth remarking that the thought experiment as originally presented might be run in reverse' (Burge, 1979: 84).

> I intend malfunctions to cover … failures of normal understanding—as for example when an individual believes arthritis can occur in the thigh. (Burge, 1996: 103, fn. 7)

If failures of normal understanding are, or are always indicative of, a malfunction in the mistaken individual, then errors which derive from such failures are non-brute. Jack judges that (4) and (6) are consistent because he believes (5)—that tharthritis cannot occur in the thigh. If believing that arthritis can occur in the thigh constitutes a failure of normal understanding, then presumably so too does believing that tharthritis cannot occur in the thigh. So it looks like Jack's error in judging that (4) and (6) are consistent is non-brute, for it is traceable to a malfunction in Jack.

But *are* failures of normal understanding always, or always indicative of, malfunctions within the mistaken individual? It seems clear that the answer to this question is 'No'. Suppose that a subject comes to believe that arthritis can occur in the thigh because that is what she has been told by someone whom she reasonably believes is a medical professional. In this case, the subject is surely warranted in believing that arthritis can occur in the thigh. But if her misunderstanding was, or was indicative of, some malfunction, then she would not be warranted. We can imagine a whole range of cases of this sort, cases in which a subject has abnormal understanding but in which her abnormal beliefs are warranted. The existence of such cases seems to give the lie to the claim that failures of normal understanding always are, or involve, a malfunction in the mistaken individual.[13]

Might the scenario involving Jack be such a case? I think so. Jack's failure of normal understanding consists in his believing that tharthritis cannot occur in the thigh. But it seems clear

[13] It is not clear to me why Burge says what he does about malfunctions and failures of normal understanding. It seems at odds with his own conception of warrant. Burge writes, '*Epistemic warrants* derive from a psychology's meeting norms or standards that govern good routes for realising the representational function of belief formation—the function of forming true beliefs—or for realising the representational function—preservation of truth and warrant—of certain transitions (inferences) that serve true belief. The relevant standard is for operating *representationally well* cognitively, well enough to have the right to hold the belief, or make the transition, given relevant information and relevant cognitive resources' (2013c: 3). It seems to me that in the case where *S* believes that arthritis can occur in the thigh because that is what she has been told by someone whom she reasonably believes to be a medical professional, the relevant standard is met, even though the resultant belief is false. Believing that *p* on the basis of expert testimony (given that one has no reason not to believe that *p*) surely is an instance of 'operating representationally well cognitively'.

that the fact that he believes this need not indicate any malfunction in Jack. In the case as described, Jack believes that tharthritis cannot occur in one's thigh because of a change in the content of his mental states brought about by his being slowly switched. Might this change itself constitute a malfunction? Not on Burge's existing view. Burge acknowledges a presumption, on the part of subjects who are being slowly switched, that there has been no change in their existing concepts. In the course of discussing this presumption, Burge writes:

> What seems to me clear … is that the presumption is warranted, given the thinker's information. The thinker would be entitled to take instances of the twin concept [say tharthritis] to be instances of the home concept [arthritis]. The existence of the warranted "presumption" allows mistakes (the analogues of equivocation) in argumentation that derive from slow-switches to be assimilated more to empirical errors than to unreasonable inferences. (Burge, 2013c: 15)

Subjects would not be entitled to take instances of twin concepts to be instances of home concepts if changes in concepts brought about by slow-switching constituted malfunctions. So given what Burge says in this passage, changes brought about by slow-switching cannot constitute malfunctions.

I conclude that there is no good reason to think that Jack's error in judging that he does not have a reason to give up either (4) or (6) is non-brute.

3. EPISTEMIC REASONS, SLOW-SWITCHING AND BRUTE SUCCESS

3.1 Having a Deductive Reason

In this subsection, I want to consider a third line of response to the objection from brute error which is open to Burge. I will conclude that this line of response is successful in blocking the objection. The objection from brute error rests on the assumption that because (4) and (6) are inconsistent, Jack has a reason to judge that they are inconsistent. In this subsection, I am going to consider Burge's views about the circumstances under which a subject has a deductive reason to believe that *p*. The discussion will make it clear that given Burge's views, we can expect him to

reject this assumption. Given Burge's views, Jack does not have a reason to judge that (4) and (6) are inconsistent beliefs.

In 'Self-Understanding'—the third of three lectures published together as 'Self and Self-Understanding: The Dewey Lectures (2007, 2011)'—Burge considers the conditions under which a deductive inference provides *reason support* for a conclusion. The intuitive notion of a deductive inference's providing reason support for a conclusion just is the notion of one's having a reason to deductively infer the conclusion via the inference in question. So in setting out the conditions under which a deductive inference provides reason support for a conclusion, Burge takes himself to be setting out the conditions under which one has a reason to infer the conclusion deductively.[14] According to Burge:

> For an inferential transition to provide *reason support* for a conclusion by way of a deduction, the transition must be correctly explainable in terms of deductive inference rules and competence with logical constants; the premises of the inference must be warranted; and the reasoning must be relevantly non-circular. For a deductive reason-supporting transition to be correctly explainable in terms of deductive principles, its premise must ground some rational explanation of why the conclusion is worthy of belief *for the individual*. The individual need not be able to give the explanation. But the individual's competence with logical constants must rationalise—ground a potential explanation of—why the conclusion is belief worthy for the inferrer. (Burge, 2013e: 197)

I take it that these three conditions are both necessary *and jointly sufficient* for an inferential transition to provide reason support for a conclusion by way of a deduction, and so necessary and jointly sufficient for one's having a reason to deductively infer the conclusion via the transition. The second and third conditions—that the premises of the inference be warranted and that the reasoning be relevantly non-circular—are straightforward.[15] The first condition is a

[14] Of course, many transitions which are reason-supporting are not deductive, but probabilistic inferences. However, I am going to leave the question of the conditions under which a probabilistic inference is reason-supporting to one side (for Burge's views on the conditions under which an inductive transition is reason-supporting, see his 2013b: 493-494). The case we are interested in—Jack's case—is, after all, one in which the subject reasons deductively.

[15] The inclusion of the third condition—that the reasoning be relevantly non-circular—allows Burge's account to dodge forms of *the bootstrapping objection* (see, for example, Broome, 2013: 81-82). Suppose that I am warranted in believing *p* and if *p* then *p*. If there were no requirement that the reasoning be relevantly non-circular, then these beliefs would constitute a reason for me to believe *p* (for *p* can be deductively inferred from *p* and if *p* then *p*). The fact that I believe *p* would itself give me a reason to believe *p*. But this is an unacceptable conclusion. The fact that I believe something does not itself give me a reason to believe it, even if the belief in question is warranted. For brief, critical discussion of the bootstrapping objection as Broome presents it see Gibbons (2013: 32-33).

little harder to pin down. According to it, the inference 'must be correctly explainable in terms of deductive inference rules and competence with logical constants', where an inference is so explainable only if its premises ground 'some rational explanation of why the conclusion is worthy of belief *for the individual*'. Why is 'for the individual' italicised? The following answer is suggested by what Burge goes on to say. When considering whether a particular transition is reason-supporting, we should do so against the backdrop of the subject's existing logical competence. A transition may in fact be in accordance with a valid rule of inference. But if the subject's competence with logical constants is not such that she could rely on that rule of inference in making the transition, then the transition is not reason-supporting—the subject does not have a reason to make the transition. As Burge says, it is not necessary that the subject be able to elucidate the rule, to *explain* why the conclusion follows from the premises. But the subject must be competent to make the inference.

Burge, then, appears to endorse the following principle:

> ***Having Deductive Reason***: *S* has a reason to believe that *p* by way of a deduction if and only if there is some set of warranted propositional attitudes which is present in *S*'s psychology from which *p* can be deductively inferred by *S*.

Whether *S* can deductively infer that *p* from a set of propositional attitudes which is present in her psychology will depend on *S*'s logical competence in just the way outlined above.

As noted earlier, the objection from brute error rests on the assumption that because (4) and (6) are inconsistent beliefs, Jack has a reason to judge that they are inconsistent. Given that Burge endorses *Having Deductive Reason*, we can expect him to reject this assumption. If *Having Deductive Reason* is true, Jack does not have a reason to judge that (4) and (6) are inconsistent beliefs because, given that Jack's beliefs are as earlier characterised, there is no set of propositional

attitudes present in Jack's psychology from which the claim that (4) and (6) are inconsistent beliefs can be deductively inferred by Jack. Because Jack does not have a reason to judge that (4) and (6) are inconsistent, he does not have a reason to give up either (4) or (6). And because Jack does not have a reason to give up either (4) or (6), he does not judge incorrectly when he judges that he does not have a reason to give up either (4) or (6). So if *Having Deductive Reason* is true, then Jack's case does not demonstrate that slow-switching can undermine knowledgeability of one's epistemic reasons.

I said that given that Jack's beliefs are as earlier characterised, there is no set of propositional attitudes present in Jack's psychology from which the claim that (4) and (6) can be deductively inferred by Jack. This claim might seem questionable on the following grounds. If (4) and (6) are beliefs about tharthritis, then they are *necessarily* inconsistent. But if (4) and (6) are necessarily inconsistent, then (10) expresses a necessary truth:

(10) (4) and (6) are inconsistent

But a necessary truth can be deductively inferred from *any* set of premises, so Jack can deductively infer (10) regardless of whatever else he believes. In other words, there is always a set of propositional attitudes which is present in Jack's psychology from which (10) can be validly inferred. Consequently, there is a set of propositional attitudes present in Jack's psychology from which (10) can be validly inferred when Jack judges that (4) and (6) are not inconsistent beliefs. Assuming that Jack has the necessary logical competence—that he knows that a necessary truth can be deductively inferred from any set of premises—it looks like *Having Deductive Reason* commits us to saying that Jack does in fact have a reason to judge that (4) and (6) are inconsistent. Yet this conclusion seems to run contrary to the spirit of Burge's account.

It is clear that Jack is not in a position to explain why inferring (10) is deductively valid. This is not problematic per se for as we have already noted, it is not part of Burge's view that a subject must be able to explain why a particular conclusion follows deductively from a set of premises. But presumably it matters why the individual cannot provide this explanation. In the

cases Burge has in mind, the subject recognises that the conclusion follows deductively but cannot explain why it does because they lack a reflective understanding of the relevant inference rule. Jack, in comparison, is not even in a position to recognise that (10) follows deductively from his existing attitudes. He is not in a position to do this because, although he has the requisite competency, he does not believe that (10) expresses a necessary truth.

In light of these considerations, we ought to supplement our earlier characterisation of the conditions under which *S* is able to deductively infer that *p*. That characterisation focused exclusively on *S*'s logical competency. *S* is able to deductively infer that *p*, according to our earlier characterisation, if and only if *S*'s logical competency is such that she could rely on the relevant inference rule in making the transition. But we have come to see that it is not sufficient for *S*'s being able to deductively infer that *p* that *S*'s competency is such that she could rely on the relevant inference rule. *S* is able to deductively infer that *p* only if *S* is in a position to tell, on the basis of reflection alone, that *p* can be deductively inferred from her existing propositional attitudes.

3.2 The Objection from Brute Success

The response outlined in the previous subsection succeeds in blocking the objection from brute error. If *Having Deductive Reason* is true, then Jack does not have a reason to give up either (4) or (6). In this case, he judges correctly when he judges that he does not have a reason to give up either (4) or (6). So the objection from brute error does not succeed in showing that slow-switching can undermine knowledgeability of one's epistemic reasons, given Burge's views about the individuation of mental content. In this subsection, I am going to press a different objection, which I am going to call *the objection from brute success*. This objection shows that even if *Having Deductive Reason* is true, slow-switching can undermine knowledgeability of one's epistemic reasons, given Burge's views about the individuation of mental content.

The response to the objection from brute error outlined in Subsection 3.1 refutes the claim that in Jack's case, slow-switching brings it about that Jack's judgments about his epistemic reasons—about the rational relations between his mental states—are subject to brute error. But note that it does not refute, and is not intended to refute, the claim that slow-switching brings it about that Jack's judgments about the *logical* relations between his mental states are subject to brute error. If *Having Deductive Reason* is true, then Jack judges correctly when he judges that he does not have a reason to give up either (4) or (6). But even if *Having Deductive Reason* is true, it is still the case that he is incorrect when he judges that (4) and (6) are consistent beliefs. Insofar as (4) is a belief about tharthritis, (4) and (6) are *not* consistent beliefs.

I think that it is a problem for Burge that even if *Having Deductive Reason* is true, slow-switching can bring it about that one's judgments about the logical relations between one's mental states are subject to brute error. I think that because it can bring this about, slow-switching can bring it about that one's judgments about the rational relations between one's mental states are subject to brute success. Consider, for example, the following slow-switching scenario.

Suppose that some time prior to being switched, Jack reasons in the following way:

(11) The set of beliefs which I have about the ailment in my ankles is consistent

(12) If a set of beliefs is consistent, then I do not have a reason to think that the beliefs which belong to that set cannot all be true

(13) Therefore, I do not have a reason to think that the beliefs which I have about the ailment in my ankles cannot all be true

Let us assume that (11) and (12) are both warranted beliefs. Jack's reasoning is relevantly non-circular and in accordance with the rules for deductive inference. We can assume that Jack is competent to make the inference and that he relies on the relevant inferential rule in doing so. So Jack is warranted in believing (13). Given *Having Deductive Reason*, (13) is true; there is no set of

warranted beliefs which is present in Jack's psychology from which the claim that the beliefs which Jack has about the ailment in his ankles cannot all be true can be validly inferred by Jack. Given these considerations, it seems right to say that Jack *knows* that he does not have a reason to think that the beliefs which he has about the ailment in his ankles cannot all be true.

Now suppose that Jack is being slowly switched between Earth and Twin Earth. Suppose that during a stay back on Earth, Jack performs the exact same reasoning. Again, let us assume that both (11) and (12) are warranted, that Jack is competent to make the inference and that he relies on the relevant inferential rule in doing so. In this case, it does not seem right to say that (11) constitutes knowledge—that Jack *knows* that the set of beliefs which he has about the ailment in his ankles is consistent—even though it is both warranted and true. The reason it does not seem right is that, because Jack is being slowly switched, the possibility that the set of beliefs which Jack has about the ailment in his ankles is *inconsistent* is a relevant alternative (if Jack happened to be on Twin Earth, it would be inconsistent) and for all Jack knows on the basis of reflection alone, it *is* inconsistent. It seems that in this case, (11) is subject to brute success.

If Jack's warrant for (11) is insufficiently strong to constitute knowledge, then the same must be true of Jack's warrant for (13), Jack's judgment that he does not have a reason to think that the beliefs which he has about the ailment in his ankles cannot all be true. There are cases in which the strength of one's warrant for a belief can exceed the strength of the individual warrants for the premises from which the belief is inferred.[16] But it is difficult to see how that could be so in the present case. Given that Jack reasons to (13) by way of (11), it seems that the strength of Jack's warrant for (13) cannot exceed the strength of Jack's warrant for (11). If Jack's warrant for (11) is not sufficiently strong to constitute knowledge, then Jack's warrant for (13) cannot be sufficiently strong to constitute knowledge. But (13) is a judgment about the epistemic reasons which Jack has. So it looks like slow-switching has brought it about that Jack's judgments about

[16] Suppose, for example, that from: (*i*) Alex has brown hair; (*ii*) Sarah has brown hair; (*iii*) Kate has brown hair, I infer (*iv*) Someone has brown hair. It is plausible to think that the strength of my warrant for (*iv*) exceeds the strength of my individual warrants for (*i*), (*ii*) and (*iii*).

his epistemic reasons are subject to brute success. So it looks like slow-switching has undermined the knowledgeability of Jack's epistemic reasons.

An example may serve to highlight the basic point. Suppose that while driving through fake barn country, I reason as follows:

(14) This structure is a barn

(15) Barns are designed to house livestock and store crops

(16) Therefore, this structure is designed to house livestock and store crops

Even if (14) is true, it does not seem right to say that I *know* that this structure is a barn. It does not seem right because, given my circumstances, the possibility that this structure is not a barn is a relevant alternative and for all I know, it isn't a barn. If (14) is true, then (16) is true. But it does not seem right to say that I *know* that (16) is true. The reason it does not seem right is that I reason to (16) by way of (14), my warrant for which is insufficient to constitute knowledge.[17]

I said that in the case in which he is being slowly switched, Jack does not know on Earth that (11) is true because the possibility that (11) is false is a relevant alternative, and for all Jack knows on the basis of reflection alone, (11) *is* false. We know from our discussion in Chapter 2,

[17] Of course, there is a difference between this case and the case involving Jack. In this case, if (14) *were* false—if the structure were a facsimile—then (16) would also be false, even if I happened to believe that (14) and (15) were true. If this structure were in fact a facsimile, then even if I believed it to be a barn and even if I believed that barns are designed to house livestock and store crops, this structure would not be a structure designed to house livestock and store crops. But if (11) were false, but Jack still believed (11) and (12) to be true, (13) would still be true. If (11) were false, but Jack still believed (11) and (12) to be true, it would still be true that Jack does not have a reason to think that the beliefs he has about the ailment in his ankles cannot all be true. It would still be true because there would not be a set of warranted beliefs present in Jack's psychology from which the claim that the beliefs which he has about the ailment in his ankles cannot all be true could be validly inferred by Jack.
I grant that this is a point of difference between the two cases. But I do not think that it is a relevant difference. In cases (like Jack's) where one's reasoning is such that the strength of one's warrant for the conclusion cannot exceed the strength of one's warrants for the individual premises, the question whether the conclusion would have been false if certain of the premises had been false is irrelevant. In such cases, if one infers the conclusion by way of a premise one's warrant for which is insufficient to constitute knowledge, one's warrant for the conclusion is also insufficient to constitute knowledge *whether or not* the conclusion would have been false if the premises were false.

Section 2 that Burge denies that a subject must rule out relevant alternatives to her thinking that *p* before she can know that she is thinking that *p*. Couldn't Burge make an analogous move in the present case? Couldn't Burge simply deny that in order to know that the set of beliefs which he has about the ailment in his ankles is consistent, Jack must rule out the possibility that they are inconsistent? I do not think so. Slow-switching can bring it about that a subject, who was thinking that *p* on Earth, is thinking that *q* on Twin Earth. But for reasons made clear in Chapter 2, Section 2 it cannot bring it about that she judges that she is thinking that *p* when she is in fact thinking that *q* (or vice versa). In other words, slow-switching cannot undermine Jack's capacity to reliably judge whether he is thinking that *p* or that *q*. Things are different in the case of one's judgments about the consistency or inconsistency of certain of one's beliefs. Slow-switching can bring it about that a set of beliefs which is consistent on Earth, is inconsistent on Twin Earth. But it can also bring it about that a subject judges that the set is consistent when it is in fact inconsistent. Jack's case is a case in point. If the set of beliefs which Jack has about the ailment in his ankles was inconsistent—if Jack was on Twin Earth—he would still judge that it is consistent. In Jack's case, slow-switching has undermined his capacity to reliably judge whether the beliefs he has about the ailment in his ankles is consistent or inconsistent. For this reason, it seems right to think that Jack must rule out the possibility that the set of beliefs which he has about the ailment in his ankles is inconsistent before he can know that it is consistent.

If in the case where he is being slowly switched, Jack's capacity to reliably judge whether the set of beliefs which he has about the ailment in his ankles is consistent is undermined, then why think that on Earth, (11)—Jack's judgment that the set of beliefs which he has about the ailment in his ankles is consistent—is warranted? Why not think that the unreliability undermines his warrant for (11)? The unreliability would undermine his warrant for (11) if it were traceable to some malfunction or irrationality within Jack or to carelessness on Jack's part. But this is not so in the case as described. In the case as described, the unreliability is ultimately traceable to the fact that slow-switching has brought it about that Jack has thoughts involving the concept tharthritis, a concept which he misunderstands. For reasons given in Chapter 3, Subsection 2.3, it is not

plausible to think that this involves some malfunction or irrationality in Jack or indicates carelessness on Jack's part.

Nothing that I have said shows that (13) can be subject to brute error. Indeed, it is difficult to see how judgments like (13)—judgments about whether or not I have a reason to believe a particular proposition—*could* be subject to brute error, given that *Having Deductive Reason* is true. If I am reflecting carefully and reasoning well, then it seems that I cannot but be correct about whether there is a set of warranted propositional attitudes which is present in my psychology from which I can deductively infer a particular proposition. But if judgments about whether I have a reason to believe a particular proposition cannot be subject to brute error, then how could they be subject to brute success?

There is no in principle reason to think that the possibility of brute success entails the possibility of brute error. Instances of brute success are instances in which one's judgment is warranted but one's warrant is insufficiently strong for the judgment to constitute knowledge. One's warrant for a particular judgment can be insufficiently strong for the judgment to constitute knowledge even if that judgment cannot be false, given that the subject is reflecting carefully and reasoning well. Indeed, one's warrant for a particular judgment can be insufficiently strong to constitute knowledge even though one *is* reflecting carefully and reasoning well. One might reason to the judgment by way of a premise which although warranted, is not warranted in a way that suffices for knowledge.

An example which does not involve slow-switching may help to make the point. Suppose that I reason to the conclusion that Sarah will be happy via the following two premises: tomorrow will be sunny; if tomorrow will be sunny, then Sarah will be happy. Call this argument, *A*. It is plausible that the following judgment cannot be subject to brute error:

(17) *A* is valid

It seems that as long as I am reflecting carefully and reasoning well, I cannot be wrong about whether *A* is valid. But the judgment that *A* is valid can nevertheless be subject to brute success. Suppose, for example, that I reason to (17) by way of (18) and (19):

(18) It is not the case that the premises in *A* can be true and yet the conclusion false

(19) If it is not the case that the premises in *A* can be true and yet the conclusion false, then *A* is valid

Suppose that in addition to being true, (18) is warranted, but that my warrant is insufficient to constitute knowledge. Suppose, for example, that I am generally able to tell whether the premises in arguments can be true and yet their conclusions false, but that momentary irrationality interferes with this ability. In spite of the irrationality, I just happen to judge correctly. We might think that in this case, (18) is warranted, but not in a way which suffices for knowledge.[18] But if my warrant for (18) does not suffice for knowledge, then, given that I reason to (17) by way of (18), it seems that the same must be true of my warrant for (17). So it looks like this is a case in which the judgment that *A* is valid is subject to brute success. The judgment is both true and warranted but does not constitute knowledge.

