



The
University
Of
Sheffield.

Content Externalism and The Problem of Self-Knowledge

PhD Thesis

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

Simon Michael Stubbs

**University of Sheffield
Department of Philosophy**

Date of Submission: 04/2021

ABSTRACT: The thesis examines the worry that celebrated developments in Anglophone philosophy that fall under the labels ‘semantic externalism’, ‘content externalism’, ‘anti-individualism’, etc. are incompatible with a purported long-held notion regarding self-knowledge. The thesis examines the worry that the developments that fall under the labels ‘semantic externalism’, etc. may be incompatible with the notion that we know a priori what the content of our thoughts are. Although the thesis finds that the most prominent arguments for such incompatibility in the existing philosophical literature prove to be unconvincing, it does present an original argument for such incompatibility, and does treat the prospect of such incompatibility as a live possibility. The thesis argues that although the developments may indeed be incompatible with the notion that we know a priori what the content of our thoughts are, that need not also mean that they will be incompatible with the notion that we have a privileged epistemic access to the content of our own thoughts. The thesis suggests that what has motivated the worry is a conflation of those two things. It seems to be incontrovertible that we have a privileged epistemic access to the content of our own thoughts. It need not, however, seem *as* incontrovertible that we must be able to know a priori what the content of our own thoughts are. It is more feasible to reject the latter notion, and – the thesis ventures – we can do so without rejecting the former.

Contents

Introduction.....	7
0.0: <i>What Keeps Me Awake at Night?</i>	7
0.1: Outline of Thesis.....	9
0.2: Two Omissions	9
0.3: An Assumption About Thoughts and Phenomenology.....	10
Chapter 1: Content Externalism: Burge V. Kripke.....	13
1.0: Introduction to Chapter 1	13
1.1: Definition of Content Externalism.....	13
1.2: Burgean Externalism.....	15
1.21: Ascription and Oblique Occurrence: Burge’s Framework.....	15
1.22: Content Externalism via “Incomplete Understanding”.....	16
1.23: Content Externalism via “Ignorance of Expert Knowledge”	19
1.24: Content Externalism via “Nonstandard Theory”	20
1.3: Kripkean Externalism	23
1.31: Content Externalism via Millianism	23
1.32: Millianism Extended to Natural Kind Terms.....	28
1.4: Burge (and Kripke) on Singular Thought	28
Chapter 2: Slow Switching I: The Redeployment Thesis.....	35
2.0: Introduction to Chapter 2	35
2.1: The Redeployment Thesis and the Incomplete Argument for Compatibilism	35
2.2: The Burgean Candidate	38
2.3: The ‘No Provenance’ Objection and the ‘No Phenomenology’ Problem: A Cliff Hanger.....	40
Chapter 3: Slow Switching II: The Discrimination Argument.....	43
3.0: Introduction to Chapter 1	43
3.1: The Discrimination Argument	44
3.2: Responses to the Discrimination Argument	47
3.21: Warfield’s Response	47
3.22: The Reliabilist Response.....	49
3.23: Brown’s Response I: Brown’s Overall Strategy	56
3.24: Brown’s Response II: Natural Kind Anti-Individualism and Recognitional Capacities	57
3.25: Brown’s Response III: Perceptual Demonstrative Singular Thought: The Problem Case... 60	