It is true that (13)—Jack's judgment that he does not have a reason to think that the beliefs he has about the ailment in his ankles cannot all be true—cannot be subject to brute error. But it does not follow that (13) cannot be subject to brute success. It does not follow because, as we have seen, the possibility of brute success does not entail the possibility of brute error.

3.3 The Objection from Brute Success, *Lay Knowledge* and *False Belief*

[18] This case is meant to be analogous to the case of brute success described in Chapter 3, fn. 5. In that case, my visual system malfunctions, but not in a way that undermines its general reliability. My perceptual state happens to be veridical in spite of the malfunction. Recall that Burge wants to say that in that case, the malfunction undermines knowledge but not entitlement. Given the similarities between the perceptual case and the case described here, Burge should say the analogous thing about the present case.

If the objection from brute success is sound, then, given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of one's epistemic reasons, even if *Having Deductive Reason* is true.

The objection has purchase only insofar as slow-switching brings it about that Jack's judgments about the logical relations between his mental states are subject to brute error. As was made clear in Chapter 3, Subsection 2.2, in the case as described slow-switching brings it about that Jack's judgments about the logical relations between his mental states are subject to error by bringing it about that Jack has thoughts involving a concept which he misunderstands, namely, the concept tharthritis. Could Burge not respond to the objection from brute success, then, by simply rejecting *Misunderstanding*? If it is not possible for subjects to have thoughts involving concepts which they misunderstand, then in the case as described slow-switching would not bring it about that Jack has thoughts involving the concept tharthritis. And if it would not bring it about that Jack has thoughts involving the concept tharthritis, then there is no reason to think that it would bring it about that Jack's judgments about the logical relations between his mental states are subject to brute error.

As made clear in Chapter 1, Subsection 1.2, *Misunderstanding* is an assumption at work in the arthritis case, one of three thought experiments which Burge offers in support of externalism's driving intuition. If Burge were to reject *Misunderstanding*, then he would need to give up the arthritis case and the scope of his externalism would be compromised as a result. It would no longer include subjects who both misunderstand a concept and defer to the relevant class of expert speakers. But one might think that this narrowing of scope is a cost that Burge can bear. After all, Burge could reject *Misunderstanding* without rejecting either the water or sofa cases, two further thought experiments which he offers in support of externalism's driving intuition.

While they do not rely on *Misunderstanding*, recall that the water and sofa cases do rely, respectively, on two further claims—*Lay Knowledge*, the claim that a subject who is ignorant of certain normative characterisations associated with the word '*C*' can nevertheless have thoughts

involving the concept $\underline{C}$, and *False Belief*, the claim that if a proposition, *p*, is a meaning-giving characterisation associated with '*C*', then an otherwise competent subject who believes that *p* is false can still have thoughts involving the concept $\underline{C}$. Does the objection from brute success apply if we substitute *Misunderstanding* with either of these two further claims, given a Burgean framework for thinking about the individuation of mental states? If the answer is 'Yes', then clearly, rejecting *Misunderstanding* will not be sufficient to block the objection from brute success.

In the discussion that follows, I am going to defend two claims. The first claim is that it is not obvious that the objection from brute success applies if we substitute *Lay Knowledge* for *Misunderstanding*. We can imagine cases in which slow-switching brings it about that a subject has thoughts involving the concept $\underline{C}$ even though they are ignorant of certain normative characterisations associated with '*C*'. But these are not obviously cases in which slow-switching can bring about the sorts of errors in the subject's judgments about the logical relations between her mental states which I have shown that it can in cases where the subject misunderstands the relevant concept. The second claim I am going to defend is that whether the objection from brute success applies if we substitute *False Belief* for *Misunderstanding* depends on whether extensive, direct causal contact with newly encountered samples is sufficient to bring it about that a subject who possesses relevant discriminatory knowledge acquires a second, distinct concept.

Consider *Lay Knowledge*. Recall from our discussion in Chapter 1, Subsection 1.2 that in paradigmatic cases of misunderstanding, a subject is in error about the meaning-giving characterisations associated with the relevant term. In paradigmatic cases where a subject lacks expert knowledge, in comparison, the subject is merely ignorant of certain normative characterisations associated with the relevant term. We can imagine cases in which slow-switching brings it about that a subject comes to have thoughts involving the concept $\underline{C}$ even though she is ignorant of certain normative characterisations associated with '*C*'. But these do not seem to be cases in which slow-switching can bring about the sorts of errors in one's judgments about the logical relations between one's mental states which I have shown that it can in cases involving misunderstanding. The reason is that mere ignorance of normative characterisations does not seem

to generate the sorts of inconsistencies between one's mental states which are generated in cases involving misunderstanding.

Consider, for example, a slow-switching case involving Carl. On Earth, Carl has various thoughts involving the concept water. Suppose, for example, that he believes:

(20) This glass contains water

Nevertheless, Carl lacks expert knowledge about water. In particular, he is ignorant of water's molecular structure. It is not that Carl believes that the molecular structure of water is *not* H_2O. He simply does not have a view either way.

Suppose that Carl *did* believe:

(21) The molecular structure of water is H_2O

In this case, reflection on (20) and (21) might lead him to infer:

(22) This glass contains a liquid the molecular structure of which is H_2O

But Carl does not believe (21), so he is not going to infer (22).

Now imagine that Carl is slowly switched. After spending time on Twin Earth we can, for the reasons given in Chapter 2, Section 1, expect Burge to agree that Carl will acquire the concept twater. He will, however, lack expert knowledge about twater. Specifically, Carl will be ignorant of twater's molecular structure. So this slow-switching case involving Carl is an example of a case in which slow-switching brings it about that a subject has thoughts involving *C* (twater) even though he is ignorant of certain normative characterisations associated with '*C*' ('twater').

On Twin Earth, Carl believes:

(23) This glass contains *twater*

If Carl believed (21) on Earth, then on Twin Earth he would believe:

(24) The molecular structure of *twater* is H_2O

Reflection on (23) and (24) might then lead him to infer:

(25) This glass contains a liquid the molecular structure of which is H_2O

If Carl were to infer (25), then slow-switching would have generated the sort of inconsistency between Carl's beliefs which it generates in Jack's case (insofar as (23) is a belief about *twater*, (23) and (25) are inconsistent beliefs). But Carl does not believe (21) on Earth. Because he does not believe (21) on Earth, slow-switching is not going to bring it about that he believes (24) on Twin Earth. Because he does not believe (24) on Twin Earth, Carl is not going to infer (25). So slow-switching is not going to generate the sort of inconsistencies between Carl's beliefs which it generates in Jack's case. Consequently, there is no reason to think that it will bring it about that Carl's judgments about the logical relations between his mental states are subject to the sorts of errors to which Jack's judgments are subject.

Now let us consider *False Belief*, the claim that if a proposition, *p*, is a meaning-giving characterisation associated with '*C*', then an otherwise competent subject who believes that *p* is false can still have thoughts involving the concept *C*. Recall from the discussion in Chapter 1, Subsection 1.2 that what differentiates, on the one hand, otherwise competent subjects who have thoughts involving *C* in spite of their believing that *p*, which is a meaning-giving characterisation associated with *p*, is false, and on the other, subjects who misunderstand *C*, is that the latter (but not the former) think that the experts believe about *C*'s what they believe about *C*'s. Subjects of the former sort have a non-standard theory about *C*'s. They understand that the experts believe that *p* expresses a necessary truth about *C*'s. But they think that the experts are mistaken.

Can we imagine slow-switching cases in which an otherwise competent subject comes to have thoughts involving the concept *C* in spite of the fact that she has a non-standard theory about *C*'s? As I pointed out in Chapter 2, Section 1 it is unclear whether in slow-switching cases involving a subject who possesses relevant discriminatory knowledge and comes to have extensive,

direct causal contact with Twin Earth samples (but does not come to rely on Twin Earth communal standards), the subject's concept will remain constant, on Burge's considered view, or whether on that view, slow-switching will bring it about that the subject acquires a second, distinct concept. The answer to the question at the top of this paragraph will depend on what exactly Burge's considered view happens to be. Those subjects who could plausibly come to have thoughts involving the concept $\underline{C}$ in spite of the fact that they have a non-standard theory about *C*'s are typically going to be subjects who possess relevant discriminatory knowledge. If Burge's considered view is that the concepts of such subjects will remain constant in slow-switching cases of the relevant sort, then it is going to be difficult to imagine cases in which slow-switching brings it about that an otherwise competent subject has thoughts involving the concept $\underline{C}$, in spite of the fact that she has a non-standard theory about *C*'s. If, however, it is Burge's considered view that subjects who possess relevant discriminatory knowledge will acquire a second, distinct concept in slow-switching cases of the relevant sort, then I think we *can* imagine cases in which slow-switching brings it about that an otherwise competent subject has thoughts involving the concept $\underline{C}$ in spite of the fact that she has a non-standard theory about *C*'s.

For example, suppose that Jill, an inhabitant of Twin Earth, believes (incorrectly) that most competent speakers within her community think that safos are items of furniture meant primarily for sitting. Jill herself is convinced that safos are in fact religious artefacts. I take it that Burge would agree that on Twin Earth, Jill uses the term 'sofa' to express the concept $\underline{\text{safo}}$.[19] Now suppose that Jill is slowly switched between Twin Earth and Earth, where 'sofa' refers to sofas, objects which are (and are believed by the experts to be) items of furniture meant primarily for sitting. Suppose that Jill comes to have extensive, direct causal contact with sofas. How, if at all, will the switch affect Jill's 'sofa' concept? If it is Burge's considered view that subjects who possess relevant discriminatory knowledge will acquire a second, distinct concept in cases where

[19] Jill's situation is, after all, indiscernible from Twin Tom's situation in the sofa case (See Chapter 1, Subsection 1.2).

they come to have extensive, direct causal contact with newly encountered kinds, then we should say that on Earth Jill will use the term ‘sofa’ to express the concept sofa.

Now let us ask a further question. Given that it will result in her acquiring a second, distinct concept, can slow-switching bring about the sorts of errors in Jill’s judgments about the logical relations between her mental states which I have shown that it can in cases involving misunderstanding? I think that the answer to this question is ‘Yes’. Suppose that on Twin Earth, before her slow-switching ordeal begins, Jill reasons as follows:

(26) My most prized possession is a safo

(27) Safos are not items of furniture meant primarily for sitting

(28) Therefore, my most prized possession is not an item of furniture meant primarily for sitting

Insofar as (26) is a belief about safos, (26) and (28) are clearly consistent beliefs. Now suppose that Jill is slowly switched between Twin Earth and Earth. If a slow-switching case involving Jill will be disjoint type, then on Earth Jill believes:

(29) My most prized possession is a *sofa*

(30) *Sofas* are not items of furniture meant primarily for sitting

(31) Therefore, my most prized possession is not an item of furniture meant primarily for sitting

Insofar as (29) is a belief about sofas, (29) and (31) are inconsistent beliefs. So if it is Burge’s considered view that slow-switching will result in a subject like Jill acquiring a second, distinct concept, then slow-switching *can* bring about the sorts of errors in one’s judgments about the logical relations between one’s mental states which I have shown that it can in cases involving misunderstanding. Consequently, the objection from brute success has purchase against such a case.

The preceding discussion has not shown definitively that the objection from brute success does not apply if we substitute *Lay Knowledge* for *Misunderstanding*. But it has shown that if it *does* apply, it does not do so in the form that it does in cases involving misunderstanding. This is because in cases involving subjects who merely lack expert knowledge, slow-switching is not going to bring about the sorts of inconsistencies between the subjects' mental states which I have shown that it can in cases involving misunderstanding. Further, the discussion has shown that whether the objection from brute success applies if we substitute *False Belief* for *Misunderstanding* will depend on whether in slow-switching cases involving a subject who possesses relevant discriminatory knowledge and comes to have extensive, direct causal contact with Twin Earth samples (but does not come to rely on Twin Earth communal standards), the subject's concept will remain constant, on Burge's considered view, or whether on that view, slow-switching will bring it about that the subject acquires a second, distinct concept. If Burge's view is the latter, then the objection from brute success does apply, given this substitution.

To return to the question which prompted this discussion, whether rejecting *Misunderstanding* is sufficient to block the objection form brute success depends on two things: first, on whether there are sophisticated cases involving subjects who merely lack expert knowledge in which slow-switching can bring about the sorts of inconsistencies between the subjects' mental states which I have shown that it can in cases involving misunderstanding; and second, on Burge's considered view about whether slow-switching will affect the mental states of subjects who possesses relevant discriminatory knowledge and come to have extensive, direct causal contact with Twin Earth samples (but do not come to rely on Twin Earth communal standards).

CONCLUSION

In this chapter, I have argued that given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of one's epistemic reasons, even if

Having Deductive Reason is true. I have argued that this is a prima facie troubling result for Burge, given his views about the sorts of things that epistemic reasons are. On Burge's view, epistemic reasons are rational relations between mental states. Given this view, if slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of the rational relations between one's mental states. But it seems to be part of our intuitive picture of self-knowledge that the knowledgeability of the rational relations between one's mental states is not sensitive in this way to changes in one's context.

Chapter 4

WITTGENSTEIN, EXTERNALISM AND SLOW-SWITCHING

INTRODUCTION

The discussion in Chapters 1, 2 and 3 has focused on questions concerning the compatibility of the Burgean framework with guiding intuitions about self-knowledge. In this chapter, the focus of the discussion will shift to the Wittgensteinian framework. My aims in this chapter are essentially twofold: first, to show that Wittgenstein accepts the driving intuition about self-knowledge, that is, the claim that we normally know groundlessly what we are thinking; and second, to consider whether slow-switching objections have initial plausibility as objections to certain of the thought experiments which we considered in Chapter 1, Subsections 1.2 and 1.3, given a Wittgensteinian framework for thinking about the individuation of mental states.

I pursue the first aim in Section 1. Having established that Wittgenstein would accept the driving intuition about self-knowledge, I go on to consider how Wittgenstein might respond to the swift argument for inconsistency between externalism's driving intuition and the driving intuition about self-knowledge (as outlined at the beginning of Chapter 2, Section 1). I defend the view that Wittgenstein's response is, for all intents and purposes, the same as Burge's response. Like Burge, Wittgenstein rejects the assumption (on which the swift argument rests) that one knows that one is thinking that *p* only if one has confirmed that all the various enabling conditions for one's thinking that *p* obtain. I pursue the second aim in Sections 2 and 3. In Section 2, I consider whether slow-switching objections have initial plausibility as objections to the two thoughts experiments belonging to Wittgenstein which I discussed in Chapter 1, Subsection 1.3, given Wittgenstein's views about the way in which mental content is individuated. I conclude that they do not. In Section 3, I consider whether slow-switching objections have initial plausibility as objections to the arthritis case, given Wittgenstein's views. Again, my conclusion is that they do not. In the course of reaching this conclusion, I ask whether Wittgenstein accepts *Misunderstanding*, the claim that a subject can have thoughts involving concepts which she misunderstands. I argue that there is a strand in Wittgenstein's later thought which is at odds with this claim.

1. WITTGENSTEIN AND THE DRIVING INTUITION ABOUT SELF-KNOWLEDGE

Certain remarks of Wittgenstein's can seem suggestive of views about self-knowledge which are deeply counter-intuitive. Take, for instance, the following passage:

> I can know what someone else is thinking, not what I am thinking.
> It is correct to say "I know what you are thinking", and wrong to say "I know what I am thinking". (A whole cloud of philosophy condenses into a drop of grammar.) (PI, *Philosophy of Psychology—A Fragment*, hereafter PPF, §315)

It is correct to say 'I know what you are thinking' and wrong to say 'I know what I am thinking'. If this passage is encountered in isolation then one might think that the reason Wittgenstein regards it as wrong to say 'I know what I am thinking' is that he thinks that judgments of this sort—one's judgments about one's occurrent mental states—are routinely vulnerable to error. Interpreted in this way Wittgenstein's point is that it is wrong for me to claim to know what I am thinking because most of the time I do *not* know. But in fact, the opposite is Wittgenstein's view. Wittgenstein thinks that in most cases, there is something deeply suspect about the idea that one's judgments about one's occurrent mental states might be subject to error:

> If I ask someone "whom do you expect?" and after receiving the answer ask again "Are you sure that you don't expect someone else?" then, in most cases, this question would be regarded as absurd, and the answer will be something like "Surely, I must know whom I expect". (BB 21)

But if Wittgenstein regards it as in most cases absurd to think that someone might be wrong about whom they expect—or, by the same token, what they believe, desire, intend, and so on—then why does he think that it is incorrect to say 'I know what I am thinking'? The reason is that Wittgenstein has an idiosyncratic view concerning the conditions which need to obtain in order for us to be able to speak intelligibly about a person's knowing that *p*. Peter Hacker summarises this view nicely in the following passage:

> It makes sense to say of a person that he knows that such-and-such is the case only if it also makes *sense* to deny that he does … if there is no such thing as A's being ignorant of *p*, i.e. if it is unintelligible that *p* should be the case, yet A does *not* know it, then 'A knows that *p*' says nothing about A's knowledge. (Hacker, 1990: 57)

Knowledge, as Wittgenstein is thinking about it, is conceptually related to a battery of epistemic notions, including ignorance, doubt, error, and so on. If we cannot talk intelligibly about *S*'s being ignorant of (in doubt or error about) *p*, then we cannot talk intelligibly about *S*'s knowing that *p*, on

Wittgenstein's view. Since Wittgenstein thinks that in most cases we cannot talk intelligibly about *S*'s being ignorant of (in doubt or error about) her own mental states, he thinks that in most cases we cannot talk intelligibly about *S*'s knowing her own mental states. To summarise, the reason Wittgenstein thinks that it is incorrect to say, 'I know what I am thinking' is not that he thinks that usually we are wrong about what we are thinking; it is because he thinks that the conditions for speaking intelligibly about knowledge are not usually met in the case of judgments about one's occurrent mental states.[1]

Very few contemporary philosophers will be inclined to agree with Wittgenstein on this point. If we think of knowledge as most contemporary philosophers are inclined to think about it, as warranted, true belief, then it does not follow from the fact that in a particular case we cannot talk intelligibly about *S*'s being, say, ignorant about whether she is thinking that *p* that in that case it is incorrect for *S* to say 'I know that I am thinking that *p*'.[2] I do not intend to advocate for Wittgenstein's view here. The important point for present purposes is that Wittgenstein, like Burge, thinks that judgments about the content of our occurrent mental states are normally correct. We can agree with Wittgenstein on this point without buying into his views concerning the conditions requisite for speaking intelligibly about a person's knowing that *p*.

Putting aside these views, it is fair to describe Wittgenstein as endorsing the claim that we normally know what we are thinking. But the driving intuition about self-knowledge is not simply that we normally know what we are thinking, but that we normally do so *groundlessly*. Does Wittgenstein endorse this further claim? In the case of one's knowledge of one's mental images and sensations, Wittgenstein's views are clear enough:

---

[1] It is important to note that Wittgenstein's point is not fundamentally about the conditions under which it is intelligible or justified to ascribe knowledge of *p* to *S*; it is a point about the conditions under which *S knows* that *p*. Wittgenstein thinks that our being able to talk intelligibly about *S*'s knowing that *p* is a necessary condition for its being true that *S* knows that *p*. If such talk is unintelligible in a particular context, then in that context there is simply nothing that could count as *S*'s knowing that *p*, on Wittgenstein's view. Because the conditions for speaking intelligibly about knowledge are not usually met in the case of one's judgments about one's occurrent mental states, it is not usually true that one knows what one is thinking, on that view.

[2] Indeed, very few contemporary philosophers would be inclined to accept the further claim that in most cases, we cannot intelligibly talk about *S*'s being ignorant about whether she is thinking that *p*.

> What is the criterion for the sameness of two images?—What is the criterion for the redness of an image? For me, when it's someone else's image: what he says and does.—For myself, when it's my image: nothing. And what goes for "red" also goes for "same" (PI §377).

> What I do is not, of course, to identify my sensation by criteria: but to repeat an expression (PI, 2nd edition §290).

I do not know that my mental image is red or that my sensation is a sensation of pain on the basis of criteria, on Wittgenstein's view. Rather, I know groundlessly in both cases. There is no reason to think that Wittgenstein would not be prepared to say the same thing in the case of one's knowledge of one's mental states (indeed, *Zettel* §7, which I cite below, is confirmation of this in the case of intention). So I think it is clear that the answer to our question is 'Yes'. Wittgenstein does think that we normally know groundlessly what we are thinking. Leaving aside his idiosyncratic views about knowledge, it is fair to describe Wittgenstein as endorsing the driving intuition about self-knowledge.

But now we might press the swift argument for inconsistency. Our discussion in Chapter 1, Subsection 1.3 revealed that on Wittgenstein's view, whether or not *S* is thinking that *p* depends not on facts about what *S* is doing or what is going on within *S* at the time (where these facts are described in terms which do not take it for granted that *S is* thinking that *p*), but on facts about the broader context. But, we might ask, if *S*'s thinking that *p* depends on the broader context, how *could S* know groundlessly whether she is thinking that *p*? If the thought that *p* is partly individuated by some contextual fact, *R*, then *R*'s obtaining is a necessary condition for *S*'s thinking that *p*. If *R*'s obtaining is a necessary condition for *S*'s thinking that *p*, then *S* knows that she is thinking that *p* only if she has confirmed that *R* does in fact obtain. But if she must confirm that *R* does in fact obtain before she can know that she is thinking that *p*, then she cannot know groundlessly that she is thinking that *p*. If she judges that she is thinking that *p* without having first confirmed that *R* does in fact obtain, then her judgment that she is thinking that *p* is not warranted and consequently she does not know that she is thinking that *p*.

Recall that Burge's response to the swift argument for inconsistency involves rejecting the assumption (on which the argument rests) that one knows that one is thinking that *p* only if one has

confirmed that all the various enabling conditions for one's thinking that *p* obtains. There is evidence that Wittgenstein would respond in much the same way. I want to consider two passages in particular. Here is the first passage:

> If I have two friends with the same name and am writing one of them a letter, what does the fact that I am not writing it to the other consist in? In the content? But that might fit either. (I haven't yet written the address.) Well, the connexion might be in the antecedents. But in that case it may also be in what *follows* the writing. If someone asks me "Which of the two are you writing to?" and I answer him, do I infer the answer from the antecedents? Don't I give it almost as I say "I have toothache"? (*Zettel* §7)

According to Wittgenstein, what makes it the case that I am writing to one friend and not the other may include facts about what went on before I began writing the letter—for instance, the fact that earlier in the day I had remembered that I must reply to friend *A* (as opposed to remembering that I must reply to friend *B*). But as he goes on to say, I can (and normally do) know which friend it is that I am writing to without having to infer it from those facts. The point which Wittgenstein is making here is essentially the same as the point which Burge makes in response to the swift argument for inconsistency. On Wittgenstein's view, as on Burge's, one does not need to check that the enabling conditions for one's thinking that *p* obtain before one can know that one is thinking that *p*.[3]

A similar point is made in the second passage I want to consider. This passage is from §192 of *Remarks on the Philosophy of Psychology*, Volume 1. In an earlier remark (§172), Wittgenstein asks the reader to imagine people who only think out loud. He goes on in §192:

> One might want to make the following objection against the fiction about people who only think out loud: Suppose such a one were to say "As I left the house, I said to myself 'I must go to the baker.'" Couldn't he be asked "Did you really *mean* those words? For you might have said them as practice in elocution, or as a quotation, or as a joke, or in order to mislead someone." That is true. But was what he was doing a matter of the experience that accompanied the words? What speaks for such an

[3] Of course, if what makes it the case that I am writing to *A* and not *B* is indeed the fact that earlier I remembered that I must reply to friend *A* (as opposed to friend *B*), then it is likely to be the case that I *do* know that the enabling conditions for my intending to write to *A* obtain. In comparison, this is unlikely in the sorts of cases that Burge discusses. This difference does not, however, bring into question my basic point, which is that on Wittgenstein's view, as on Burge's, a subject does not *need* to know that the enabling conditions for her thinking that *p* obtain before she can know that she is thinking that *p*.

> assertion? Presumably, that the one who is asked may reply "I meant the sentence *like this*" without inferring this from external circumstances.

Now Wittgenstein does not in fact think that whether or not the person actually meant by the words 'I must go to the baker' that he must go to the baker is determined by whether or not he had an experience of a particular kind as he spoke the words. But in this passage he acknowledges that certain considerations can lend credence to such a view, namely, the fact that in general we do not infer what we meant by our words from external circumstances. As before, those external circumstances may play a role in making it the case that we meant what we did by our words. But we generally do not have to consult such facts in order to know what it is that we do mean.

2. WITTGENSTEIN'S EXTERNALISM AND SLOW-SWITCHING OBJECTIONS

As noted in Chapter 2, Section 1 there are certain circumstances in which we do require that subjects have checked that the enabling conditions for its being the case that *p* obtain before we describe them as knowing that *p*. Such cases are those in which there is some alternative to its being the case that *p*—its being the case that *q*, for example—which is relevant given the circumstances. Proponents of slow-switching objections take these considerations to show that if externalism's driving intuition is true, then in cases where *S*'s thinking that *q* is a relevant alternative to her thinking that *p*—cases in which *S* is being slowly switched, for example—*S* cannot know groundlessly that she is thinking that *p*.

In Chapter 2, Section 2 we considered Burge's response to slow-switching objections. Recall that this response culminates in the claim that slow-switching cannot undermine the knowledgeability of one's mental states. On Burge's view, slow-switching cannot bring it about either that one's judgments about one's mental states are subject to error or that those judgments, although true, do not constitute knowledge. The argument in support of this claim has two components. The first component involves defending a claim about the accuracy in slow-switching cases of one's judgments about one's mental states. The claim is that slow-switching

cannot bring it about that these judgments are subject to error. The second component involves defending a claim about the source of one's warrant for one's judgments about one's mental states. On Burge's view, this warrant is not a justification which is grounded in one's ability to identify one's thoughts or to discriminate them from relevant alternatives. Rather, it takes the form of an entitlement which derives from one's identity as a critical reasoner. On Burge's view, *Discrimination*—the claim that, where *S*'s thinking that *q* is a relevant alternative, *S* is warranted in judging on the basis of reflection alone that she is thinking that *p* only if she is able to discriminate on the basis of reflection alone between instances in which she is thinking that *p* and instances in which she is thinking that *q*—sets the bar for self-knowledge too high. A subject can be warranted in judging on the basis of reflection alone that she is thinking that *p*, Burge maintains, even though she lacks an ability to discriminate on the basis of reflection alone between instances in which she is thinking the thought she is actually thinking and instances in which she is thinking some relevant alternative thought.

In this section, I want to consider whether slow-switching objections have initial plausibility as objections to the two thought experiments belonging to Wittgenstein which I discussed in Chapter 1, Subsection 1.3.[4] Neither of the thought experiments, I noted, is addressed primarily to either the content of a subject's mental states or the concepts which she possesses. But I argued that given Wittgenstein's positive picture of the individuation of mental content, there is no principled reason why Wittgenstein does not offer thought experiments structurally identical to the arguments put forward in the passages I considered, but which are addressed primarily to mental content or concepts instead of mental attitudes and activities.

For instance, the following thought experiment—modelled on the thought experiment which he actually puts forward in *Investigations* §584—seems congenial to Wittgenstein's view. Suppose that Sally belongs to a community in which there is an extant practice of using the word

[4] My focus will be on whether slow-switching objections have initial plausibility as objections to the thought experiment taken from *Philosophical Investigations* §584. I will consider whether slow-switching objections have initial plausibility as objections to the thought experiment taken from *Remarks on the Foundations of Mathematics* VI, §35 only briefly.

'money' to express the concept money. Suppose further that Sally is a competent participant in this practice, that is, that she has undergone the appropriate training and uses the term 'money' regularly, confidently and more than often correctly. Now suppose that Twin Sally belongs to a Twin Earth community in which there is not an extant practice of using the word 'money' to express the concept money. Let us suppose that in Twin Sally's community, there is no institution of money. People simply trade goods and services. Suppose that in Twin Sally's community, the word 'money' refers to works of art which are indiscernible in appearance, feel, and so on, from bank notes and coins in Sally's community. Suppose that these works of art are among the things which are traded for goods and services. On Twin Earth, the term 'money' does not express the concept money, but some other concept—the concept twin money, say. Let us suppose that Twin Sally is a competent participant within her community's practice. She has undergone the appropriate training and she uses the term 'money' regularly, confidently and more than often correctly.

Obviously, there are likely to be significant differences between Sally and Twin Sally's respective histories and their non-intentionally described intrinsic properties over time. But this is not a problem for the thought experiment. As our discussion in Chapter 1, Subsection 1.3 made clear, unlike the arthritis, water and sofa cases, the thought experiments which Wittgenstein sets out in *Remarks on the Foundations of Mathematics* VI, §35 and *Investigations* §584, respectively, do not involve us imagining subjects who have led indistinguishable lives non-intentionally described. Rather, they involve us imagining subjects who for a given period of time (two minutes in the case of the first thought experiment and one minute in the case of the second) are engaged in the same behaviour and have the same things come before their mind (where these are described in terms which do not beg the question at issue). So let us suppose that for a period of two minutes, Sally and Twin Sally are engaged in the same behaviour and have the same things come before their mind (where this behaviour and these things are described in terms which do not take it for granted that Sally is having thoughts about money or that Twin Sally is having thoughts about twin

money). Suppose that during this period, both Sally and Twin Sally utter the words 'N.N. will come and bring me some money today'.

We can readily imagine a broader context which would prompt us to say that Sally uses these words to express the belief that N.N. will come and bring her some money today. We know that Sally is a competent speaker with respect to the term 'money'. Let us suppose that she uttered these words in the context of a conversation about money, perhaps in response to a question about the state of her finances. Suppose further that she was searching through her purse as she answered the question. Given this broader context, it seems right to say that Sally uses the words 'N.N. will come and bring me some money today' to express the belief that N.N. will come and bring her some money today. In comparison, we can readily imagine a broader context which would prompt us to say that Twin Sally uses these words to express the belief that N.N. will come and bring her some *twin money* today. We know that Twin Sally is a competent speaker with respect to the term 'money'. Perhaps Twin Sally uttered these words in the context of a conversation about tradeable goods, in response to a question about whether she had any such goods. Suppose that she was searching through the pouch in which she keeps such goods as she answered the question. Given this context, it seems right to say that Twin Sally uses the words 'N.N. will come and bring me some money today' to express the belief that N.N. will come and bring her some twin money today. The conclusion of the thought experiment is that two subjects, who for a period of two minutes have the same things come before their minds, who engage in the same verbal and non-verbal behaviour, where these are described in terms which do not beg the question at issue, might nevertheless differ with respect to the content of their mental states.

What might the internalist say about such a case? The internalist may well agree with this conclusion but argue that the difference in Sally and Twin Sally's respective mental states is nevertheless ultimately traceable to differences in their respective intrinsic properties. For example, Sally and Twin Sally are disposed to respond differently to questioning about what 'money' is or what it is that they mean by the term 'money'. It is this difference, the internalist

may insist, which ultimately accounts for the difference in Sally and Twin Sally's respective mental states.

In reply, Wittgenstein is likely to make two points. First, (as we shall see) Wittgenstein is likely to agree that the ways in which Sally and Twin Sally are disposed to respond to questioning about what they mean by 'money' play an important role in individuating their respective mental states. But as was made clear in Chapter 1, Subsection 1.3, it is only under an intentional description that a subject's dispositions play an individuating role, on Wittgenstein's view. And on that view, whether a particular disposition is correctly intentionally characterised in a given way will itself depend on facts about the broader context. So insofar as Sally and Twin Sally's dispositions to respond to questioning about what they mean by 'money' play an individuating role with respect to their mental states, those dispositions are *not* intrinsic properties of Sally and Twin Sally, according to Wittgenstein. Second, Wittgenstein is likely to deny that the difference between Sally and Twin Sally's mental states can be wholly explained in terms of a difference in their (intentionally described) dispositions. He is likely to insist that facts about their immediate context—for example, the fact that Sally spoke the words 'N.N. will come and bring me some money today' in response to a question about her finances, the fact that Twin Sally spoke these words in response to a question about tradeable goods in her possession, and so on—also play an important individuating role. These features of the immediate contexts are not intrinsic properties of Sally or Twin Sally, so they are not considerations which the internalist can identify as playing an individuating role.

We have been considering a thought experiment, modelled on the thought experiment that Wittgenstein actually puts forward in *Investigations* §584, but which is addressed primarily to mental states. Now let us imagine the following slow-switching scenario. Suppose that unawares, Sally is slowly switched between her actual community and Twin Sally's community on Twin Earth. Let us suppose that while on Twin Earth, Sally has extensive, direct causal contact with samples of twin money. Will slow-switching affect the content of the thought which Sally expresses when she says 'N.N. will come and bring me some money today'?

First, let us consider how Burge might respond to this question. In Chapter 2, Section 1 I considered what Burge might say about how, if at all, slow-switching will affect the content of subjects' mental states. I claimed that it is unclear whether Burge thinks that slow-switching cases in which a subject who possesses relevant discriminatory knowledge comes to have extensive, direct causal contact with Twin Earth samples will be cases in which the subject's concept remains constant or whether he thinks that in such cases, the subject will acquire a second, distinct concept. The case under consideration is a case of just this sort. We can expect that Sally possesses knowledge which could enable her to discriminate between samples of money and samples of twin money.[5] Consequently, it is unclear how Burge will respond to our question. Whether Burge thinks that slow-switching will affect the content of the thought which Sally expresses when on Twin Earth she says 'N.N. will come and bring me some money today' will depend on his views about such cases more generally. If he thinks that in such cases the subject will acquire a second, distinct concept, then he will answer 'Yes'. His view will be that when on Twin Earth Sally says 'N.N. will come and bring me some money today', she expresses the belief that N.N. will come and bring her some *twin money* today. If, however, he thinks that in such cases the subject's concept will remain constant, then he will answer 'No'. His view will be that on Twin Earth, as on Earth, Sally uses the sentence 'N.N. will come and bring me some money today' to express the belief that N.N. will come and bring her some *money* today.

Will slow-switching affect the content of the thought which Sally expresses when on Twin Earth she says 'N.N. will come and bring me some money today' on Wittgenstein's view? I think the answer to this question is 'No'. According to Wittgenstein's positive picture of the individuation of mental states, whether a subject, say, believes that *p* is not determined by what the subject is doing or what is going on within the subject at the time (where these are described in terms which do not take it for granted that the subject believes that *p*); it depends, Wittgenstein thinks, on facts about the broader context. These facts include facts about the subject's abilities.

[5] By way of a reminder, I say '*could* enable' because the subject may not be in a position to discriminate the samples, even though she possesses knowledge of the appropriate kind. She may not, for example, know how to carry out the relevant tests.

Wittgenstein is likely to identify the fact that Sally possesses knowledge which could enable her to discriminate between samples of money and samples of twin money as a reason to think that when on Twin Earth she says 'N.N. will come and bring me some money today', she expresses the thought that N.N. will come and bring her some *money* today. But I think that there are further facts about Sally upon which Wittgenstein is likely to place even more weight.

We know from our discussion in Chapter 1, Subsection 1.3 that on Wittgenstein's view, facts about the way in which a subject is disposed to respond to questioning about what she thinks, intends, understands, and so on, play an individuating role with respect to what it is that she does think, intend or understand. I noted that a certain class of dispositional facts are of special importance to Wittgenstein, namely, facts about how a subject is disposed to express the relevant state in words. The fact that if we were to ask *S* what she is thinking, she would tell us that she is thinking about how much colder it is this year than last, is a criterion for her actually thinking about how much colder it is this year than last. In cases like Sally's, however—cases in which we want to know what a subject *means*—the content of the subject's state remains unclear despite her already having expressed her state in words. In cases of this sort, the special importance lies with facts about how the subject is disposed to respond to questioning about what they mean by their words (Budd, 1989; Glock & Preston, 1995). Suppose that *S* remarks 'I hope it will stop soon', but it is unclear from her remark whether she means the music coming from the other room or the pain in her leg. The fact that were we to ask her which she means, *S* would say that she means the music coming from the other room is a criterion for her having meant that she hoped the music coming from the other room will stop soon. Wittgenstein makes the analogous point about meaning one person as opposed to another when he writes:

> What makes this utterance into an utterance about *him*?—Nothing in it or simultaneous with it ('behind it'). If you want to know whom he meant, ask him! (PPF ii, §17)

To be clear, Wittgenstein is not advancing a semantic claim (Budd, 1989). He is not suggesting that statements like 'I meant that *p*' *really mean* 'If you had asked me what I meant, I would have told you I meant that *p*'. Nor is he advancing a metaphysical claim. He is not

suggesting that to mean that *p just is* to be disposed to respond to questioning about what one meant by saying that one meant that *p*. On Wittgenstein's view, the claim that I have such-and-such a disposition is a hypothesis (BB 142), something which can only be confirmed (or disconfirmed) by empirical testing. But it is not a hypothesis that when I said 'I hope it will stop soon' I meant that I hope that the music in the other room will stop soon. I do not need to wait on empirical testing in order to know that that was what I meant.[6]

It is true that the fact that I am disposed to respond to questioning about what I meant by saying that I meant that *p* is a criterion for my having meant that *p*. But the fact that I am disposed to give that response does not *entail* that I meant that *p*. Nor does the fact that I am disposed to respond by saying that I meant, say, that *q* entail that I did *not* mean that *p*. According to the received view of Wittgenstein's notion of a criterion, a criterion for a particular mental state is neither a necessary nor sufficient condition for having that state. Like empirical evidence, a criterion constitutes defeasible grounds for judging that the state in question obtains.[7] Unlike empirical evidence, it is a necessary as opposed to a contingent truth—it belongs to the *grammar* of our psychological concepts, as Wittgenstein would put it—that the behaviour in question constitutes good grounds for judging that the relevant state obtains.

The fact that *S* is disposed to respond to questioning about what she meant by saying, for instance, that when she said 'I hope it will stop soon' she meant that she hoped the music in the other room will stop soon, is not the only criterion for her having meant this, on Wittgenstein's view. On that view, further criteria might include the fact that she covered her ears when she spoke the words, the fact that earlier she had complained about the music in the other room, and so on. In paradigmatic cases, we can expect facts about how a subject is disposed to respond to

[6] Indeed, there is a further reason why we should not say that to mean that *p just is* to be disposed to respond to questioning about what one means by saying that one meant that *p*. As Saul Kripke makes clear, 'the relation of meaning and intention to future action is *normative*, not *descriptive*' (Kripke, 1982: 37). What follows from the fact that I meant that I hope that the music in the other room will stop soon is not that I *will* go on to act in such-and-such a way, but that if I intend to accord with my past meaning then I *ought to*.

[7] For an alternative reading according to which criteria do *not* constitute defeasible grounds, on Wittgenstein's view, see McDowell (1982). For an overview of the debate, see Witherspoon (2011).

questioning about her meaning to mesh with facts about her abilities, facts about what she did when she spoke the words, facts about what she said or did earlier, and so on; in such cases, we can expect that all these various facts about the subject will speak in favour of our describing her as having meant, say, that she hoped the music in the other room will stop soon. But we can imagine cases in which these criteria might come into conflict.

I am not going to attempt an overview of all the various ways in which they might come into conflict, but let us consider one relatively straightforward case. Suppose that a subject is disposed to respond to questioning about what she meant when she said 'I hope it will stop soon' by saying that she meant that she hoped the music in the other room will stop soon, but that what she did when, or perhaps before, she spoke the words suggests otherwise. Suppose, for example, that as she spoke the words, she clutched her leg, or that earlier she had complained about the pain in her leg. I take it that this is a case in which we should say that there are considerations which at the very least weigh against the fact that she is disposed to respond to questioning by saying that she meant that she hopes that the music in the other room will stop soon.

The thought that facts about the way in which a subject is disposed to respond to questioning about what she means may play a role in determining what it is that she does mean, supports the claim that when on Twin Earth Sally says 'N.N. will come and bring me some money today' she expresses the belief that N.N. will come and bring her some money today (and not the belief that N.N. will come and bring her some twin money today). Suppose that on Twin Earth, we were to ask Sally what she means by 'money'. How might she respond? It is likely that she will respond by saying that she means currency or that she means notes and coins. If facts about the way in which a subject is disposed to respond to questioning about what she means play a role in determining what it is that she does mean, then the fact that Sally would respond in this way to our questioning about her meaning is a criterion for her having meant that N.N. will come and bring her some money today. To reiterate a point made earlier, from a Wittgensteinian perspective, this thought—the thought that the way in which Sally is disposed to respond to questioning about her meaning plays an individuating role with respect to her mental states—is distinctly externalist.

According to Wittgenstein, it is under an intentional description that dispositions play an individuating role. It is the fact that Sally is disposed to respond to questioning by *saying* that she means currency which play a role in individuating her mental states, on Wittgenstein's view. And on that view, whether Sally's disposition is correctly intentionally characterised in this way will itself depend on facts about the broader context.

I do not deny that there is likely to be *some* differences between the way in which Sally is disposed on Earth and Twin Earth, respectively, to explain what she means by 'money'. Defining a term ostensively is, on Wittgenstein's view, one perfectly good way of explaining what one means by that term. We can expect there to be a difference between the way Sally is disposed to ostensively define the term 'money' on Earth and the way she is disposed to ostensively define the term on Twin Earth. On Earth, we can expect Sally to be disposed to say that by 'money' she means *this* sort of thing, while pointing to a sample of *money*. On Twin Earth, we can expect Sally to be disposed to say that by 'money' she means *this* sort of thing, while pointing to a sample of *twin money*.

Wittgenstein would surely acknowledge that there is likely to be this difference, but I do not think that he would deem it sufficient for a difference between the concept which Sally uses the term 'money' to express on Earth and the concept which she uses it to express on Twin Earth. Consider a similar case. Suppose that a subject who is competent with respect to the term 'coin' is disposed to ostensively define the term 'coin' by pointing to actual coins. Now suppose that her situation changes so that she is disposed to ostensively define the term 'coin' by pointing, not to actual coins, but to counterfeit coins. Suppose, however, that the subject is ignorant of this change (perhaps thieves have replaced her collection of genuine coins with a collection of look-alikes). Should we say that the subject now uses the term 'coin' to express a concept the extension of which includes coins *and* counterfeit coins? I think that it is clear that Wittgenstein's answer would be 'No'. As a competent speaker, we can assume that the subject possesses knowledge which could enable her to discriminate genuine coins from look-alikes. Moreover, if she were made aware of the change in the way in which she is disposed to ostensively define the term 'coin',

she would likely disavow those explanations of what she means by 'coin' which involved her pointing to counterfeit coins. I think that given these considerations, Wittgenstein would say that the subject uses the term 'coin' to express the concept coin both before and after the change in the way they are disposed to ostensively define the term. Given the similarities between this case and the case involving Sally, I think that we can conclude that Wittgenstein would say the analogous thing in Sally's case.[8]

I said that the fact that Sally is disposed to respond to questioning by saying that by 'money' she means currency or that she means notes and coins is a criterion for her having meant that N.N. will come and bring her some money today. But it does not follow automatically from the fact that Sally is disposed to respond to questioning in this way that this *is* what she meant. If facts about Sally's abilities, about the things Sally did when she spoke the words, about the things she said or did earlier, and so on, suggest that Sally means something different by the term 'money', then the fact that she is disposed to respond to questioning about her meaning in the way that she is may not be decisive in determining what it is that she does mean. But nothing about Sally's abilities, about what she did when she spoke the words 'N.N will come and bring me some money today' or about the things that she said or did earlier, speak in favour of her meaning something other than money by 'money'. They certainly do not speak in favour of her using the term 'money' to mean twin money.

In summary, Wittgenstein is likely to take the view that slow-switching will not affect the contents of Sally's mental states. Wittgenstein is likely to say that on Twin Earth, as on Earth, Sally uses the utterance 'N.N will come and bring me some money today' to express the thought that N.N. will come and bring her some money today. In defending this response, Wittgenstein is

[8] In the case under consideration, the subject possesses knowledge which could enable her to discriminate between genuine coins and counterfeits. What might Wittgenstein say about analogous cases in which the subject does *not* possess such knowledge? Suppose, for example, that a subject's collection of aluminium is replaced by samples of twaluminium. Suppose further that the subject lacks knowledge which could enable her to discriminate samples of aluminium from samples of twaluminium. I think that Wittgenstein is unlikely to say that in cases of this sort, a change in the way in which the subject is disposed to ostensively define the term 'aluminium' will be sufficient to bring about a change in the concept which she uses the term to express.

likely to emphasise facts about the way in which Sally is disposed to respond to questioning about what she means by 'money'.