3.26: Brown’s Response IV: Recognition-Based Singular Thought.....	61
3.27: Brown’s Response V: Social Anti-Individualism (a.k.a. Burgean Externalism) and Illusion Cases	65
Chapter 4: Inferring from Mind to World I: Boghossian’s Reductio.....	67
4.0: Introduction to Chapter 4.....	67
4.1: Boghossian’s Reductio: Preliminaries and the Basic Argument.....	67
4.11: Preliminaries I: Content and ‘Concept’	67
4.12: Preliminaries II: Content and ‘Concept’ and Twin Earth	69
4.13: The Basic Argument	71
4.2: Three Problems for Boghossian’s Reductio	73
4.21: The Three Problems.....	73
4.22: First Problem I: Korman’s Counterexamples	73
4.23: First Problem II: Boghossian on the Language of Thought Hypothesis.....	76
4.24: First Problem III: An Alternative Argument for Why Thirsty Oscar’s Concept Could not be Non-Atomic	77
4.25: Second Problem I: Two Ways in Which Thirsty Oscar’s Concept Might be Atomic	78
2.26: Second Problem II: Thirsty Oscar’s Concept Does Not Need Application Conditions: McLaughlin and Tye’s “Concrete Proposal”	79
4.27: Second Problem III: Three Ways in which Thirsty Oscar’s Concept Could have Application Conditions	80
4.28: Second Problem IV: Thirsty Oscar’s Concept is ‘Motley’	81
4.29: Second Problem V: Thirsty Oscar’s Concept is Necessarily Vacuous	81
4.210: Second Problem VI: Necessary Vacuity and the Three Burgean Paradigms.....	86
4.211: Second Problem VII: Illusory Bouncers	87
4.212: The Third Problem: Boghossian’s Account of the Provenance of Content Externalism...	89
Chapter 5: Inferring from Mind to World II: Brown’s Reductio	93
5.0: Introduction to Chapter 5.....	93
5.1: Brown’s Reductio: The Basic Argument.....	94
5.2: Responding to Brown’s Reductio	96
5.21: Doubting Q	96
5.22: Doubting R I: The Concept ‘gilver’	99
5.23: Doubting R II: The Concept ‘zofa’	100
5.24: How can you Know A Priori that you’re ‘Agnostic’?	102
Conclusion to Chapters 3-5	109
Chapter 6: Inferring from Mind to World III: An Additional Reductio.....	113
6.0: Introduction to Chapter 6.....	113
6.1: The Basic Argument	113

6.2: Could ‘Aristotle’ be Shorthand for a Definite Description in the Event that he did not Exist?	114
6.3: Could ‘Aristotle’ Still Refer Directly even if Aristotle did not Exist (if You’re a Millian)?...	118
Chapter 7: Three <i>Broad</i> Positions on the Provenance of Our Self-Ascriptions of Thought and Other Mental States	121
7.0: Introduction to Chapter 7	121
7.1: Outline of Chapter 7	121
7.2: McKinsey and Brown and the Supposed Equivalence of A Priori and Privileged Access.....	122
7.3: Three Broad Positions Concerning the Method by Which We Self-Ascribe Mental States....	124
7.31: The First Position.....	124
7.32: The Second Position	130
7.33: The Third Position	132
7.34: Why be a Third Positionist?.....	133
7.4: A Third-Positionist Account of the Method by Which We Self-Ascribe Thought.....	134
7.5: Loose Ends.....	143
7.51: The Three Reductios	143
7.52: The Problem of the ‘Empty’ Case	145
7.53: Slow Switching	146
7.54: Comparison with Burge’s ‘Cogito-Like Judgements’ Account of Thought Self-Ascription: A Cliff Hanger Resolved.....	146
7.55: The Redeployment Thesis.....	147
Conclusion to Thesis	149
References	151

Introduction

0.0: *What Keeps Me Awake at Night?*

Thoughts are strange. They are insubstantial, formless things that hardly seem to be there at all. Yet, they are things that we are very much aware of having; they are a constant and incessant feature of our waking lives. Our awareness of our thoughts constitutes one of the most humdrum aspects of human experience. But it also, paradoxically, seems to constitute one of the most profound mysteries.

I lie awake at night. The lights are out. There are no audible noises. My senses are almost entirely deprived of any stimulus. But I cannot sleep. My thoughts will not allow me. I am being viciously and mercilessly harangued. Thoughts about the regrettable events of the preceding day keep fizzing up and forcing themselves before my conscious mind, precluding any chance of sleep. What I am experiencing here is not really something that I am experiencing. These thoughts that I cannot rid myself of produce no experiential phenomenology. There is no distinctive what it is like to think them.¹ Yet I am being kept awake nonetheless.

There are certain claims that are standardly made regarding the nature of our awareness of our thoughts.

One such claim is that a subject is able to *know a priori* what it is she is thinking. A subject should be able to *know a priori* what the *content* of her thoughts are. A subject should be able to judge correctly what she is thinking without investigating the world around her, without such judgements having what the subject's *senses* as part of their justificatory basis. When I am suffering my nocturnal torment, I may be judging that I should not have done a certain thing. As part of making that judgement I may make judgements about the consequences of that regrettable action. For instance, I

¹ Or at least arguably there isn't – see below §0.3.

may remember the hurt look on the face of a close personal acquaintance, and thus judge that my relations with that acquaintance will forever be soured. Such a judgement will be *a posteriori*, and thus not *a priori*. Thus, if I judge, based on that, that I should not have done what it is I have done, then that judgement would not be *a priori*. However, this should not be the case when it comes to my *second order* judgement that I am thinking about how I should not have done that regrettable thing. This second order judgment should not, in any way, hinge on judgements that I've made about the external world.