In reaching the conclusion that Wittgenstein would deny that the fact that she is being slowly switched will affect the content of Sally's thoughts, I am in apparent disagreement with Philip Pettit (1983). Pettit considers a slow-switching case not entirely dissimilar from the one involving Sally. In Pettit's slow-switching case, a competent subject is switched from a community in which 'plus' means plus to a Twin Earth community in which it means something subtly different, where 'the difference shows up only when the arguments to the plus function are both over a million' (Pettit, 1983: 452). Pettit and I agree that the subject who is being slowly switched will not, on Wittgenstein's view, come to have mental states involving the Twin Earth concept: 'in relation to such contents the person would have to be said to lack proper grasp' (Pettit, 1983: 453). But we appear to disagree in that I think that Wittgenstein will say that after the switch, the subject will continue to have thoughts involving the concept plus. According to Pettit, it is likely that Wittgenstein will say that the subject 'does not have beliefs with determinate contents, but is simply confused' (Pettit, 1983: 453).

I describe my disagreement with Pettit as apparent because I think our disagreement is really about the nature of slow-switching cases. Before he considers Wittgenstein's response to the 'plus' case, Pettit draws attention to the important distinction between being *located* in a particular socio-linguistic community and being *positioned* in that community. To be located in a particular socio-linguistic community is just to be physically present in that community. To be positioned in a socio-linguistic community is to be accountable to the community's rules and conventions; it is for the community's practices to function as a 'frame of reference for one's behaviour' (Pettit, 1983: 452). Given these characterisations, a subject can be located in a particular community without being positioned in that community, and vice versa.

Pettit thinks that:

> As we consider the thought experiment in connection with … Wittgenstein, we must be careful to see that it is positioning, and not just location, that we are varying. (Pettit, 1983: 452)

Why must we? On Pettit's view, it is an assumption in slow-switching cases that a subject's positioning varies as their location varies:

> In Burge's experiment we imagine the subject located, now here, now there, but at the same time we imaginatively position him now here, now there. What it is important to recognise however is that we might have fixed positioning independently of location and that had we done so we would have had a different result. We might have imagined the subject located in the counterfactual society but imaginatively positioned him in ours: this is reasonable so long as the difference between the contexts has not registered on him. If we had done this, then it would have been natural to continue to credit him with arthritis-beliefs rather than ascribing beliefs involving a concept other than that of arthritis. (Pettit, 1983: 452)

I think that Pettit is conflating Burge's original thought experiment for externalism's driving intuition with the various slow-switching cases which have been formulated as responses to that experiment. Burge's original thought experiment (summarised in Chapter 1, subsection 1.2) does not involve us imagining Alf located now here, now there (now on Earth, now on Twin Earth). It involves us imagining Alf, located on Earth, and a non-intentionally described twin (or Alf under counterfactual conditions) located on Twin Earth. Quite apart from that, however, I think Pettit is wrong to claim that it is an assumption in slow-switching cases (or indeed in Burge's original thought experiment) that we must position (in Pettit's sense) the subject in the community in which she happens to be located. Rather, this is a question which is left open when such cases are posed. I think that if Pettit were to agree with me about the nature of slow-switching cases, then our views about what Wittgenstein would say concerning the mental states of a subject who is being slowly switched may well be more closely aligned.

We have been considering a slow-switching scenario modelled on the thought experiment which Wittgenstein sets out in *Investigations* §584. I have argued that given Wittgenstein's views about the individuation of mental content, slow-switching will not affect the mental states of the subject in the scenario as described. Given that this scenario is modelled on the thought experiment in *Investigations* §584, we can conclude that slow-switching objections do not have initial plausibility as objections to the thought experiment which Wittgenstein sets out in

Investigations §584. Such objections rely on slow-switching bringing about some change in the content of a subject's mental states. If slow-switching does not bring about a change, then it does not introduce relevant alternatives and so cannot so much as threaten to undermine the knowledgeability of those states.

In Chapter 1, Subsection 1.3 I suggested that the thought experiment which Wittgenstein sets out in *Remarks on the Foundations of Mathematics* VI, §35 also supports externalism's driving intuition. I am not going to consider in detail a slow-switching scenario modelled on this thought experiment. But I think that Wittgenstein will deny that in such a scenario, slow-switching will affect the subject's mental states or the activity in which he is engaged. I think he will do so for the same sorts of reasons that he is likely to deny that slow-switching will affect the content of Sally's mental states. The thought experiment set out in *Remarks on the Foundations of Mathematics* VI, §35 involved us imagining a mathematician in England and a twin situated in some other context but who, for a period of two minutes, has the same events, processes and states go on within him as go on within the mathematician (where these are described in term which do not take it for granted that either is calculating). Wittgenstein's suggestion was that (while we can easily imagine a past and a continuation of the two minutes which would prompt us to describe what the mathematician is doing as calculating) we can imagine a past and a continuation of the two minutes which would prompt us to describe what the two-minute man is doing as something other than calculating.

Now, suppose that the mathematician in England is slowly switched between England and this twin context. Is the mathematician calculating in this twin context? The mathematician is disposed to do the same things when he is 'calculating' in this twin context as he is disposed to do on Earth. He is also disposed to *explain* what he is doing when he is 'calculating' in this twin context in exactly the same way as he is disposed to explain what he is doing when he is calculating on Earth. Moreover, insofar as he is a mathematician, the subject presumably possesses knowledge which could enable him to discriminate the activity of calculating from whatever activity it is in which the subjects in this twin context engage. In light of these

considerations, I think that Wittgenstein's answer to our question will be 'Yes'. What the mathematician is doing when he engages in the relevant behaviour in this twin context is calculating.

If this is right, then slow-switching objections do not have initial plausibility as objections either to the thought experiment in *Investigations* §584 or the thought experiment in *Remarks on the Foundations of Mathematics* VI, §35. Such objections rely on slow-switching bringing about some change in the content of a subject's mental states. But slow-switching will not do so in scenarios modelled on either of these thought experiments, given Wittgenstein's views about the individuation of mental states.

3. WITTGENSTEIN, SLOW-SWITCHING AND *MISUNDERSTANDING*

In the thought experiment for externalism which I sketched in the previous section, it was assumed that Sally was a competent speaker with respect to the term 'money'. In Chapter 1, Subsection 1.2 we considered three thought experiments which Burge offers in support of externalism's driving intuition. One of these thought experiments—the arthritis case—involved a subject who is not competent with respect to the relevant term. Alf, the subject in the arthritis case, misunderstands the concept arthritis. What, if anything, might Wittgenstein say about slow-switching cases involving subjects like Alf?[9]

In order to answer this question, we first need to consider what, if anything, Wittgenstein might say about the mental content of subjects like Alf on Earth before they are switched. To do

[9] Because of constraints on space, I am not going to consider what Wittgenstein might say about a slow-switching case involving Carl, the subject in Burge's water case. But in the context of the present discussion, what Wittgenstein might say about such a case is of less interest than what he might say about a slow-switching case involving Alf. Ultimately, I am interested in whether the objection from brute success arises, given a Wittgensteinian framework for thinking about the individuation of mental states. The objection as formulated in Chapter 3, Subsection 3.2 relies on *Misunderstanding*, the principle at work in the arthritis case. If the conclusions I reached in Chapter 3, Subsection 3.3 are sound, then it is not obvious that the objection applies if we substitute *Misunderstanding* with *Lay Knowledge*, the principle at work in the water case.

this, we need to clarify Wittgenstein's views regarding misunderstanding. Specifically, we want to know whether a subject can have thoughts involving concepts which she misunderstands, on Wittgenstein's view. We can approach this question by considering Wittgenstein's views on the nature of understanding. In Chapter 1, Subsection 1.3 we reached some conclusions about the sorts of things which may make it the case that *S* understands that *p*, on Wittgenstein's view. These include facts about *S*'s immediate context, facts about *S*'s abilities, about the things *S* said or did, and so on. But let us ask a different question: What *is* it to understand that *p*, according to Wittgenstein? On Wittgenstein's view, understanding is not a mental event, process or state. Rather, it is akin to an ability (PI §150). To understand the meaning of a word, for example, is to be able to use that word correctly:

> "Understanding a word" may mean: *knowing* how it is used; *being able to* apply it. (PG 47)

Why does Wittgenstein say '*may* mean'? This passage bears an obvious resemblance to *Investigations* §43:

> For a large class of cases of the employment of the word "meaning"—though not for *all*—this word can be explained this way: the meaning of a word is its use in the language.

Here too there is a qualification: '…though not for *all*…' Paul Horwich argues, convincingly I think, that 'the restriction is to a certain sense of "meaning", rather than to a certain subset of words' (Horwich, 2010: 117). The word 'meaning' is ambiguous. It could mean, for example, referent, intended interpretation or implicature. But it could also mean the literal semantic meaning of a word within a language. According to Horwich, 'It is in this [latter] sense—which [Wittgenstein] takes to be the most fundamental of them—that the meaning of a word, *every* word, is its use' (Horwich, 2010: 117). The term 'understanding' is also ambiguous. Not every instance in which we describe a subject as understanding a word is one in which we mean that the subject is able to apply the word correctly. For example, we sometimes mean that the subject knows why a particular word was used (instead of some other), that is, that she understands the speaker's intention in uttering it. Understanding a word in this sense is akin to, say, understanding a poem

(PI §531). In the passage from *Philosophical Grammar*, as in *Investigations* §43, the restriction is to a certain sense of 'understanding', rather than to a certain subset of words.

To understand the meaning of a word is to be able to use that word correctly, on Wittgenstein's view. What does Wittgenstein think it is to understand a concept? As Glock makes clear, in his later work Wittgenstein is thinking about concept possession in much the same way as he is thinking about grasp of linguistic meaning (Glock, 2010: 100). So for Wittgenstein, to grasp the concept *C* is, like grasping the meaning of the word '*C*', to be able to use '*C*' correctly. It is to be able to participate in particular language games: 'I say: the person who cannot play this game does not have this concept' (*Remarks on Colour* §115).[10]

What I have said thus far may lead one to think that Wittgenstein offers a positive account of understanding a concept or the meaning of a word, where a positive account is one which includes some non-circular explication of the notion of understanding a concept or the notion of understanding the meaning of a word. This is not in fact the case, as consideration of Wittgenstein's conception of language will show. On Wittgenstein's view, language is a rule-governed activity; each word, expression or proposition in the language is associated with rules which govern its use. They govern its use in the sense that they stipulate how the word, expression or proposition *ought* to be used. These rules are expressed by what Wittgenstein calls *grammatical propositions*. An example of a grammatical proposition is 'Red is a colour'. In most contexts, Wittgenstein thinks, this proposition expresses a rule for the use of the term 'red'; it 'instructs' one to use the term 'red' as one would the name of a colour.

We have noted that on Wittgenstein's view, to understand the meaning of the word '*C*' is to be able to use '*C*' correctly. For example, to understand the meaning of the word 'red' is to be able to use the word 'red' correctly. If language is rule-governed, then to be able to use 'red' correctly is, presumably, to grasp those rules which govern the use of 'red'. So if to understand the meaning of 'red' is to be able to use 'red' correctly and language is rule-governed, then to

[10] See also PI §384: 'You learned the *concept* 'pain' in learning language'.

understand the meaning of 'red' is to grasp these rules. But it is plausible that to grasp the rules governing the use of a word just is to grasp that word's meaning, on Wittgenstein's view. If this is right, then the claim that to understand the meaning of the word 'red' is to be able to use the word correctly amounts to the claim that to understand the meaning of the word 'red' is to grasp its meaning. Clearly, this latter claim is circular, for the notion of understanding the meaning of a word and the notion of grasping the meaning of a word are equivalent.

A similar point holds true for Wittgenstein's views about understanding a concept. On Wittgenstein's view, to understand the concept *C* is, like understanding the meaning of the word '*C*', to be able to use '*C*' correctly. For example, to understand the concept red is to be able to use the word 'red' correctly. So if to be able to use 'red' correctly is to grasp those rules which govern the use of 'red', then to understand the concept red is to grasp those rules. But given that we are thinking of concept possession on the model of grasp of linguistic meaning, it is plausible that to grasp the rules governing the use of 'red' just is to grasp the concept red. Consequently, the claim that to understand the concept red is to be able to use 'red' correctly amounts to the claim that to understand the concept red is to grasp the concept red, which is obviously circular. Wittgenstein has something to say about what it is to understand a concept or the meaning of a word. But what he has to say does not amount to a positive account in the relevant sense.

Our focus thus far has been on Wittgenstein's conception of understanding. What is it to *mis*understand the meaning of a word, on Wittgenstein's view? If understanding the meaning of a word is to be able to use it correctly, then presumably someone who misunderstands the meaning of a word is someone who is *not* able to use it correctly. But not being able to use a word correctly cannot be a sufficient condition for misunderstanding its meaning, for we want to retain a distinction between misunderstanding the meaning of a word and, for instance, being ignorant of what a word means. If not being able to use a word correctly was sufficient for misunderstanding its meaning, then someone to whom a word meant absolutely nothing would count as misunderstanding it.

To understand the meaning of the word 'red' is to be able to use 'red' correctly, on Wittgenstein's view. But to be able to use 'red' correctly is to grasp the rules governing the use of 'red'. So presumably, someone who misunderstands the meaning of the word 'red' is someone who has failed to grasp the rules governing the use of the word 'red'. More specifically, we can say that on Wittgenstein's view, to misunderstand the meaning of a word is to be in *error* about the rules governing the word's use. What is it to misunderstand a concept, according to Wittgenstein? Given that Wittgenstein is thinking about concept possession on the model of grasp of linguistic meaning, we can say that on Wittgenstein's view, to misunderstand the concept C is to be in error about the rules governing the use of the word '*C*'. To misunderstand the concept red, for example, is to be in error about the rules governing the use of the word 'red'.

As in the case of understanding, we should not expect these characterisations to ground positive accounts of what it is to misunderstand a concept or the meaning of a word. Take, for example, the claim that to misunderstand the concept C is to be in error about the rules governing the use of the word '*C*'. To be in error about the rules governing the use of the word '*C*' is to be in error about the meaning of '*C*', on Wittgenstein's view. Given this consideration, and given that we are to think of concept possession on the model of grasp of linguistic meaning, to be in error about the rules governing the use of a word is, presumably, to be in error about the concept which that word expresses. For example, to be in error about the rules governing the use of the word 'red' is, presumably, to be in error about the concept red. If this is right, then the claim that to misunderstand the concept C is to be in error about the rules governing the use of the word '*C*' amounts to the claim that to misunderstand the concept C is to be in error about the concept which the word '*C*' expresses. But the notion of misunderstanding a concept and the notion of being in error about the concept which a word expresses are too closely connected for this claim to amount to a non-circular explication of the notion of misunderstanding a concept.

I want to make one final point before moving on. I take it that someone who is in error about the rules governing the use of a word is disposed to use that word incorrectly. Moreover, I take it that they are disposed to use the word incorrectly precisely *because* they are in error about

the rules governing its use. Suppose that I am in error about the rules governing the use of the term 'condone'. Suppose I think that it is in accordance with those rules to describe someone who strongly disapproves of certain behaviour (who condemns it) as condoning it. In this case, I am disposed to use the word 'condone' incorrectly. I am disposed to describe people who in fact condemn particular behaviour as condoning that behaviour. More specifically, I am disposed to use the word incorrectly *because* I am in error about the rules governing its use. Given that to misunderstand the concept condone is to be in error about the rules governing the use of the word 'condone', I take it that someone who misunderstands the concept condone is disposed to use the word 'condone' incorrectly because they are in error about the rules governing its use.

If all it is to be disposed to use a word incorrectly is to be disposed to apply that word to things to which it does not in fact apply, then not every case in which one is disposed to use a word incorrectly is a case in which one is so disposed *because* one is in error about the rules governing the use of that word. For instance, I might be disposed to wrongly describe Jones as condoning certain behaviour, not because I misunderstand the term 'condone', but because I am mistaken about Jones' attitude to that behaviour. Perhaps Jones is very good at concealing his disapproval. In this case, I am disposed to use the word 'condone' incorrectly. But I am not disposed to use it incorrectly *because* I am in error about the rules governing its use. In comparison, someone who is in error about the rules governing the use of the word 'condone'—someone who misunderstands the concept condone—is disposed to use 'condone' incorrectly because they are in error about the rules governing its use.

Now that we have a sense of what it is to misunderstand a concept on Wittgenstein's view, we can ask what Wittgenstein would say about the mental content of subjects like Alf on Earth before they are switched. Specifically, we want to know whether Wittgenstein would agree that on Earth before he is switched, Alf has thoughts involving the concept arthritis, even though it is a concept which he misunderstands.

Let us ask a prior question: Does Wittgenstein accept *Misunderstanding*? Does Wittgenstein think that a subject who misunderstands a concept can nevertheless have thoughts involving that concept? Different strands in Wittgenstein's thought can seem to pull against one another with respect to this question. One strand which can seem to pull in the direction of a 'Yes' has to do with Wittgenstein's views about the relation between questions in a philosophical context about whether a subject has a particular concept or is thinking a particular thought and our everyday practice of attributing concepts and thoughts to one another. In Chapter 1, Subsection 1.2 we noted that on Burge's view, standard practice regarding belief attributions is instructive with respect to a subject's actual mental content. Wittgenstein would agree with Burge on this point. Like Burge, Wittgenstein thinks that questions like 'Is *S* thinking that *p*?' are answerable to our everyday practice of attributing the belief that *p* to one another. That is to say, we should address such questions by recalling the criteria which would warrant an ascription of a belief that *p* to a subject in an everyday context. Now it seems to be a feature of our everyday practice that we *do* attribute beliefs involving the concept $\underline{C}$ to subjects who misunderstand $\underline{C}$. So, given that he thinks that our everyday practice of ascribing beliefs constitutes the standard, it seems that Wittgenstein would—or at least should—agree that subjects can have thoughts involving concepts which they misunderstand.

I said that Wittgenstein is likely to agree with Burge about the significance of our everyday practice of attributing concepts and thoughts to one another within the context of philosophical reflection. It is important, however, to note a difference between the way in which Burge and Wittgenstein conceive of the authority of standard usage.[11] On Burge's view, standard usage is revisable in principle. If considerations come to light which show that usage to be in error, then on Burge's view it ought to be modified accordingly (Burge, 1979: 102). In comparison, on Wittgenstein's view the very idea that considerations might come to light which show standard usage to be in error is incoherent. As noted above, according to Wittgenstein rules govern standard usage in the sense that they determine how words, expressions and propositions within that

[11] I do not have the space to do more than simply note this difference here.

language ought to be used. These rules are expressed by grammatical propositions. The very idea that grammatical propositions might be true or false simply gets no purchase within a Wittgensteinian framework. Grammatical propositions determine the myriad of ways in which what we say is beholden to the world for its truth or falsity. But grammatical propositions are not themselves beholden to the world. Grammar is not, on Wittgenstein's view, accountable to a more fundamental reality. It is in precisely this sense that grammar is arbitrary, according to Wittgenstein (PI §497).[12]

To return to the main line of discussion, we have identified a strand of thought in Wittgenstein's later work which seems to support *Misunderstanding*, namely, the thought that our ordinary practices of belief attribution are instructive when thinking in a philosophical context about whether a subject has thoughts involving a particular concept. Yet certain other strands in Wittgenstein's later thought speak in favour of rejecting *Misunderstanding*. I am going to focus on one strand in particular.

In Chapter 1, Subsection 1.3 I noted that it is a recurring theme throughout *Investigations*, and Wittgenstein's later work more generally, that considered in isolation—that is, apart from its application—any image is variously interpretable. This theme is explored in *Investigations* §139 and §140. In §139 Wittgenstein asks whether understanding a word—the word 'cube', say—could consist in an image coming before one's mind. His answer is that it could not, precisely because any image that comes before one's mind can be variously applied. He goes on in §141 to ask whether an image *and* its application could come before one's mind. Wittgenstein's answer is that it can, 'only we need to become clearer about our application of *this* expression' (that is, the application of the expression 'an application comes before one's mind'). Wittgenstein imagines a scenario in which he is trying to communicate a particular method for applying a picture of a cube to another person. He goes on:

> … let's ask ourselves in what case we'd say that the method I mean comes before his mind.

[12] For an in depth discussion of this component of Wittgenstein's view, see Forster (2004).

> Now evidently we accept two different kinds of criteria for this: on the one hand, the picture (of whatever kind) that he visualised at some time or other; on the other, the application which—in the course of time—he makes of this image. (PI §141)

Considered apart from its application, any image or picture which comes before the mind can be variously interpreted, on Wittgenstein's view. As the discussion in Chapter 1, Subsection 1.3 made clear, it is for this reason that he thinks that it cannot be in virtue of a particular picture or image's coming before one's mind that one's understanding has the content which it does. Yet in *Investigations* §141 he acknowledges that there are nevertheless circumstances in which the fact that a particular picture or image comes before one's mind is a *criterion* for one's understanding that *p*, say (as opposed to that *q*). Wittgenstein goes on to identify a second kind of criteria, namely, the application or use which one makes of this picture or image.

It is not peculiar to grasping methods of application or to understanding considered more generally that there are these two kinds of criteria, on Wittgenstein's view. Elsewhere, Wittgenstein makes the same basic point with respect to meaning something by a word. For example:

> The important point is to see that the meaning of a word can be represented in two different ways: (1) by an image or picture, or something which corresponds to the word, (2) by the use of the word—which also comes to the use of the picture. (*Wittgenstein's Lectures on the Foundations of Mathematics*, hereafter LFM, 190)

Here Wittgenstein puts the point in terms of *representing* one's meaning, but he would be prepared to rephrase that point in terms of criteria for meaning. Wittgenstein acknowledges that there are at least two different sorts of criteria for one's meaning, say, cube by the word 'cube': first, the picture or image that one associates with the word 'cube'; and second, the use which one goes on to make of this picture or image.

Now clearly these two kinds of criteria might conflict with one another. Suppose, for example, that you and I associate the same image with the word 'cube' but apply the word differently. What should we say in such cases? Do you and I mean the same thing by 'cube' or do we mean different things? Wittgenstein's own view is clear enough:

> What is essential now is to see that the same thing may be in our minds when we hear the word and yet the application still be different. Has it the *same* meaning both times? I think we would deny that. (PI §140)

Wittgenstein acknowledges that there are circumstances in which the fact that you and I associate the same image with the word 'cube' is a criterion for our using the word 'cube' to mean the same thing. But he thinks that in cases where we associate the same image with the word but go on to use the word 'cube' in different ways, this criterion is defeated. In such cases, we use the word 'cube' to mean different things.

Wittgenstein's point is not exclusively about meaning. Consider, for example, the following passage:

> Suppose you say, "What does it mean for a man to understand a sign?"—You might say, "It means he gets hold of a certain idea."
> Then if two people—Lewy and I—get hold of the same idea of 'two', we both understand it in the same way.—Suppose he had got hold of the same idea of 'two' as I, whatever that means. What if he used it differently in future? Would I still say he has got hold of the same idea? You might say, "Yes, he's got hold of the same idea, but applies it differently."
> Suppose someone said, "Couldn't there be telepathy, and I know that Lewy has got hold of the same idea as I have? Or a medium might tell us." Would we say it was understood in the same way if it was applied in different ways? In fact it is clear that under those circumstances whatever the medium saw or said would be irrelevant to the question.
> 'Having the same idea' is only interesting if (a) we have a criterion for having the same idea, (b) this guarantees that we use the word in the same way. (LFM 23-24)

In this passage, Wittgenstein is making the same basic point in relation to understanding which in *Investigations* §140 he makes in relation to meaning something by a word. In certain contexts, the fact that Lewy and Wittgenstein associate the same image or idea with the word 'two' is a criterion for their understanding 'two' in the same way. But suppose that Wittgenstein and Lewy associate the same image or idea with the word 'two' but go on to use it differently. Wittgenstein asks whether in such a case we would say that he and Lewy understand 'two' in the same way. Although he does not say so explicitly, I take it that his answer is 'No' (''Having the same idea' is only interesting if … this guarantees that we use the word in the same way'). In such a case, Wittgenstein wants to say, Lewy and he understand different things by the word 'two'.

Of course, it is not plausible that just *any* difference in application suffices for a difference in understanding. But Wittgenstein does not mean to suggest that it does. Wittgenstein's primary concern in the passage I have just cited is to displace a particular picture of what it is to grasp the meaning of a word. According to the picture Wittgenstein is trying to displace, to grasp the meaning of a word is to grasp something—an image or idea—which is the source of correct use (PI §146), something which, when considered apart from its application, suffices to determine how it is that the particular word ought to be used. In the course of undermining this picture Wittgenstein is inclined to say things like ''Having the same idea' is only interesting if … this guarantees that we use the word in the same way'. But claims of this sort must be understood within the context in which they are made. The context in which they are made is one in which Wittgenstein is concerned with undermining a picture of what it is to understand the meaning of a word which in his view neglects the connection between understanding and use. Wittgenstein does not mean that 'Having the same idea' is only interesting if this guarantees that we use the word in *exactly* the same way. Indeed, within his own preferred picture of understanding, Wittgenstein would acknowledge that there *are* cases in which two subjects who have the same idea or image in mind understand a word to mean the same thing, even though they are disposed to use it differently.

For example, suppose that Smith and Jones make a habit of watching the sunset together. Suppose that on such occasions, Smith is disposed to say things like 'What a striking shade a red!' Suppose that Jones, in comparison, is disposed to say nothing at all. I do not think that it is Wittgenstein's considered view that this difference in the way in which they are disposed to apply the word 'red' suffices for a difference in the way in which Smith and Jones understand 'red'. Wittgenstein would agree that in spite of this difference, Smith and Jones may understand 'red' to mean the same thing.

What sorts of differences in dispositions of application does Wittgenstein think *do* suffice for differences in understanding? Smith and Jones are disposed to use the word 'red' differently. But neither is disposed to use it incorrectly. Imagine a modified scenario. Suppose that Smith is disposed to use 'red' in ways which accord with the rules governing its use, but Jones is not.

Perhaps Jones is disposed to apply the word 'red' to both red *and* orange objects. Suppose further that Jones' disposition is ultimately to be explained in terms of his being in error about the rules governing the use of 'red'. In this case, I think that the difference between the way Smith and Jones are disposed to apply the word 'red' *would* suffice for a difference in understanding, on Wittgenstein's considered view. More generally, I think that any case in which the difference between the ways in which *A* and *B* are disposed to apply the word '*C*' is such that, first, *A* is disposed to apply '*C*' correctly but *B* is disposed to apply it incorrectly (or vice versa), and second, *B*'s disposition is ultimately to be explained in terms of her being in error about the rules governing the use of '*C*', is a case in which *A* and *B* understand the word '*C*' to mean different things, respectively, on Wittgenstein's view. Given that Wittgenstein is thinking of concept possession on the model of grasp of linguistic meaning we can say that on Wittgenstein's view, any case in which these two conditions are satisfied is a case in which *A* and *B* use the term '*C*' to express different concepts.

As I noted earlier, someone who is in error about the rules governing the use of a word is disposed to use that word incorrectly because they are in error about those rules. Someone who is in error about the rules governing the use of the word 'condone', for example, is disposed to use the word 'condone' incorrectly because they are in error about those rules. So if what I have said in the previous paragraph is correct, then someone who is in error about the rules governing the use of the word 'condone' is someone who uses that word to express a concept which is different to the concept which those who are not in error about those rules use 'condone' to express. But note that if someone is in error about the rules governing the use of the word 'condone', then that person misunderstands the concept <u>condone</u>, on Wittgenstein's view. Recall that on that view, to misunderstand the concept <u>*C*</u> is to be in error about the rules governing the use of the word '*C*'. So if what I have said in the previous paragraph is correct, then someone who misunderstands the concept <u>condone</u> is someone who uses the word 'condone' to express a concept which is different to the concept which those who do not misunderstand the concept use 'condone' to express.

If this is right, then we have located a strand of thought in Wittgenstein's later work which is at odds with *Misunderstanding*, with the idea that a subject can have thoughts involving concepts which she misunderstands. The strand of thought (which finds expression in the passage from *Lectures on the Foundations of Mathematics* cited above) is that any case in which the difference between the ways in which *A* and *B* are disposed to apply the word '*C*' is such that, first, *A* is disposed to apply '*C*' correctly but *B* is disposed to apply it incorrectly (or vice versa), and second, *B*'s disposition is ultimately to be explained in terms of her being in error about the rules governing the use of '*C*', is a case in which *A* and *B* understand the word '*C*' to mean different things, respectively. For the sake of convenience, I shall refer to this strand of thought hereafter as *Sufficient Difference*.

Let me summarise the reasoning that leads to the conclusion that *Sufficient Difference* is at odds with *Misunderstanding*. We know from our earlier discussion that any case in which one misunderstands the concept $\underline{C}$ is a case in which one is in error about the rules governing the use of the word '*C*', on Wittgenstein's view. We also know that any case in which one is in error about the rules governing the use of the word '*C*' is a case in which one is disposed to use '*C*' incorrectly because one is in error about the rules governing its use. So any case in which one misunderstands the concept $\underline{C}$ is a case in which one is disposed to use the term '*C*' incorrectly because one is in error about the rules governing its use. If *Sufficient Difference* is true, then any case in which one is disposed to incorrectly use the term '*C*' because one is in error about the rules governing the use of '*C*' is a case in which one understands '*C*' to mean something different to what those who are disposed to use '*C*' correctly understand it to mean. Given that Wittgenstein is thinking about concept possession on the model of grasp of linguistic meaning, we can say that any such case is one in which one uses '*C*' to express a concept which is different to the concept which those who are disposed to apply '*C*' correctly use it to express. Putting this together we can conclude that if *Sufficient Difference* is true, then any case in which one misunderstands the concept $\underline{C}$ is a case in which one uses the term '*C*' to express a concept which is different to the concept which those who do not misunderstand use '*C*' to express. In other words, if *Sufficient Difference* is true, then

Misunderstanding is false; a subject cannot have thoughts involving concepts which she misunderstands.

Here is a summary of the discussion thus far. We have identified a strand of thought in Wittgenstein's later work which seems to support *Misunderstanding*, namely, the thought that our ordinary practices of belief attribution are instructive when thinking in a philosophical context about whether a subject has thoughts involving a particular concept. We have also identified a strand of thought which is at odds with *Misunderstanding*, namely, *Sufficient Difference*.

In fact, I do not think that the thought that our ordinary practices of belief attribution are instructive when thinking in a philosophical context about whether a subject has thoughts involving a particular concept supports *Misunderstanding*. One could accept this claim and agree that we routinely ascribe beliefs involving the concept *C* to subjects who evidently misunderstand *C*, and yet reject *Misunderstanding*. This is also the position of Hanjo Glock and John Preston. Glock and Preston acknowledge that we routinely ascribe beliefs involving *C* to subjects who clearly misunderstand *C*. They acknowledge, for example, that it would not be unusual to ascribe to Alf the belief that he has arthritis in his thigh. But, they insist, there is a question about exactly what such an ascription amounts to. Glock and Preston think that if asked to say precisely what it is that Alf believes, 'we would answer it by saying something like "He believes that he has some rheumatoid inflammation in his thigh"' (Glock & Preston, 1995: 519). Donald Davidson agrees. He writes:

> Suppose that I, who thinks the word 'arthritis' applies to inflammation of the joints only if caused by calcium deposits, and my friend Arthur, who knows better, both sincerely utter to Smith the words 'Carl has arthritis' … If Smith (unspoiled by philosophy) reports to still another party (perhaps a distant doctor attempting a diagnosis on the basis of a telephone report) that Arthur and I both have said, and believe, that Carl has arthritis, he may actively mislead *his* hearer. If this danger were to arise, Smith, alert to the facts, would not simply say 'Arthur and Davidson both believe Carl has arthritis'; he would add something like 'But Davidson thinks arthritis must be caused by calcium deposits'. The need to make this addition I take to show that the simple attribution was not quite right; there was a relevant difference in the thoughts Arthur and I expressed when we said 'Carl has arthritis'. (Davidson, 2001c: 27-28)

Glock, Preston and Davidson accept (at least for argument's sake) that standard practice is instructive when thinking in a philosophical context about a subject's actual mental content. But all three think that Burge gives a less than complete characterisation of that practice. On their view, in cases where it is clear that *S* misunderstands *C* but we nevertheless find it natural to ascribe beliefs involving *C* to *S*, we are usually prepared to revise our original ascriptions when asked to specify the content of *S*'s beliefs exactly. Our readiness to revise our original ascriptions in cases involving misunderstanding is a feature of standard practice which, on their view, Burge overlooks.

I agree with Glock, Preston and Davidson. And if *Sufficient Difference* is a strand in Wittgenstein's later thought, then there is evidence that Wittgenstein would also agree. Burge incompletely characterises our ordinary practice of belief attribution. Because Burge's characterisation is incomplete, one can accept the claim that standard practice is instructive when thinking in a philosophical context about whether a subject has thoughts involving a particular concept and yet reject the claim that a subject can have thoughts involving concepts which she misunderstands. In other words, the former claim does not support *Misunderstanding*.

The question which prompted this discussion was 'Would Wittgenstein agree that on Earth before he is switched, Alf has thoughts involving the concept arthritis?' We began by identifying a strand of thought in Wittgenstein's later work which seemed to support *Misunderstanding*, but which on closer inspection turned out not to do so. We went on to identify a second strand of thought in Wittgenstein's work, namely, *Sufficient Difference*, which is at odds with *Misunderstanding*. To the extent that *Sufficient Difference* is at odds with *Misunderstanding*, it is evidence that the answer to our question is 'No', that Wittgenstein would *not* agree that on Earth before he is switched, Alf has thoughts involving the concept arthritis.

What concept *does* Alf use the term 'arthritis' to express on Earth, on Wittgenstein's view? It is not clear how Wittgenstein will answer this question. However he answers it, I think his view will be that Alf uses 'arthritis' to express the same concept on Twin Earth as he uses it to express

on Earth. On Twin Earth, Alf is disposed to respond to questioning about what he means by 'arthritis' in the same way as he is disposed to respond on Earth. Given the role which Wittgenstein thinks a subject's dispositions to respond to questioning about her meaning play in individuating her mental states, it is likely to be Wittgenstein's view that Alf's 'arthritis' concept remains constant in cases in which he is being slowly switched. If this is right, then slow-switching objections do not have initial plausibility as objections to the arthritis case, given a Wittgensteinian framework for thinking about the individuation of mental states. Such objections rely on slow-switching bringing about some change in the content of a subject's mental states. But slow-switching will not do so in Alf's case, given that framework.

CONCLUSION

Two aims have guided discussion in this chapter. The first aim was to show that Wittgenstein accepts the driving intuition about self-knowledge, that is, the claim that we normally know groundlessly what we are thinking. This aim was pursued in Section 1. I went on to consider how Wittgenstein might respond to the swift argument for inconsistency between externalism's driving intuition and the driving intuition about self-knowledge (as outlined at the beginning of Chapter 2, Section 1). I defended the view that Wittgenstein's response is effectively the same as Burge's response. The second aim was to consider whether slow-switching objections have initial plausibility as objections to certain of the thought experiments which we considered in Chapter 1, Subsections 1.2 and 1.3, given a Wittgensteinian framework for thinking about the individuation of mental states. In Section 2, I considered whether slow-switching objections have initial plausibility as objections to the two thoughts experiments belonging to Wittgenstein which I discussed in Chapter 1, Subsection 1.3, given Wittgenstein's views about the way in which mental content is individuated. I concluded that they do not. In Section 3, I considered whether slow-switching objections have initial plausibility as objections to the arthritis case, given Wittgenstein's views. Again, I concluded that they do not. In the course of reaching this conclusion, I defended

the view that there is a strand in Wittgenstein's later thought which is at odds with *Misunderstanding*, with the claim that a subject can have thoughts involving concepts which she misunderstands.

Chapter 5

WITTGENSTEIN'S EXTERNALISM AND THE OBJECTION FROM BRUTE SUCCESS

INTRODUCTION

My aims in this chapter are threefold. In Chapter 4, I concluded that slow-switching objections do not have initial plausibility either as objections to Sally's case or to the arthritis case, given a Wittgensteinian framework for thinking about the individuation of mental states. My first aim in Chapter 5 is to consider how Wittgenstein might respond should it turn out that there is some further slow-switching scenario with respect to which slow-switching objections *do* have initial plausibility, given this framework. I defend the view that Wittgenstein's response is likely to be similar in form to Burge's response (which I outlined in Chapter 2, Section 2). Wittgenstein is likely to be sympathetic to the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. He is also likely to reject the claim that one's warrant for judgments about one's mental states depends on one's being able to discriminate one's thoughts from relevant alternatives.

My second aim is to consider whether Wittgenstein offers a positive account of one's warrant for judgments about one's mental states and if so how that account ought to be understood. In relation to this second aim, I defend John McDowell's interpretation of Wittgenstein on self-knowledge.

My third aim is to defend the view that the objection from brute success has no purchase against an externalism which incorporates *Sufficient Difference*, the strand of thought which in Chapter 4, Section 3 I found a basis for in Wittgenstein's later work. The objection from brute success as presented in Chapter 3 relies on *Misunderstanding*, on the claim that a subject can have thoughts involving concepts which she misunderstands. But on an externalism which incorporates *Sufficient Difference*, *Misunderstanding* is false. Consequently, the objection from brute success as presented in Chapter 3 has no purchase against such an externalism. In Subsection 2.1, I consider whether there are good, independent reasons for challenging the assessment of the content of Jack's mental states on Twin Earth offered by Wittgenstein's externalism. I conclude that there is not. In Subsection 2.2, I consider whether the objection from brute success applies if

Misunderstanding is substituted with *False Belief*, given a Wittgensteinian framework. I note two considerations which support the view that it does not.

1. WITTGENSTEIN AND OUR WARRANT FOR SELF-KNOWLEDGE

1.1 Wittgenstein and Slow-Switching Objections

In Chapter 4, I defended the view that given a Wittgensteinian framework for thinking about the individuation of mental states, slow-switching objections do not have initial plausibility either as objections to Sally's case or to the arthritis case. Slow-switching objections depend on slow-switching bringing about regular changes in the content of a subject's mental states. But slow-switching will not bring this about in either Sally or Alf's case, given a Wittgensteinian framework. To be clear, I do not take the discussion in Chapter 4 to have shown that given a Wittgensteinian framework, there is *no* slow-switching scenario with respect to which slow-switching objections have initial plausibility. I do not take that discussion to have ruled out the possibility that there is some scenario in which we can expect that slow-switching *will* bring about regular changes in the content of a subject's mental states, given this framework.

Suppose that such a scenario is formulated. The proponent of slow-switching objections might then press the line of questioning which in Chapter 2, Section 1 we imagined her pressing against Burge: Given that slow-switching will bring about regular changes in the subject's mental states, how can the subject know groundlessly in cases where she is being slow-switched what she is thinking? In order for the subject to know groundlessly that she is thinking that *p*, say, she must be able to discriminate, on the basis of reflection alone, between instances in which she is thinking that *p* and instances in which she is thinking some relevant alternative thought. But in any case in which slow-switching would result in regular changes in the content of the subject's mental states, this apparently is not something which the subject is in a position to do.

In Chapter 2, Section 2 we noted that Burge's response to this line of questioning has two components. The first component involves defending a claim about the accuracy of one's judgments about one's mental states in slow-switching scenarios, namely, the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. The second component involves defending a claim about the source of one's warrant for judgments about one's mental states. According to this claim, one's warrant for first-person judgments takes the form of an entitlement and derives from one's identity as a critical reasoner, *not* from one's ability to identify one's thoughts or to discriminate them from relevant alternatives. Together, these components support the claim that slow-switching cannot undermine the knowledgeability of one's mental states. On Burge's view, slow-switching cannot bring it about, either that one's judgments about one's mental states are subject to error, or that those judgments, although true, do not constitute knowledge.

We can expect Wittgenstein's response to slow-switching objections to be similar in form. With regards to the first component, Wittgenstein does not in fact make the points which Burge makes (about the self-verifying nature of cogito-like judgments or the role of pure preservative memory in preserving the content of one's earlier states) in the course of defending the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. But these points are certainly consistent with Wittgenstein's general view. With regards to the second component, we can expect that Wittgenstein, like Burge, will reject *Discrimination*. Wittgenstein would almost certainly disagree that one's being warranted in judging that one is thinking that *p* depends on one's having the sort of discriminatory abilities which that claim requires. Whether Wittgenstein offers any *positive* account of one's warrant for judgments about one's mental states—and if so, where that account might be located with respect to the scheme for classifying accounts of self-knowledge which I set out in Chapter 1, Subsection 2.2—are questions to which I now turn.

1.2 Wright and McDowell on Wittgenstein on Self-Knowledge

There is disagreement about Wittgenstein's account of our warrant for self-knowledge. Perhaps the most widely accepted interpretation of that account is the interpretation offered by Crispin Wright, which in Chapter 1, Subsection 2.2 I introduced as an example of an account which is non-epistemic and non-substantive. Recall that on Wright's interpretation, it is part of the grammar of mental states that my sincere avowal that I am thinking that *p* has default standing; in the absence of any reason to reject it, my avowal is to be counted as correct. On this view—hereafter *the default view*—the relation between one's mental states and one's judgments about one's mental states is not epistemic but, rather, *constitutive*. In the absence of evidence to the contrary, my sincere avowal (or the fact that I am disposed to sincerely avow) that I am thinking that *p makes it the case* that I am in fact thinking that *p*. Accordingly, the authority of one's judgments about one's mental states is accounted for in terms of their constitutive role, and not in terms of 'an associated epistemologically privileged relation in which the subject stands to [those states]' (Wright, 2001: 312).[13]

In Chapter 2, Section 3 I situated Burge's account of our warrant for first-person judgments with respect to the scheme for classifying accounts of self-knowledge which I set out in Chapter 1, Subsection 2.2. I concluded that Burge's account is most fairly described as epistemic and substantive. If Wright's reading is accurate, then clearly, Wittgenstein's account differs from Burge's account in both respects. On Wright's reading, Wittgenstein's account is non-epistemic because it identifies some non-epistemic feature of one's judgments about one's mental states—their default standing—as grounding one's warrant for those judgements. It is non-substantive because it denies that one's judgments about one's mental states involve a cognitive achievement. If Wright is correct, Wittgenstein does not think that one's judgments about one's mental states involve the detection of states of affairs which exist even partly independently of one's judging (under ideal conditions) that they do. One's sincerely judging (or one's being disposed to sincerely

[13] See the surrounding passages for evidence of Wright's attribution of the default view to Wittgenstein.

judge) that one is thinking that *p* is, on Wright's reading, wholly constitutive of one's actually thinking that *p*.[14]

Indeed, if the default view *is* Wittgenstein's, then there might appear to be the following further point of difference between Wittgenstein's account of our warrant for self-knowledge and Burge's account. Burge's account entails a claim about the accuracy of one's judgments about one's *reasons and reasoning*. According to Burge, one's warrant for one's judgments about one's mental states takes the form of an entitlement which derives from one's identity as a critical reasoner. But in order to be a critical reasoner, Burge thinks, one's judgments about one's reasons and reasoning must normally be correct. So on Burge's view, it follows from the fact that one's judgments about one's mental states are warranted that one's judgments about one's reasons and reasoning are normally correct. But it is not obvious that this follows on the default view. More exactly, it is not obvious that on the default view it follows from the fact that one's judgments about one's mental states are warranted that one's judgments about one's *epistemic* reasons are normally correct. After all, judgments about, say, whether I have a reason to believe that *p* do not have the sort of default standing which is enjoyed by my judgments about my mental states. It is not constitutive of my having a reason to believe that *p* that I sincerely judge (or am disposed to

[14] At one point Wright suggests that it is a sufficient but not a *necessary* condition for one's being in a particular mental state that one sincerely avows that one is (Wright, 2001: 313-314). If this is right, then it is not the case that one's sincerely judging (or one's being disposed to sincerely judge) that one is thinking that *p* is *wholly* constitutive of one's actually thinking that *p*, on the default view (for it is consistent with that view that one might believe that *p* even though one does not sincerely judge, or is not disposed to sincerely judge, that one believes that *p*).
But this suggestion seems inconsistent with other features of Wright's view. Specifically, Wright's suggestion seems at odds with the line of thought which Wright thinks motivates the default view. As I go on to make clear, Wright thinks that only the default view is in a position to explain what he calls the 'disposition-like theoreticity' of mental states. Mental states are disposition-like in the sense that the question of whether or not I have, say, a particular intention—like the question of whether or not I have a particular disposition—is sensitive to what I go on to do or to say. If we conceive of mental states as occurrent states of consciousness, then, Wright thinks, we are at a loss to explain this characteristic. The default view, in comparison, has a ready explanation, according to Wright: Whether a subject has, say, an intention to φ is sensitive to how they go on because what it is that they go on to say about their intention (or what it is that they are disposed to go on to say about their intention) plays a role in *constituting* that intention.
It follows from Wright's suggestion that one might intend to φ even though one does not avow (and is not disposed to avow) that one intends to φ. But if the disposition-like theoreticity of mental states really is to be accounted for in terms of the constitutive nature of first-person judgments, then presumably, if one intends to φ but does not avow (and is not disposed to avow) that intention, one's intention cannot be disposition-like. But if it is not disposition-like, then in what sense is one's intention to φ really an *intention* at all?

sincerely judge) that I do. One might think that judgments about one's *explanatory* reasons—judgments about the reasons for which, say, one believes or intends what one does—have a kind of default standing. For example, one might think that it is constitutive of my believing that *p* because of *q* that I sincerely judge (or am disposed to sincerely judge) that I believe that *p* because of *q*. But the analogous claim does not seem to be true of judgments about one's epistemic reasons. So on a first pass, it might look as though it is possible for one's judgments about one's mental states to be warranted and yet for one's judgments about what one has reason to believe to be normally incorrect, if the default view is true.