There is a worry (Boghossian, 1989; 1998; McKinsey, 1991; Brown, 1995) that has sparked substantial debate – much of which features in this thesis (e.g. Brown, 2004; Korman, 2006; McLaughlin and Tye, 1998a; Warfield, 1992)². The worry is that this claim may be incompatible with certain celebrated developments in Anglophone philosophy instigated, mostly circa 1970s, by Kripke, Putnam, Burge, et al (Kripke, 1982; Putnam, 1975; Burge, 1979; Kaplan, 1989; Salmon, 1982; et al) that fall under the labels ‘content externalism’, ‘semantic externalism’, and ‘anti-individualism’. It is thought that if this is the case then that constitutes a grave problem for those developments. If the developments are incompatible with this first claim, then it must either be the case that the claim is false or that the developments are false. It is supposed that it can't be the case that the claim is false. Thus, if the developments are incompatible with the claim, then it must be the case that the developments are false.

There is also another claim that is standardly made regarding the nature of our awareness of our thoughts.

This second claim concerns an *epistemic asymmetry*. The claim is that a subject is better able to judge what it is she is thinking, what the *content* of her thoughts are than anyone else could possibly be. There is an asymmetry with respect to the epistemic relationship a subject has with her thoughts and the epistemic relationships that a third party may have with that subject's thoughts. We can observe the external behaviour of others and abduce what they might be thinking. But we can't know what's on their mind as well as we can know what is on our own mind.

In this thesis, I assess whether the developments that fall under the labels ‘semantic externalism’, etc. and the first – *a priori* – claim are indeed incompatible. I also assess whether their being incompatible should be deemed problematic. I suggest that a conflation of the first – *a priori* – claim with the second – privilege – claim may be the root of the concern. The first – *a priori* – claim, I suggest, may not be as incontrovertible as the second – privilege – claim. Thus, although it may be the

² Also see Beebe, 2001; Brueckner, 1992; 2002; 2004; 2008; Falvey, 2000; Falvey and Owens, 1994; Gerken, 2008; Gibbons, 1996; Goldberg, 2003; 2015; Kallestrup, 2011; Lasonen-Aarnio, 2006; Ludlow, 1997; McKinsey, 2007; McLaughlin and Tye, 1998a; 1998b; Noordhof, 2005;; Sawyer, 1998; Warfield, 1997; et al. And also see the anthologies: Ludlow and Martin, 1998; Wright, Smith, and MacDonald, 1998; Nuccetelli, 2003; Goldberg, 2015.

case that the developments that fall under the labels ‘content externalism’, etc. being incompatible with the second – privilege – claim could constitute a grave problem for those developments, their being incompatible with the first – a priori – claim possibly does not. If we cannot deny the second – privilege – claim, then we might be able to deny the first – a priori – claim, and we might be able to do so without denying the second – privilege – claim.

0.1: Outline of Thesis

The thesis comprises seven chapters.

In Chapter 1, I present a definition of what I shall call *content externalism*. I go on to characterise two prominent ways in which it is possible to be a content externalist.

In Chapters 2, 3, 4 and 5, I examine the debate that lasted from the late eighties (Davidson, 1987; Burge, 1988; Boghossian, 1989) to the mid noughties (Brown, 2004; Korman, 2006; et al) over whether content externalism is compatible with the claim that we have an a priori access to the contents of our thoughts. There have been two major strands to this debate. The first strand has concerned what have become known as *slow-switch* cases. The second strand has concerned potential reductios that putatively demonstrate that if content externalism, or a given form of content externalism, is true, and if it is indeed the case that we have an a priori access to our thoughts, then absurd consequences would follow. Chapters 2 and 3 consider the first strand. Chapters 4 and 5 consider the second.

In Chapter 6, I present an original argument for the incompatibility of what I call *Kripkean externalism* and the claim that we have an a priori access to our own thoughts, which takes the form of a reductio argument similar to those considered in Chapters 4 and 5.