In fact, I do not think that this is a possibility, given the default view. According to Wright, the success of the language game involving the ascription of mental states to oneself and others rests on:

> … the contingency that taking the apparent self-conceptions of others seriously, in the sense involved in crediting their apparent beliefs about their intentional states, as expressed in their avowals, with authority, almost always tends to result in an overall picture of their psychology which is more illuminating—as it happens, *enormously* more illuminating—than anything which might be gleaned by respecting all the data *except* the subject's self-testimony. And that in turn rests on the contingency that we are, each of us, ceaselessly but—on the proposed conception—subcognitively moved to opinions concerning our own intentional states which will indeed give good service to others in their attempt to understand us. (Wright, 2001: 313)[15]

Arguably, neither of the contingencies Wright identifies in this passage would obtain if we were not normally correct about our epistemic reasons. If we were normally incorrect about our epistemic reasons, then it is not clear that taking the apparent self-conceptions of others seriously *would* almost always tend to result in a more illuminating picture of their psychology. It is not clear that our opinions concerning our own mental states *would* give good service to others in their attempt to understand us. Why not? If *S* was normally incorrect about her epistemic reasons, then taking *S*'s judgments about her epistemic reasons at face value might result in confusion about what it is that *S* thinks. After all, *S*'s judgments about what she has a reason to think or to do are a

[15] Wright's suggestion that we are 'subcognitively moved to opinions concerning our own intentional states' makes it sound as though our opinions about our mental states are mere hunches, which just happen to be normally correct. But this does not seem right. It seems that we are usually in a position to give some *rationale* for our self-ascriptions.

basis on which we ascribe mental states to *S*. If *S* sincerely judges that she has a reason to intend to φ, then that is grounds for thinking that *S* has beliefs and desires which give her a reason to intend to φ. If *S*'s judgment about her reasons is incorrect, then our own judgments about *S*'s beliefs and desires are also likely to be inaccurate. If *S*'s judgments about what she has reason to think or to do were *normally* incorrect, then our judgments about *S*'s mental states would be incorrect a significant amount of the time. Arguably, we would gain a better picture of *S*'s psychology in such a case if we were to ignore those judgments altogether.

In such a case, the contingencies which Wright identifies in the passage above would fail to hold. But if these contingencies failed to hold, then the authority of one's judgments about one's mental states would, presumably, be undermined. So I think that it is true on the default view, as it is on Burge's, that if one's judgments about one's mental states are authoritative, then one's judgments about one's reasons and reasoning must normally be correct.

Of course, the default view offers no account of *how* it is that our judgments about our epistemic reasons could normally be correct. We might think that this counts against the default view, considered as a general account of our warrant for first-person judgments. Presumably, the proponent of the default view will want to offer some such account. That account cannot appeal to the constitutive nature of one's judgments about one's epistemic reasons, for as I pointed out above, it is not plausible that one's judgments about one's epistemic reasons have the sort of default standing which, according to the default view, is enjoyed by one's judgments about one's mental states. Rather, it seems that any such account must present knowledge of one's epistemic reasons as substantive. But if it should turn out that the proponent of the default view ought to endorse an account of how it is that we know our epistemic reasons which presents such knowledge as substantive, then it is difficult to see what her rationale might be for holding knowledge of one's mental states to be non-substantive. Why not simply extend the account of knowledge of one's epistemic reasons to knowledge of one's mental states?

In any case, I want to consider a different question: Is Wright correct in attributing the default view to Wittgenstein? On Wright's reading, Wittgenstein's views about self-knowledge involve both a positive and a negative component. The positive component finds expression in the default view. The core of the negative component is Wittgenstein's rejection of what Wright calls the *Cartesian idea of the mind* (Wright, 2001: 293). Central to this idea is the notion that we each of us have privileged observational access to the contents of our minds. The objects of inner observation—sensations and intentional states alike—are conceived of as occurrent phenomena of consciousness. On Wright's reading, Wittgenstein is hostile to the Cartesian idea because he is hostile to the notion that mental states are occurrent phenomena of consciousness. Wittgenstein stands opposed to this notion, Wright thinks, in large part because he thinks that if we conceive of mental states as occurrent phenomena of consciousness, then we are at a loss to explain what Wright calls their 'disposition-like theoreticity' (Wright, 1998: 30). It is part of Wittgenstein's view, Wright thinks, that the question of whether or not I have, say, a particular intention—like the question of whether or not I have a particular disposition—is sensitive to what I go on to do or to say. If we think of mental states as occurrent phenomena of consciousness, then we are inclined to account for this sensitivity by positing normative links which mental states bear to future ways of going on.

Indeed, it seems to be part of our pre-theoretical conception of mental states that they *do* bear these sorts of normative links. For example, it seems to be part of our pre-theoretical conception that if I, say, intend to go to Paris, then only my going to Paris will accord with my intention. In this way, my intention normatively constrains my future action. Wright, however, takes it to be one of the lessons of *Investigations* that my intention to go to Paris could not sustain these sorts of normative links to future ways of going on if it were indeed an occurrent phenomenon of consciousness, for Wright takes it to be Wittgenstein's considered view that *any* occurrent phenomenon of consciousness can be variously interpreted. If my intention were indeed an occurrent phenomenon of consciousness, then for any future way of going on, there would be *some* interpretation of my intention according to which that way of going on accorded with it. But

if for any conceivable way of going on, there is some interpretation of my intention according to which that way of going on accords with it, then my intention does *not* constrain my future action. So on Wright's reading, Wittgenstein works his way from the observation that whether I have a particular intention is sensitive to what I go on to do or say, to the conclusion that intentions—and mental states more generally—are not occurrent phenomena of consciousness. In Wright's own words, a fundamental point in Wittgenstein's polemic against the Cartesian idea is:

> … that the details of the phenomena of consciousness which may be associated with understanding, expecting, intending, hoping, etc., are neither in general called upon, nor able, to sustain the kinds of internal connection with aspects of a subject's (subsequent) doings and reactions which mental states of this kind essentially sustain. (Wright, 2001: 296)

On Wright's reading, then, in rejecting the Cartesian idea of the mind, Wittgenstein is at the same time rejecting the claim that mental states are occurrent phenomena of consciousness which bear normative links to future ways of going on. On what is, according to Wright, Wittgenstein's preferred view—the default view—the disposition-like theoreticity of mental states is explained in terms of the constitutive nature of one's judgments about one's mental states. Whether a subject has, say, an intention to go to Paris is sensitive to how they go on because what it is that they go on to say about their intention—what they go on to say about its content, about whether such and such an action conforms with it—plays a role in constituting that intention.

But *does* Wittgenstein want to deny that mental states are occurrent phenomena of consciousness? According to John McDowell, the answer is 'No'. According to McDowell, the claim that mental states are occurrent phenomena of consciousness is one which, appropriately understood, Wittgenstein would be happy to accept. On McDowell's view, Wittgenstein's challenge to the Cartesian idea consists in his bringing into question a certain conception of what it is for something to be an occurrent phenomenon of consciousness. According to the conception which Wittgenstein regards as problematic, it is only under some interpretation that an occurrent state of consciousness can bear normative links to future ways of going on.[16] According to this

[16] McDowell refers to this as *the master thesis* (McDowell, 1998b: 270).

conception, considered in themselves such states are normatively inert, as it were. It is this assumption, McDowell thinks, which Wittgenstein ultimately wishes us to reject. Once we reject this assumption, it does not remain for us to give some positive account of how mental states *could* bear normative links to future ways of going on. Having jettisoned those ways of thinking which made it appear problematic, the idea that mental states are occurrent states of consciousness which bear normative links to future ways of going on, '… can fall into place as simply part of a way of thinking that we are now able to take in our stride' (McDowell, 1998b: 274). As McDowell puts it:

> Wittgenstein's thought does not leave untouched the picture of the introspectable as a domain of self-containedly knowable states of affairs, only externally related to anything outside themselves, and expel the intentional from that domain. The key argument generalises so as to undermine that picture of the introspectable. Once we understand that, we can see that there is no need to be suspicious of including intentionality among the occurrent phenomena of consciousness. (McDowell, 1998a: 303)

I do not have the space to give a comprehensive comparative analysis of Wright and McDowell's readings of Wittgenstein. I do, however, think that on the whole, McDowell's reading is a more accurate portrayal of what Wittgenstein is up to in *Investigations* than is Wright's. There are several passages which support McDowell's contention that Wittgenstein's response to questions about how mental states could have the intentional properties which they seemingly do is to challenge the assumptions on which the questions are based (and not, as it is on Wright's reading, to put forward some positive account of the intentionality of mental states). One such passage (which McDowell does not cite) is *Investigations* §428:

> "A thought—what a strange thing!"—but it does not strike us as strange when we are thinking. A thought does not strike us as mysterious while we are thinking, but only when we say, as it were retrospectively, "How was that possible?" How was it possible for a thought to deal with *this very* object? It seems to us as if we had captured reality with the thought.

Wittgenstein's suggestion in this passage is that there is not anything genuinely mysterious about thoughts having the intentional properties that they do. There is, however, a way of thinking about thoughts which can make it *seem* mysterious. Although he does not in fact do so, it would not be unusual or out of keeping with what he says elsewhere for Wittgenstein to go on by encouraging us to challenge this way of thinking, as opposed to developing a theory of intentionality.

What is Wittgenstein's positive account of first-person authority on McDowell's reading? On McDowell's reading, Wittgenstein declines to offer a positive account (McDowell, 1998c: 58). On that reading, Wittgenstein's response to the question 'How can self-knowledge be groundless yet authoritative?' is similar in form to his response to the question 'How can occurrent states of consciousness bear normative links to ways of going on?' Wittgenstein's view is that there really is no special problem about how a subject can have groundless, warranted self-knowledge. We can be led to think that there is a problem by thinking about the phenomena in a particular way, for example, by having in mind a certain conception of what it is for knowledge to be groundless. McDowell imagines someone who thinks: first, that only non-inferential knowledge is groundless; second, that only observational knowledge is non-inferential; and third, that self-knowledge is not observational. Such a person would obviously have difficulty conceiving how self-knowledge might be groundless. The Wittgensteinian response, McDowell thinks, is not to advance a theory which explains how self-knowledge could be groundless though not observational, but to challenge the assumption that only observational knowledge is groundless.

We might think that if Wittgenstein declines to put forward a positive account of self-knowledge, then it is a mistake to try and situate him with respect to the schema for classifying accounts of self-knowledge which I set out in Chapter 1, Subsection 2.2. In fact, I do not think that this is the case. If McDowell's reading is accurate, then it is not the case that there is not *anything* which Wittgenstein thinks we can or should say about our capacity for self-knowledge. At a minimum, we should feel free to think of self-knowledge as groundless knowledge. More generally, though, we should 'take in our stride' those claims about our capacity for self-knowledge which belong to our pre-theoretical conception. Is it part of our pre-theoretical conception that self-knowledge is substantive or epistemic (in the senses of 'substantive' and 'epistemic' described in Chapter 1, Subsection 2.2)? It seems clear that it is part of our pre-theoretical conception that self-knowledge is substantive, that it involves the detection of states of affairs which exist partly independently of one's judging (under ideal conditions) that they do. We do not tend to think that one's judgments about one's mental states are wholly constitutive of those

states (as, for example, the default view claims). It is less clear whether it is part of our pre-theoretical conception that self-knowledge is epistemic, that is, whether it belongs to that conception that one's warrant for judgments about one's mental states derives from some epistemic feature of those judgments. It may be that our pre-theoretical conception is not definite enough for us to decide this question. In any case, I agree with McDowell when he suggests that on Wittgenstein's view, the notion that we each of us have privileged observational access to the contents of our own minds is a philosophical perversion of our pre-theoretical thinking. To summarise, the picture of self-knowledge which Wittgenstein would endorse on McDowell's reading is a picture according to which self-knowledge is substantive. It is less clear to me whether it is a picture according to which self-knowledge is epistemic.

Is it part of that picture that our judgments about what we are thinking are warranted only if our judgments about our reasons and reasoning are normally correct? I think the answer to this question is 'Yes'. I noted above that on Wright's reading, one's warrant for judgments about one's mental states rests on a deep contingency. On McDowell's view, Wright's reading misplaces the deep contingency:

> The contingency is that human beings are capable of acquiring a mastery of the concepts that figure in the relevant self-conceptions; since the default status of avowals is partly constitutive of the concepts, this mastery includes the ability to apply the concepts to oneself in a way that meshes with, for instance, one's subsequent performance in the manner required for the self-conception in question to help make sense of one, but without one's needing to wait and make sure of the mesh before one can know that the concept applies. (McDowell, 1998a: 318-319)

On McDowell's reading, it is Wittgenstein's view that our warrant for judgments about our mental states rests on the contingent fact that we are able to become competent participants within a practice. Becoming competent participants within a practice includes mastering the use of intentional concepts in self-ascriptions. So on Wittgenstein's view as interpreted by McDowell, our warrant for judgments about our mental states rests on the contingent fact that we are able to master the use of intentional concepts in self-ascriptions. Arguably, this contingency would not obtain if we were not normally right about our epistemic and explanatory reasons. It would seem to be partly constitutive of mastering the use of an intentional concept in self-ascriptions that one is

normally right both about the rational relations which thoughts involving that concept bear to other thoughts and about the reasons for which one in fact believes or intends what one does. So I think that it is true on McDowell's reading, as it is on Wright's, that if one's judgments about one's mental states are warranted, then one's judgments about one's reasons and reasoning must normally be correct.

Earlier, I said that on McDowell's reading, Wittgenstein declines to give a positive account of how it is that one's judgments about one's mental states could be both groundless and authoritative. But what I have just said might make it look as though McDowell's Wittgenstein *does* offer something like a positive account after all. If it is Wittgenstein's view that one's warrant for judgments about one's mental states rests on the contingent fact that one is able to become a competent participant within a practice, then why not say that on that view, one's warrant for judgments about one's mental states derives from one's identity as such a participant?

McDowell would likely take issue with the term 'derives'. He might prefer to say that on Wittgenstein's view, being warranted in one's judgments about one's mental states *involves* being a competent participant within a practice. But the question still stands. McDowell may well respond in something like the following way: 'The answer to your question depends on what is involved in giving a positive account of groundless self-knowledge. In the context of the disagreement between Wright and myself, a positive account is one which explains how groundless self-knowledge is so much as possible. Such an account will include an explanation of how it is so much as possible that one's judgments about one's mental states could normally be correct. Wright's Wittgenstein offers such an explanation. According to Wright's Wittgenstein, one's judgments about one's mental states are normally correct in virtue of the constitutive nature of those judgments. My Wittgenstein, in comparison, offers no such explanation. On my reading, Wittgenstein thinks that being warranted in one's self-ascriptions involves being a competent participant in a practice, where being a competent participant in a practice itself involves mastering the use of intentional concepts in self-ascriptions. But on my reading, Wittgenstein offers no account of how it is that we are able to possess this mastery. Having undergone the appropriate

training, we usually come to employ intentional concepts in accurate self-ascriptions. But my Wittgenstein would dissuade us from seeking a philosophical account of how our coming to employ intentional concepts in accurate self-ascriptions having undergone the appropriate training is so much as possible. If a positive account of groundless self-knowledge is an account which includes an explanation of how it is that one's judgments about one's mental states could normally be correct, then it is right to say that my Wittgenstein declines to give such an account'.

Is this a satisfying response? It is true that if a positive account is one which includes an explanation of how it is that one's judgments about one's mental states could normally be correct, then McDowell's Wittgenstein does not give a positive account. But then it is not obvious that this is a point of difference between McDowell's Wittgenstein and Wright's. It is not part of the default view that we are bound to count a subject's sincere self-ascriptions as correct come what may. In cases where there is conflicting evidence, in cases where, for example, a subject's sincere self-ascriptions are deeply at odds with her behaviour, we may challenge those ascriptions. In other words, for one's sincere self-ascriptions to play a constitutive role, on the default view, there has to be a mesh—of the sort which McDowell mentions in the passage above—between those ascriptions and one's behaviour. Now Wright's Wittgenstein offers us no account of how it is so much as possible that our sincere self-ascriptions normally *do* mesh with our behaviour. If a positive account of groundless self-knowledge is an account which includes an explanation of how it is that one's judgments about one's mental states could normally be correct, then it is not clear that Wright's Wittgenstein *does* offer a positive account. Wright's Wittgenstein neglects to explain how it is so much as possible that our sincere self-ascriptions mesh with our behaviour to the extent necessary for those ascriptions to play the sort of constitutive role which they commonly do, according to the default view.

Insofar as a satisfying response to our question is one which makes it clear why the account of groundless self-knowledge offered by Wright's Wittgenstein qualifies as a positive account, but the account put forward by McDowell's Wittgenstein does not, the answer to our question which I offered on McDowell's behalf is not a satisfying response.

2. WITTGENSTEIN'S EXTERNALISM AND THE OBJECTION FROM BRUTE SUCCESS

2.1 Assessing Wittgenstein's Response to the Objection from Brute Success

In Chapter 3, Subsection 3.2 I set out what I called the objection from brute success. I argued that given Burge's views about the way in which mental content is individuated, slow-switching can undermine knowledgeability of one's epistemic reasons. This is a troubling result, I suggested, given Burge's views about the sorts of things that epistemic reasons are. We have been considering some of the points of similarity and difference between Burge and Wittgenstein's respective views about the individuation of mental states and our warrant for self-knowledge. With these points of similarity and difference in mind, let us ask whether the objection from brute success as formulated in Chapter 3, Subsection 2.2 has purchase against Wittgenstein's externalism. Can slow-switching undermine knowledgeability of Jack's epistemic reasons, given Wittgenstein's views about the individuation of mental content?

If the conclusion I reached in Chapter 4, Section 3 is sound, then the answer to this question is 'No'. The objection from brute success relies on its being possible for slow-switching to bring it about that a subject is wrong about the logical relations between her mental states. It is only because slow-switching can bring this about that it can bring it about that one's warrant for judgments about the rational relations between one's mental states is insufficiently strong to constitute knowledge. As was made clear in Chapter 3, Subsection 2.2 slow-switching brings it about that Jack is wrong about the logical relations between his mental states only insofar as it brings it about that Jack has thoughts involving a concept which he misunderstands, namely, the concept tharthritis. Jack judges that the set of beliefs which he has about the ailment in his ankles is consistent when it is in fact inconsistent. But the set of beliefs which Jack has about the ailment in his ankles is inconsistent only because Jack has thoughts involving the concept tharthritis, a concept which he misunderstands. To this extent, the objection from brute success as formulated in Chapter 3, Subsection 2.2 relies on *Misunderstanding*.

In Chapter 4, Section 3 I argued that there is a strand of thought in Wittgenstein's later work which is at odds with *Misunderstanding*. The strand of thought—which I called *Sufficient Difference*—is that any case in which the difference between the ways in which *A* and *B* are disposed to apply the word '*C*' is such that, first, *A* is disposed to apply '*C*' correctly but *B* is disposed to apply it incorrectly (or vice versa), and second, *B*'s disposition is ultimately to be explained in terms of her being in error about the rules governing the use of '*C*', is a case in which *A* and *B* understand the word '*C*' to mean different things, respectively. But cases in which these two conditions are satisfied are cases in which *B* misunderstands the concept *C* (given my characterisation in Chapter 4, Section 3 of what it is to misunderstand a concept on Wittgenstein's view). So given an externalist view which builds on *Sufficient Difference*, it is not true that a subject can have thoughts involving concepts which she misunderstands. On such a view, *Misunderstanding* is false. Since the objection from brute success as formulated in Chapter 3, Subsection 2.2 relies on *Misunderstanding*, that objection has no purchase against such a view.

A view which builds on *Sufficient Difference* supports a negative claim about Jack's mental states on Twin Earth. The negative claim is that Jack does not use the term 'arthritis' on Twin Earth to express the concept tharthritis. It can only be true that Jack uses the term 'arthritis' on Twin Earth to express the concept tharthritis if subjects can have thoughts involving concepts which they misunderstand. But on a view which builds on *Sufficient Difference*, subjects cannot have thoughts involving concepts which they misunderstand. Do Wittgenstein's views about the individuation of mental content support a corresponding positive claim, a claim about which concept, if any, Jack *does* use the term 'arthritis' to express on Twin Earth? I think so. On Earth, before he is switched, Jack has grasped the meaning-giving characterisations associated with the term 'arthritis'. For this reason, Wittgenstein will likely agree that on Earth, before he is switched, Jack uses 'arthritis' to express the concept arthritis. On Twin Earth, Jack is disposed to respond to questioning about what he means by 'arthritis' in the exact same way as he is disposed to respond to such questioning on Earth. Given the role which Wittgenstein thinks a subject's dispositions to respond to questioning about her meaning play in individuating her mental states, Wittgenstein's

view is likely to be that on Twin Earth, as on Earth, Jack uses the term ‘arthritis’ to express the concept arthritis.

In summary, Wittgenstein’s externalism supports both a positive and a negative claim about the concept which Jack uses the term ‘arthritis’ to express on Twin Earth:

> *Negative claim*: On Twin Earth, Jack does not use the term ‘arthritis’ to express the concept tharthritis.
>
> *Positive claim*: On Twin Earth, Jack uses the term ‘arthritis’ to express the concept arthritis.

Of course, it does not follow, either from the fact that strands in Wittgenstein’s later thought support these claims, or that a view that endorses them is not vulnerable to the objection from brute success (or indeed both of these considerations together) that they are claims which one ought to make all things considered. The strands in Wittgenstein’s later thought might be misguided. There might be good, independent reasons for thinking that the claims are false. In the remainder of this subsection I am going to weigh some of the considerations which one might think militate against the negative and positive claims.

Let us begin with the negative claim. Are there good, independent reasons for thinking that this claim is false, that is, that on Twin Earth Jack *does* use the term ‘arthritis’ to express the concept tharthritis? One might think that the fact that Jack defers to the experts is such a reason. Given enough time in his new location, it is reasonable to assume that the experts to whom Jack defers will come to be the experts on Twin Earth. One might think that this is a good, independent reason for thinking that given enough time on Twin Earth, Jack will come to use the term ‘arthritis’ to express the concept tharthritis.

Whether or not one thinks that the fact that Jack defers to the experts is a consideration which counts against the negative claim is going to depend on one’s views about the significance of deferential relations. *Deference* encapsulates one way of thinking about their significance. On

the way of thinking which is encapsulated in *Deference*—on what we might call the received view—deferring to the experts can bring it about that one uses a word to express that concept which the experts standardly use it to express. But this is not the only possible view. For example, according to the view favoured by Akeel Bilgrami, 'All that deference amounts to … is that [the subject] will change his linguistic behaviour and adopt [the behaviour of the experts]. He will start speaking as they do' (Bilgrami, 2012: 108). On an alternative view of this sort, *Deference* is false; deferring to the experts *cannot* bring it about that one uses a word to express that concept which the experts standardly use it to express. What it can do is bring it about that a subject adjusts his use of a word so that it is in closer accord with the use which is made by the experts.

The existence of alternative, 'deflationary' views of this sort bring into question the claim that the fact that Jack defers to more competent speakers when he uses the term 'arthritis' is a consideration which speaks in favour of his using the term 'arthritis' on Twin Earth to express the concept tharthritis. If one wants to defend this claim then one must demonstrate that there are good, independent reasons for preferring the received view over the alternative view which I have sketched. I am not going to try and reach a conclusion here about whether there are such reasons. So I leave open the question whether there are good, independent reasons for rejecting the negative claim.

Let us turn to the positive claim. Are there good, independent reasons for thinking that this claim is false, that is, that on Twin Earth Jack does *not* use the term 'arthritis' to express the concept arthritis? In 'Individualism and the Mental', Burge raises a number of objections to various methods for reinterpreting the first phase of the arthritis case. I am not going to consider every method or objection which Burge discusses. Instead, I am going to focus on his objections to what I will call the *object-level method*. In general, this method involves defending the claim that a subject who misunderstands the concept *C* possesses some other concept in place of *C*—the concept *C** say, where *C** captures the subject's misconception. It is in accordance with this method to claim that on Twin Earth, Jack uses the term 'arthritis' to express the concept arthritis. Instead of attributing to Jack the concept tharthritis, a concept he misunderstands, we attribute the

concept arthritis, which captures his misconception. I am going to consider Burge's objections to the application of the object-level method to Alf's understanding on Earth, since it is in this connection that Burge originally raises the objections. But my conclusions will apply with equal force to the application of the object-level method to Jack's understanding on Twin Earth.[17]

Burge claims that the object-level method is not a component of ordinary practice and he raises four objections to it which he thinks explains why this is so. The first objection is expressed in the following passage:

> Suppose we are to reinterpret the attribution of [Alf's] erroneous belief that he has arthritis in the thigh. We make up a term 'tharthritis' that covers arthritis and whatever it is he has in his thigh. The appropriate restrictions on the application of this term and of the patient's supposed notion are unclear. Is just any problem in the thigh that the patient wants to call 'arthritis' to count as tharthritis? Are other ailments covered? What would decide? The problem is that there are no recognised standards governing the application of the new term. In such cases, the method is patently *ad hoc*. (Burge, 1979: 94)

In fact, the issue being raised in this passage strikes me as more of a concern than an objection. Suppose that we introduce the term 'tharthritis' to capture Alf's misconception. There are no recognised standards which determine how the term is to be applied or which concept it expresses, for the term has no currency in the language which Alf speaks. But if the introduction of the term is to serve any purpose, then there *must* be some fact of the matter about how it is to be applied or which concept it expresses. If it is not recognised standards which play this determining role, Burge asks, then what does?

Burge is of course right to say that there are no recognised standards in Alf's language which govern the application of 'tharthritis'. But established standards of application are not the only things which could play the determining role which Burge has in mind. Presumably, facts about Alf—about the way in which he is disposed to use the term 'arthritis', about the way he is

[17] Burge thinks that the object-level method is often invoked in conjunction with another method, which he calls the *metalinguistic reinterpretation method* (Burge, 1979: 96). This method counts the error of subjects who misunderstand a particular concept as metalinguistic. For example, according to this method, Alf's error is not that he believes that arthritis can occur in the thigh, but that he believes that the word 'arthritis' applies in English to an ailment in the thigh. I am not going to consider Burge's objections to this method, since my interest is only in defending the claim that Jack has thoughts involving the concept arthritis (and not further claims about what exactly his error consists in).

disposed to respond to questioning about what he means by 'arthritis', and so on—are what determine how the term 'tharthritis' ought to be applied and which concept the term expresses on Alf's lips. Of course, when it is clear that someone has misunderstood a particular term, we do not, as a rule, take the time to investigate their usage or question them thoroughly about their meaning. We usually correct their misunderstanding and move on. But it does not follow that there is no fact of the matter about what the person means by the term in question or which concept they use it to express.

Burge's second objection is as follows:

> The method's willingness to invoke new terminology whenever conceptual error or partial understanding occurs is *ad hoc* in another sense. It proliferates terminology without evident theoretical reward. We do not engender better understanding of the patient by inventing a new word and saying that he thought (correctly) that tharthritis can occur outside joints. It is simpler and equally informative to construe him as thinking that arthritis may occur outside joints. When we are making other attributions that do not directly display the error, we must simply bear the deviant belief in mind, so as not to assume that all of the patient's inferences involving the notion would be normal. (Burge, 1979: 94)

But this objection seems misplaced. Burge seems to assume that the object-level method advocates the introduction of a novel term to cover Alf's misconception. In fact, this is not—or at least need not be—part of the object-level method. The proponent of the object-level method can agree that it is appropriate in the present case to attribute to Alf beliefs about arthritis (indeed, they can agree that, for the very reasons that Burge points out, novel terms should not, as a general rule, be introduced to cover subjects' misconceptions).[18] The claim which they challenge is that such attributions ought to be taken literally. The proponent of the object-level method might challenge this claim by taking the line of argument which I ascribed to Glock, Preston and Davidson in Chapter 4, Section 3. They might point out that in cases where it is clear that a subject misunderstands a concept, we are usually prepared to revise our original ascriptions when asked to specify the content of the subject's beliefs exactly.

[18] See Nordby (2004: 59) on this point.

Burge's third objection is as follows:

> The method of object-level reinterpretation often fails to give a plausible account of the evidence on which we base mental attributions. When caught in the sorts of errors we have been discussing, the subject does not normally respond by saying that his views had been misunderstood … In such cases, the subject will ordinarily give no evidence of having maintained a true object-level belief. In examples like ours, he typically admits his mistake, changes his views, and leaves it at that. Thus the subject's own behavioural dispositions and inferences often fail to support the method. (Burge, 1979: 94-95)

I do not think that the fact that Alf responds in the way that Burge describes—by admitting his mistake, changing his views and leaving it at that—is evidence against his believing that his tharthritis has spread to his thigh. As Glock and Preston make clear, all that Alf's response shows is that subjects who are deferential usually adjust their use of a term when it is clear that it is in conflict with the use which is made of the term by the experts. (Glock & Preston, 1995: 519-520) Given that Alf is deferential, I see no reason to expect that he would respond to the doctor's correction by claiming that he has been misunderstood, even if we agreed that the belief he expressed *did* involve the concept tharthritis. After all, he aims to use the term 'arthritis' to express the concept which subjects like his doctor standardly use it to express.

Burge's fourth objection is as follows:

> On this new interpretation, the patient is right in thinking that he has tharthritis in the ankle and wrists. His belief that it has lodged in the thigh is true. His fear is realised. But these attributions are out of keeping with the way we do and should view his actual beliefs and fears. His belief is not true, and his fear is not realised. He will be relieved when he is told that one cannot have arthritis in the thigh …When told that arthritis cannot occur in the thigh, the patient does not decide that his fears were realised, but that perhaps he should not have had those fears. (Burge, 1979: 95)

I take it that the basic thought is as follows. If Alf really does believe that his tharthritis has spread to his thigh, then his belief is true. His fear is realised. But Alf will be relieved when his doctor tells him that one cannot have arthritis in one's thigh. If Alf's fear is realised, then how are we to account for his reacting in this way?

Of course, Alf may feel *some* relief when he learns that one cannot have arthritis in one's thigh. But the news is also likely to leave Alf unsatisfied. After all, the pains in his leg indicate

that there is *something* wrong with his thigh. If it isn't arthritis, then what is it? Moreover, when he learns that he has a rheumatoid ailment in his thigh he may well feel as though his fear *has* been realised. If accounting for the feeling of relief which Alf experiences when he learns that one cannot have arthritis in one's thigh is a problem for the proponent of the object-level method, then presumably the opponent of that method faces a similar problem accounting for the fact that Alf will likely react in this way to the news that he has a rheumatoid ailment in his thigh.

I conclude that none of the four objections which Burge raises against the object-level method counts against that method. None counts against the application of that method to Alf, that is, against the claim that Alf uses the term 'arthritis' on Earth to express the concept tharthritis. Consequently, none counts against the application of that method to Jack, that is, against the claim that Jack uses the term 'arthritis' on Twin Earth to express the concept arthritis. Earlier in this chapter, I suggested that Wittgenstein's externalism supports the positive claim that on Twin Earth, Jack uses the term 'arthritis' to express the concept arthritis.. We have seen that this claim stands in the face of Burge's objections.

2.2 Wittgenstein, the Objection from Brute Success and *False Belief*

In Chapter 3, Subsection 3.3 I considered whether the objection from brute success applies if *Misunderstanding* is substituted with either *Lay Knowledge* or *False Belief*, given a Burgean framework for thinking about the individuation of mental states. I noted that whether the objection applies if *Misunderstanding* is substituted with *False Belief*—the claim that if a proposition, *p*, is a meaning-giving characterisation associated with *p*, then an otherwise competent subject who believes that *p* is false can still have thoughts involving the concept *C*—is likely to depend on Burge's considered view about slow-switching cases involving subjects who possess relevant discriminatory knowledge and come to have extensive, direct causal contact with Twin Earth samples (but do not come to rely on Twin Earth communal standards). If it is Burge's considered view that slow-switching will typically result in such subjects acquiring a second, distinct concept,

then the objection from brute success *does* apply. In this case, slow-switching can bring it about that an otherwise competent subject, like Jill, who believes that 'Sofas are items of furniture meant primarily for sitting' expresses a falsehood about sofas, nevertheless has thoughts involving the concept sofa. And as I demonstrated in Chapter 3, Subsection 3.3, if slow-switching can bring this about, then it can bring about the sorts of errors in Jill's judgments about the logical relations between her mental states on which the objection from brute success depends. If, however, it is Burge's considered view that in such cases the subjects' concepts will remain constant, then the objection does not apply (for in this case, slow-switching cannot bring about the requisite sort of errors).

The discussion in Chapter 5, Subsection 2.1 focused on whether the objection from brute success has application in Jack's case, given a Wittgensteinian framework for thinking about the individuation of mental states. I defended the view that it does not, precisely because there are strands of thought within that framework which are at odds with *Misunderstanding*. Reflection on the discussion in Chapter 3, Subsection 3.3 may prompt us to ask whether, given a Wittgensteinian framework, the objection from brute success has application if *Misunderstanding* is substituted with *False Belief*.

I do not have the space to consider this question in all its ramifications here. But I want to note two considerations which suggest that the answer is 'No'. The first consideration relates to Wittgenstein's views about the effects of slow-switching on subjects' mental states in the relevant sort of cases. In Chapter 4, Section 1 I considered what Wittgenstein might say about how, if at all, slow-switching will affect the mental states of subjects like Sally, subjects who possess relevant discriminatory knowledge and come to have extensive, direct causal contact with Twin Earth samples (but do not come to rely on Twin Earth communal standards). I concluded that Wittgenstein's view is likely to be that slow-switching will not bring about any change in such cases. Given Wittgenstein's views about the individuating role of a subject's dispositions to respond to questioning about her meaning, Wittgenstein is likely to take the view that Sally uses the term 'money' to express the same concept on Twin Earth as she does on Earth. If this is right,

then the objection from brute success probably does not apply if *Misunderstanding* is substituted with *False Belief*, given a Wittgensteinian framework for thinking about the individuation of mental states. If slow-switching is not likely to bring about a change in the mental states of subjects like Sally, then it is going to be difficult to imagine cases in which slow-switching brings it about that an otherwise competent subject has thoughts involving the concept $\underline{C}$ even though she has a non-standard theory about *C*'s. If slow-switching is not likely to bring this about, then it probably cannot bring about the sorts of errors in one's judgments about the logical relations between one's mental states on which the objection from brute success depends.

The second consideration is a strand of thought in Wittgenstein's later work which is prima facie at odds with the idea that one can believe, of a proposition which is meaning-giving, that it is false. Clearly, if Wittgenstein does not accept this idea, then Jill cannot believe that the proposition that sofas are items of furniture meant primarily for sitting is false, given a Wittgensteinian framework (given that the proposition is indeed a meaning-giving characterisation associated with 'sofa'). But if Jill cannot believe that this proposition is false, then slow-switching cannot bring it about that she does. And if slow-switching cannot bring this about, then it cannot bring about the sorts of errors in Jill's judgments about the logical relations between her mental states on which the objection from brute success depends.

The strand of thought is a claim about the negation of grammatical propositions, which are the analogues to Burge's meaning-giving characterisations. Insofar as grammatical propositions determine what words, expressions and propositions can be used to say, their negations do not, on Wittgenstein's view, express genuine possibilities. Take, for example, the proposition 'Red is a colour'. On Wittgenstein's view, the negation of this proposition—'Red is *not* a colour'—is not strictly speaking false, but nonsensical. It is strictly speaking nonsense because 'it employs expressions blatantly contrary to the rules for their use that are given by the grammatical proposition denied' (Backer and Hacker, 1985: 19). But if its negation is nonsense, then one cannot accurately be described as believing that the proposition 'Red is a colour' is false. If 'Not *p*' is nonsense, then one cannot believe that not *p*. Of course, it is possible that someone might not

be willing to assent to the proposition that *p*, even though they are aware that the relevant experts would assent to it. But if 'Not *p*' does not describe a genuine possibility, it is not possible for someone to believe that not *p*. If 'Not *p*' does not describe a genuine possibility, then it is more accurate to describe someone who is not willing to assent to that proposition but who is otherwise competent, as operating with a different set of concepts altogether. It seems more accurate to describe someone who is not willing to assent to the proposition 'Red is a colour' but is otherwise competent as using the term 'red' to express some concept other than the concept red.

And indeed, often this *is* how Wittgenstein describes such people. At numerous points throughout his later work, Wittgenstein imagines scenarios in which we encounter people who apparently do not accept certain propositions which have the status of grammatical propositions within our language. His point in many of these cases is that these people mean something different by their words or are operating with a different set of concepts. For example, in *Remarks on the Foundations of Mathematics* Wittgenstein imagines our encountering people who find it natural to charge for a pile of wood a price proportionate to the area the pile covers (as opposed to, say, its cubic measurement). Wittgenstein goes on to imagine trying to show these people that you do not necessarily buy more wood if you buy a pile covering a bigger area:

> I should, for instance, take a pile which was small by their ideas and, by laying the logs around, change it into a 'big' one. This *might* convince them—but perhaps they would say: "Yes, now it's a *lot* of wood and costs more"—and that would be the end of the matter.—We should presumably say in this case: they simply do not mean the same by "a lot of wood" and "a little wood" as we do; and they have a quite different system of payment from us. (RFM I, §149)

The people Wittgenstein is imagining apparently do not accept propositions which Wittgenstein would say have the status of grammatical propositions in our language, propositions like 'Pile *A* contains more wood than pile *B* if its cubic measurement is larger'. Wittgenstein's conclusion is not that these people mean by the expression 'more wood' what we mean, but that they are using the expression wrongly. His conclusion is that these people mean something different by the expression 'more wood'.

Unfortunately, I do not have the space to explore this strand of thought in more detail. Suffice it to say that like the first consideration I mentioned, it is evidence that the objection from brute success does not apply if *Misunderstanding* is substituted with *False Belief*, given a Wittgensteinian framework.

CONCLUSION

In Subsection 1.1 I considered how Wittgenstein might respond to slow-switching objections. I concluded that his response is likely to be similar in form to Burge's response. Wittgenstein is likely to be sympathetic to the claim that slow-switching cannot bring it about that one's judgments about one's mental states are subject to error. He is also likely to reject the claim that one's warrant for judgments about one's mental states depends on one's being able to discriminate one's thoughts from relevant alternatives. I went on in Subsection 1.2 to consider whether Wittgenstein offers a positive account of one's warrant for judgments about one's mental states and if so how that account ought to be understood. I defended John McDowell's interpretation of Wittgenstein's views on our warrant for self-knowledge. In Subsection 2.1 I defended the view that the objection from brute success has no purchase against Wittgenstein's externalism. The objection from brute success depends on *Misunderstanding*, on the claim that a subject can have thoughts involving concepts which she misunderstands. But an externalism which builds on the strands in Wittgenstein's later thought which I discussed in Chapter 4 will reject *Misunderstanding*. An externalism of this sort supports both a negative and a positive claim about the concept which Jack uses the term 'arthritis' to express on Twin Earth. I went on to weigh considerations which one might think militate against these claims. I left open the question whether there are good, independent reasons for rejecting the negative claim. I argued, however, that none of the objections which Burge raises against the object-level method militate against the positive claim. In Subsection 2.2 I considered whether the objection from brute success has

purchase if *Misunderstanding* is substituted with *False Belief*, given a Wittgensteinian framework. I noted two considerations which support the view that it does not.

CONCLUSION

The primary aim of this thesis has been to assess the prospects for reconciling content-externalism with self-knowledge within a Burgean and a Wittgensteinian framework, respectively. I have argued that a tension between the basic externalist intuition and a crucial guiding intuition about self-knowledge arises within a Burgean framework which does not arise within a Wittgensteinian framework.

In Chapter 1, I introduced two intuitions: first, what I called externalism's driving intuition, the thought that at least some types of mental states are at least partly individuated by the relations in which a speaker stands to a context; and second, what I called the driving intuition about self-knowledge, the thought that we normally know groundlessly what we are thinking. In Chapter 2, I explained why some philosophers have thought that there is a tension between these two intuitions. The discussion focused on one particular way in which this thought has been developed in the literature. The way of developing the thought on which I focused makes appeal to slow-switching objections. I concluded that these objections have prima facie plausibility against two of the thought experiments which Burge offers in support of externalism's driving intuition, namely, the arthritis and water cases. Burge's response to slow-switching objections, which I went on to outline, culminates in the claim that slow-switching cannot undermine the knowledgeability of one's mental states. On Burge's view, slow-switching cannot bring it about, either that one's judgments about one's mental states are subject to error, or that those judgments, although true, do not constitute knowledge.

In Chapter 3, I raised the objection from brute success. This objection shows that given Burge's views about the way in which mental content is individuated, slow-switching can undermine knowledgeability of a subject's epistemic reasons. I argued that this is a troubling result, given Burge's views about the sorts of things that epistemic reasons are. On Burge's view, epistemic reasons are rational relations between mental states. If slow-switching can undermine knowledgeability of one's epistemic reasons, then it can undermine knowledgeability of the rational relations between mental states. But the thought that the knowledgeability of the rational

relations between one's mental states is sensitive to changes in one's context in this way is at odds with our intuitive picture of self-knowledge.

I went on to argue that the objection from brute success as formulated in Chapter 3 does not have purchase against a Wittgensteinian framework. The objection as formulated in Chapter 3 depends on *Misunderstanding*. In Chapter 4, I argued that there is a strand in Wittgenstein's later thought which is at odds with *Misunderstanding*, namely, the thought that any case in which the difference between the ways in which *A* and *B* are disposed to apply the word '*C*' is such that, first, *A* is disposed to apply '*C*' correctly but *B* is disposed to apply it incorrectly (or vice versa), and second, *B*'s disposition is ultimately to be explained in terms of confusion about the rules governing the use of '*C*', is a case in which *A* and *B* understand the word '*C*' to mean different things, respectively. I called this strand of thought *Sufficient Difference*. For the reasons set out in Chapter 4, *Sufficient Difference* is at odds with the idea that a subject can have thoughts involving concepts which she misunderstands. On an externalist view which builds on *Sufficient Difference*, *Misunderstanding* is false. Since the objection from brute success as formulated in Chapter 3 relies on *Misunderstanding*, that formulation of the objection has no purchase against such a view. In Chapter 5, Subsection 2.2 I noted two considerations which support the view that a Wittgensteinian framework is not vulnerable to certain other formulations of the objection from brute success.

What ought we to conclude about the relative merits of the Burgean and Wittgensteinian frameworks? I have argued that given Burge's views about the individuation of mental content, slow-switching can undermine knowledgeability of the rational relations between a subject's mental states. But as I made clear in Chapter 3, Subsection 2.1 I do not consider this result to automatically count against the Burgean framework. The objection from brute success relies on the claim that changes in one's context cannot undermine the knowledgeability of the rational relations between one's mental states. This claim has strong intuitive appeal. But I do not think it has the status of a datum. If I am right that there is a tension between this claim and Burge's views about the individuation of mental content, it does not immediately follow that we have a reason to

revise the latter. Rather, we should take seriously the possibility that we have a reason to revise our intuitions about our capacity to know the rational relations between our mental states.

I do not have the space to investigate this possibility here. I am, therefore, not in a position to defend a view about whether Burge ought to respond to the objection from brute success by challenging the intuition that slow-switching cannot undermine the knowledgeability of the rational relations between one's mental states. But suppose that someone *did* wish to defend this view. What would an effective defence involve? At a minimum, it would explain why our intuitions about the knowledgeability of the rational relations between our mental states are less justified (or at any rate, should be given less weight in our thinking about this issue) than Burge's externalist views about the individuation of mental content. Such a defence may begin by reminding us that we should expect *some* conflict between externalist theses about the individuation of mental content and our intuitive picture of self-knowledge, given—as was noted in Chapter 1, Subsection 2.1—the extent to which that picture has informed, and in turn has been informed by, internalist intuitions about the mind. But it must go beyond this observation. Some account must be given which explains why in the case of this particular conflict, it is Burge's views about the individuation of mental content which are to be preferred. I am not claiming that such an account cannot be given. But if, by challenging the intuition that the knowledgeability of the rational relations between one's mental states is not sensitive to changes in one's context, Burge wishes to convince us that the objection from brute success presents no problems for his view, then the onus is on him to provide it.

The objection from brute success does not automatically count against Burge's view because Burge could respond to the objection by challenging the intuition about the knowledgeability of the rational relations between one's mental states on which the objection relies. Suppose, however, that it should turn out that a challenge of this sort fails. Suppose it should turn out that the objection from brute success *does* count against the Burgean framework. Still, it need not count *decisively*. There may be further considerations which count in favour of the Burgean framework. I am not claiming to have offered an *exhaustive* comparative analysis of

the Burgean and Wittgensteinian frameworks. I am certainly not ruling out the possibility that such an analysis will highlight differences which give us a reason to prefer the former. The fact that given Burge's views about the individuation of mental states, slow-switching can undermine knowledgeability of the rational relations between one's mental states must be weighed in a final evaluation of the two frameworks. But it need not weigh decisively.

These considerations should not, however, obscure this study's accomplishments. Whilst a final assessment of the implications of the objection from brute success for Burge's view must be postponed, *that* Burge's view is vulnerable to the objection is itself an interesting result. It advances an understanding of the implications of Burge's externalism, of what exactly Burge's views about the individuation of mental content commit him to. Moreover, I hope that the supporting discussion has facilitated a deepened understanding of the various components of Burge's externalism, of features of his philosophy considered more generally—for example, his views about self-knowledge, epistemic reasons, and so on—and the way in which these components and features hang together.

In addition, this thesis has revealed various substantive and hitherto unexplored similarities and differences between Burge and Wittgenstein. With regard to similarities, both Burge and Wittgenstein think that our ordinary practices of belief attribution are instructive when thinking in a philosophical context about whether a subject is in a particular mental state (although as I made clear in Chapter 4, Section 3, there is a difference in the way in which Wittgenstein and Burge conceive of the authority of standard usage). Both endorse accounts of one's warrant to judgments about one's mental states according to which, in order for those judgments to be warranted, one's judgments about one's reasons and reasoning must normally be correct.

Of course, our discussion has also uncovered several substantive differences. I have already noted one of these differences in this conclusion. Burge accepts, whereas there is evidence that Wittgenstein would reject, *Misunderstanding*, the claim that a subject can have thoughts involving concepts which she misunderstands.

A second substantive difference concerns the ways in which Burge and Wittgenstein develop the basic externalist thought. When discussing the role which external factors can play in individuating a subject's mental states, Burge focuses on relations of deference between a subject and a community of speakers and direct causal relations connecting a subject to an objective subject matter. Wittgenstein, in contrast, emphasises a different set of considerations, for example, facts about the immediate context and facts about the subject herself. Those facts about the subject which Wittgenstein thinks are of particular relevance include facts about the subject's abilities, about the things that the subject said or did at the time or earlier, about the subjects dispositions, and so on. As I made clear in Chapter 1, Subsection 1.3, the notion that, for example, facts about a subject's dispositions might play an individuating role with respect to her mental states is a distinctly externalist thought, on Wittgenstein's view. On that view, it is only under an intentional description that a subject's dispositions play an individuating role with respect to her mental states. But whether a particular disposition is correctly intentionally characterised in a given way will itself, Wittgenstein thinks, depend on facts about the broader context.

In drawing attention to these points of difference, the discussion has traced two contrasting ways in which externalism's driving intuition might be elaborated.

BIBLIOGRAPHY

Audi, R., (2001) ‘An Internalist Theory of Normative Grounds’, in *Philosophical Topics*, Vol. 29, Nos. 1 & 2, pp. 19-46

Baker, G. & Hacker, P. M. S., (1983) *Wittgenstein,* Meaning *and Understanding: Essays on the Philosophical Investigations, Volume 1* (Oxford: Basil Blackwell)

_____, (1985) *Wittgenstein, Rules, Grammar and Necessity*: *Essays on the Philosophical Investigations, Volume 2* (Oxford: Basil Blackwell)

_____, (1990) ‘Malcolm on Language and Rules’, in *Philosophy*, Vol. 65, No. 252, pp. 167-179

Bar-On, D., (2004) *Speaking My Mind: Expression and Self-Knowledge* (Oxford: Clarendon Press)

Bernecker, S., (1996) ‘Externalism and the Attitudinal Component of Self-Knowledge’, in *Noûs*, Vol. 30, No. 2, pp. 262-275

Bilgrami, A., (2012) ‘Why Meaning Intentions are Degenerate’, in C. Wright and A. Coliva (eds.), *Mind, Meaning, and Knowledge: Themes From the Philosophy of Crispin Wright* (Oxford: Oxford University Press), pp. 96-124

Boghossian, P. A., (1992) ‘Externalism and Inference’, in *Philosophical Issues*, Vol. 2, Rationality in Epistemology, pp. 11-28

_____, (1994) ‘The Transparency of Mental Content’, in *Philosophical Perspectives*, Vol. 8, Logic and Language, pp. 33-50

_____, (1998) ‘Content and Self-Knowledge’, in P. Ludlow and N. Martin (eds.), *Externalism and Self-Knowledge* (Stanford: CSLI Publications), pp. 149-173

Bonjour, L., (2002) ‘Internalism and Externalism’, in P. Moser (ed.), *The Oxford Handbook of Epistemology* (Oxford: Oxford University Press), pp. 234-263

Broome, J., (2013) *Rationality Through Reasoning*, (Oxford: Blackwell Publishing)

Brown, J., (1999) ‘Boghossian on Externalism and Privileged Access’, in *Analysis*, Vol. 59, No. 1, pp. 52-59

_____, (2000) ‘Critical Reasoning, Understanding and Self-Knowledge’, in *Philosophy and Phenomenological Research*, Vol. 61, No. 3, pp. 659-676

_____, (2004) *Anti-individualism and Knowledge* (Cambridge: MIT Press)

_____, (2009) ‘Semantic Externalism and Self-knowledge’, in B. McLaughlin, A. Beckermann & S. Walter (eds.), *The Oxford Handbook of Philosophy of Mind* (Oxford: Clarendon Press), pp. 767-780

Budd, M., (1984) ‘Wittgenstein on Meaning, Interpretation and Rules’ in *Synthese*, Vol. 58, No. 3, Essays on Wittgenstein’s Later Philosophy, pp. 303-323

_____, (1989) *Wittgenstein’s Philosophy of Psychology* (London: Routledge)

Burge, T., (1977) ‘Belief *De Re*’, in *The Journal of Philosophy*, Vol. 74, No. 6, pp. 338-362

_____, (1979) ‘Individualism and the Mental’, in *Midwest Studies in Philosophy*, Vol. 4, Iss. 1, pp. 73-121

_____, (1982) ‘Two Thought Experiments Reviewed’, in *Notre Dame Journal of Formal Logic*, Vol. 23, No. 3, pp. 284-293

_____, (1986a) ‘Individualism and Psychology’, in *The Philosophical Review*, Vol. 95, No. 1, pp. 3-45

_____, (1986b) ‘Intellectual Norms and Foundations of Mind’, in *The Journal of Philosophy*, Vol. 83, No. 12, pp. 697-720

_____, (1988) ‘Individualism and Self-Knowledge’, in *The Journal of Philosophy*, Vol. 85, No. 11, Eighty-Fifth Annual Meeting American Philosophical Association, Eastern Division, pp. 649-663

_____, (1993) ‘Content Preservation’, in *The Philosophical Review*, Vol. 102, No. 4, pp. 457-488

_____, (1996) ‘Our Entitlement to Self-Knowledge’, in *Proceedings of the Aristotelian Society*, New Series, Vol. 96, pp. 91-116

_____, (1998) ‘Reason and the First Person’, in C. Wright, B. Smith and C. Macdonald (eds.), *Knowing Our Own Minds* (Clarendon Press: Oxford), pp. 243-270

_____, (1999) ‘A Century of Deflation and a Moment about Self-Knowledge’, in *Proceedings and Addresses of the American Philosophical Association*, Vol. 73, No. 2, pp. 25-46

_____, (2003a) ‘Davidson and Forms of Anti-Individualism: Reply to Hahn’, in M. Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: The MIT Press), pp. 347-361

_____, (2003b) ‘Mental Agency in Authoritative Self-Knowledge: Reply to Kobes’, in M. Hahn & T. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: MIT Press), pp. 417-433

_____, (2003c) ‘Perceptual Entitlement’, in *Philosophy and Phenomenological Research*, Vol. 67, No. 3, pp. 503-548

_____, (2003d) ‘The Thought Experiments: Reply to Donnellan’ in M. Hahn & T. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: MIT Press), pp. 363-369

_____, (2005) ‘Frege on Sense and Linguistic Meaning’ in T. Burge (ed.), *Truth, Thought, Reason: Essays on Frege* (Oxford: Oxford University Press), pp. 242-269

_____, (2007a) ‘Cartesian Error and the Objectivity of Perception’ in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 192-207

_____, (2007b) ‘Introduction’, in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 1-31

_____, (2007c) ‘Other Bodies’, in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 82-99

_____, (2007d) ‘Postscript to “Belief *De Re*”, in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 65-81

_____, (2007e) ‘Postscript to “Individualism and the Mental”, in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 151-181

_____, (2007f) ‘Wherein is Language Social?’, in T. Burge (ed.), *Foundations of Mind: Philosophical Essays, Volume 2* (Oxford: Clarendon Press), pp. 275-290

_____, (2010) *Origins of Objectivity* (Oxford: Clarendon Press)

_____, (2013a) ‘Concepts, Conceptions, Reflective Understanding’, in T. Burge (ed.), *Cognition Through Understanding: Philosophical Essays, Volume 3* (Oxford: Oxford University Press), pp. 521-533

_____, (2013b) ‘Epistemic Warrant’, in T. Burge (ed.), *Cognition Through Understanding: Philosophical Essays, Volume 3* (Oxford: Oxford University Press), pp. 489-507

_____, (2013c) ‘Introduction’, in T. Burge (ed.), *Cognition Through Understanding: Philosophical Essays, Volume 3* (Oxford: Oxford University Press), pp. 1-52

_____, (2013d) ‘Memory and Self-Knowledge’, in T. Burge (ed.), *Cognition Through Understanding: Philosophical Essays, Volume 3* (Oxford: Oxford University Press), pp. 88-103

_____, (2013e) ‘Self and Self-Understanding: The Dewey Lectures (2007, 2011)’, in T. Burge (ed.), *Cognition Through Understanding: Philosophical Essays, Volume 3* (Oxford: Oxford University Press), pp. 140-226

_____, (2013f) ‘Some Remarks on Putnam’s Contributions to Semantics’ in *Theoria*, Vol. 79, Iss. 3, pp. 229-241

Butler, K., (1997) ‘Externalism, Internalism, and Knowledge of Content’, in *Philosophy and Phenomenological Research*, Vol. 57, No. 4, pp. 773-800

Child, W., (2002) ‘Wittgenstein’s Externalism and Modern Externalism’, in *Filosoficky Casopis*, Vol. 50, Iss. 3, pp. 459-477

_____, (2006) ‘Wittgenstein’s Externalism: Context, Self-Knowledge and the Past’, in T. Marvan (ed.), *What Determines Content?: The Internalism/Externalism Dispute* (Newcastle: Cambridge Scholars Press), pp. 198-220

_____, (2010) ‘Wittgenstein’s Externalism’, in D. Whiting (ed.), *The Later Wittgenstein on Language* (London: Palgrave Macmillan), pp. 63-80

_____, (2011) *Wittgenstein* (London: Routledge)

Clark, A. & Chalmers, D., (1998) ‘The Extended Mind’, in *Analysis*, Vol. 58, No. 1, pp. 7-19

Crane, T., (1991) ‘All the Difference in the World’, in *The Philosophical Quarterly*, Vol. 41, No. 162, pp. 1-25

_____, (2010) ‘Wittgenstein and Intentionality’, in *The Harvard Review of Philosophy*, Vol. 17, Iss. 1, pp. 88-104

Davidson, D., (2001a) ‘Epistemology Externalised’, in D. Davidson (ed.), *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press), pp. 193-204

_____, (2001b) ‘First Person Authority’, in D. Davidson (ed.), *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press), pp. 3-14

_____, (2001c) ‘Knowing One’s Own Mind’, in D. Davidson (ed.), *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press), pp. 15-38

Donnellan, K., (2003) ‘Burge’s Thought Experiments’ in M. Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: The MIT Press), pp. 59-75

Dretske, F., (1970) ‘Epistemic Operators’, in *The Journal of Philosophy*, Vol. 67, No. 24, pp. 1007-1023

Falvey, K. & Owens, J., (1994) ‘Externalism, Self-Knowledge, and Skepticism’, in *The Philosophical Review*, Vol. 103, No. 1, pp. 107-137

Forster, M., (2004) *Wittgenstein on the Arbitrariness of Grammar* (Princeton: Princeton University Press)

Georgalis, N., (1999) ‘Rethinking Burge’s Though Experiment’, in *Synthese*, Vol. 118, No. 2, pp. 145-164

Gibbons, J., (2013) *The Norm of Belief* (Oxford: Oxford University Press)

Glock, H.J., (1996) *A Wittgenstein Dictionary* (Oxford: Blackwell Publishers)

_____, (2009) ‘Concepts, Conceptual Schemes and Grammar’ in *Philosophia*, 37, pp. 653-668

_____, (2010) ‘Wittgenstein on concepts’, in A. Ahmed (ed.), *Wittgenstein’s* Philosophical Investigations: *A Critical Guide* (Cambridge: Cambridge University Press), pp. 88-108

Glock, H. J. & Preston, J. M., (1995) ‘Externalism and First-Person Authority’, in *The Monist*, Vol. 78, No. 4, pp. 515-533

Glüer, K., Wikforss, Å., (2010) ‘Es braucht die Regel nicht: Wittgenstein on Rules and Meaning’, in D. Whiting (ed.), *The Later Wittgenstein on Language* (London: Palgrave Macmillan), pp. 148-166

Goldberg, S., (2006) ‘Brown on Self-Knowledge and Discriminability’ in *Pacific Philosophical Quarterly*, Vol. 87, Iss. 3, pp. 301-314

Goldman, A., (1976) ‘Discrimination and Perceptual Knowledge’, in *The Journal of Philosophy*, Vol. 73, No. 20, pp. 771-791

Grice, H. P., (1957) ‘Meaning’, in *The Philosophical Review*, Vol. 66, No. 3, pp. 377-388

_____, (1968) ‘Utterer’s Meaning, Sentence-Meaning, and Word-Meaning’, in *Foundations of Language*, Vol. 4, No. 3, pp. 225-242

Hacker, P., (1990) *Wittgenstein, Meaning and Mind* (Oxford: Basil Blackwell)

_____, (1996) *Wittgenstein's Place in Twentieth-Century Analytic Philosophy* (Oxford: Blackwell Publishers)

_____, (2010) 'Meaning and Use', in D. Whiting (ed.), *The Later Wittgenstein on Language* (London: Palgrave Macmillan), pp. 26-44

Hahn, M., (2003) 'When Swampmen Get Arthritis: "Externalism" in Burge and Davidson', in M. Hahn & B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge: The MIT Press), pp. 29-58

Heil, J., (1988) 'Privileged Access', *Mind*, Vol. 97, No. 386, pp. 238-251

Horwich, P., (2010) 'Wittgenstein's Definition of 'Meaning' as 'Use', in D. Whiting (ed.), *The Later Wittgenstein on Language* (London: Palgrave Macmillan), pp. 17-25

Kripke, S., (1982) *Wittgenstein on Rules and Private Language* (Oxford: Basil Blackwell)

Langton, R., & Lewis, D., (1998) 'Defining 'Intrinsic'', in *Philosophy and Phenomenological Research*, Vol. 58. No. 2, pp. 333-345

Lewis, D., (1983) 'Extrinsic Properties', *Philosophical Studies*, Vol. 44, Iss. 2, pp. 197-200

_____, (1996) 'Elusive Knowledge', *Australasian Journal of Philosophy*, Vol. 74, No. 4, pp. 549-567

Ludlow, P., (1995) 'Externalism, Self-Knowledge, and the Prevalence of Slow Switching', in *Analysis*, Vol. 55, No. 1, pp. 45-49

Malcolm, N., (1986) 'Following a Rule', in *Wittgenstein: Nothing is Hidden* (Oxford: Basil Blackwell), pp. 154-181

_____, (1995) 'Wittgenstein on Language and Rules', in G. H. von Wright (ed.), *Wittgensteinian Themes* (Ithaca: Cornell University Press), pp. 145-171

McDowell, J., (1982) 'Criteria, defeasibility, and knowledge' in *Proceedings of the British Academy*, Vol. 68, pp. 455-479

_____, (1984) 'Wittgenstein on Following a Rule', in *Synthese*, Vol. 58, pp. 325-363

_____, (1998a) 'Intentionality and Interiority in Wittgenstein', in *Mind, Value & Reality* (Cambridge: Harvard University Press), pp. 297-321

_____, (1998b) 'Meaning and Intentionality in Wittgenstein's Later Philosophy', in *Mind, Value & Reality* (Cambridge: Harvard University Press), pp. 263-278

_____, (1998c) ‘Response to Crispin Wright’, in C. Wright, B. Smith and C. Macdonald (eds.), *Knowing Our Own Minds* (Clarendon Press: Oxford), pp. 47-62

McGinn, C., (1989) *Mental Content* (Oxford: Basil Blackwell)

McKinsey, M., (1991) ‘Anti-Individualism and Privileged Access’ in *Analysis*, Vol. 51, No. 1, pp. 9-16

Moran, R., (1997) ‘Self-Knowledge: Discovery, Resolution, and Undoing’, in *European Journal of Philosophy*, Vol. 5, Iss. 2, pp. 141-161

_____, (2001) *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press)

Norby, H., (2004) ‘Incorrect understanding and concept possession’, in *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, Vol. 7, Iss. 1, pp. 55-70

Owens, J., (2003) ‘Externalism, Davidson, and Knowledge of Comparative Content’, in S Nuccetelli (ed.), *New Essays on Semantic Externalism and Self-Knowledge* (Cambridge: The MIT Press), pp. 201-218

Pettit, P., (1983) ‘Wittgenstein, Individualism and the Mental’, in *Epistemology and the Philosophy of Science: Proceedings of the Seventh International Symposium* (Vienna: Holder-Pichler-Tempsky), pp. 446-455

Putnam, H., (1975a) ‘Is Semantics Possible?’ in H. Putnam (ed.), *Mind, Language and Reality: Philosophical Papers Volume 2* (Cambridge: Cambridge University Press), pp. 139-152

_____, (1975b) ‘The Meaning of ‘Meaning’’ in H. Putnam (ed.), *Mind, Language and Reality: Philosophical Papers Volume 2* (Cambridge: Cambridge University Press), pp. 215-271

Sawyer, S., (1999) ‘An Externalist Account of Introspective Knowledge’, in *Pacific Philosophical Quarterly*, Vol. 80, Iss. 4, pp. 358-378

Schiffer, S., (1992) ‘Boghossian on Externalism and Inference’, in *Philosophical Issues*, Vol. 2, Rationality in Epistemology, pp. 29-37

Schroeder, M., (2008) ‘Having Reasons’, in *Philosophical Studies*, Vol. 139, Iss. 1, pp. 57-71

Shoemaker, S., (1988) ‘On Knowing One’s Own Mind’, in *Philosophical Perspectives*, Vol. 2, *Epistemology*, pp. 183-209

Vahid, H., (2003) ‘Externalism, Slow Switching and Privileged Self-Knowledge’, in *Philosophy and Phenomenological Research*, Vol. 66, No. 2, pp. 370-388

Warfield, T. A., (1992) ‘Privileged Self-Knowledge and Externalism are Compatible’, in *Analysis*, Vol. 52, No. 4, pp. 232-237

_____, (1997) ‘Externalism, Privileged Self-Knowledge, and the Irrelevance of Slow Switching’, in *Analysis*, Vol. 57, No. 4, pp. 282-284

Wikforss, A. M., (2001) ‘Social Externalism and Conceptual Errors’, in *Philosophical Quarterly*, Vol. 51, Iss. 203, pp. 217-231

_____, (2008) ‘Self-Knowledge and Knowledge of Content’, in *Canadian Journal of Philosophy*, Vol. 38, No. 3, pp. 399-424

Williams, M., (1999) *Wittgenstein, Mind and Meaning: Toward a Social Conception of Mind* (London: Routledge)

Witherspoon, E., (2011) ‘Wittgenstein on Criteria and The problem of Other Minds’, in O. Kuusela and M. McGinn (eds.), *The Oxford Handbook of Wittgenstein* (Oxford: Oxford University Press), pp. 472-497

Wittgenstein, L., (1958) *Philosophical Investigations*, 2nd edition, G. Anscombe (trans.) (Oxford: Blackwell Publishing)

_____, (1958) *The Blue and Brown Books: Preliminary Studies for Philosophical Investigations* (Oxford: Blackwell Publishing)

_____, (1967) *Zettel*, 2nd edition, G. H. von Wright & G. E. M. Anscombe (eds.), G. E. M. Anscombe (trans.) (Oxford: Basil Blackwell)

_____, (1974) *Philosophical Grammar*, R. Rhees, ed., A. Kenny (trans.) (Oxford: Basil Blackwell)

_____, (1975) *Wittgenstein’s Lectures on the Foundations of Mathematics*, C. Diamond (ed.) (Chicago: The University of Chicago Press)

_____, (1978) *Remarks on the Foundations of Mathematics*, 3rd edition, G. H. von Wright, R. Rhees & G. E. M. Anscombe (eds.), G. E. M. Anscombe (trans.) (Oxford: Basil Blackwell)

_____, (2009) *Philosophical Investigations*, rev. 4th edition, P. Hacker & J. Schulte (eds.), G. Anscombe, P. Hacker & J. Schulte (trans.) (Oxford: Wiley Blackwell)

Wright, C., (1998) ‘Self-Knowledge: The Wittgensteinian Legacy’ in C. Wright, B. Smith and C. Macdonald (eds.), *Knowing Our Own Minds* (Clarendon Press: Oxford), pp. 13-46

_____, (2001) ‘Wittgenstein’s Later Philosophy of Mind: Sensation, Privacy, and Intention’, in *Rails to Infinity* (Cambridge: Harvard University Press), pp. 291-318