In Chapter 7, I consider how our having an a priori access to our thoughts may not be as incontrovertible as it has been made out to be. I sketch an account of the provenance of our thought self-ascriptions that builds on the observation that thoughts produce no phenomenology, and under which we’d have a privileged access to our thoughts that is not necessarily a priori.

All seven chapters are relatively self-contained and could potentially be read in any order. It is important to note, however, that the first two chapters are intended as an informal introduction to, and informal framework for, the five chapters that ensue.

0.2: Two Omissions

There are two significant topics in the literature that this present thesis does not touch upon. One is the bearing that slow-switch cases, as considered in Chapters 2 and 3, have on our understanding of

memory and ratiocination.³ The other is the bearing that Wright's et al⁴ ideas on *warrant transmission failure* have on the reductio arguments considered in Chapters 4 and 5 (and also 6). These two omissions are largely due to considerations of space.⁵ (Indeed, a full exploration of either would warrant – no pun intended – an entire PhD thesis in itself!) Both topics might be brought to bear on the specific lines of argument pursued in this present thesis. But that is something that must wait until potential sequels.

0.3: An Assumption About Thoughts and Phenomenology

I make an assumption throughout the thesis that is particularly relevant in Chapters 2 and 7, and that some may deem controversial. For me it is self-evident that thoughts do not produce phenomenology in the way that perceptions and sensations do. I appreciate, however, that not all may agree. I appreciate that not all philosophers find it to be incontrovertible that thought produces no phenomenology (e.g. see Lormand, 1996; Horgan and Tienson, 2002; Pitt, 2004; and Tye and Wright, 2011)⁶. An unequivocal acceptance that thought produces no phenomenology is not strictly essential for the line of argument explored in Chapters 2 and 7. If thoughts having no phenomenology is taken as a *given*, then the account of thought self-ascription that I give in Chapter 7 can be viewed as an explanation of how it is that we self-ascribe thoughts *given* that thoughts produce no phenomenology, as well as demonstrating how we can have a privileged access to our thoughts without having an a priori access to our thoughts. If, on the other hand, the question of whether thoughts produce their own phenomenology is considered something that is up for debate, then my account can be viewed as

³ Wright, 2003; 2003; 2011; Davies, 2003; Schiffer, 2005; et al.

⁴ E.g. Tye, 1998.

⁵ The omission of Wright et al also, along with some of the other directions that this thesis takes, stem in-part from a broad frustration with the discipline of epistemology as it has been practiced within analytic tradition philosophy. Epistemology concerns normative evaluations of our beliefs. Such evaluations necessitate examinations of the provenance of our beliefs, but the epistemologist does not view such examinations of provenance as ends in themselves. Why we believe what we believe about ourselves and the external world is possibly a more interesting question than whether we deserve a pat on the back for believing what we believe. I believe this might be especially true when it comes to the beliefs we form about ourselves. (It is only in Chapter 3 and in certain places towards the end of Chapter 5 (§5.24) where I, in order to assess the arguments of others, fully succumb to this tendency in epistemology.)

An additional reason for the *other* omission is focus. The entire focus of the thesis is on how we, if content externalism is true, discern the content of our occurrent thoughts – how we discern what we are presently thinking. If it were to consider memory and ratiocination, then this focus would become diffused.

⁶ Horgan and Tienson (2002) and Pitt (2004) put forward a *phenomenology of thought thesis* that holds that for each thinkable thought there is a distinct what it is like to think that thought that is experienced during each token of the thinking of that thought. Tye and Wright point out that such accounts “have overlooked one important point: in the relevant sense, there is no such thing as the phenomenology of thought” (Tye and Wright, 2011, p.326).

an explanation of how it is that we self-ascribe thought *if it is indeed (or were) the case* that thoughts produce no phenomenology (as well as demonstrating how we can have a privileged access to our thoughts without needing an a priori access to our thoughts).

Chapter 1: Content Externalism: Burge V. Kripke

1.0: Introduction to Chapter 1

In this first chapter I define what I call *content externalism* and explain why two prominent content externalists are content externalists. I give an account of the arguments for content externalism advanced by Burge and by Kripke.

Being sympathetic to either Burge or Kripke is not exhaustive of all the possible ways it is to be a content externalist. If one is a content externalist, then it is not necessarily the case that one is either a *Burgean externalist* or a *Kripkean externalist*. I focus on Burge and Kripke because they are two of the most prominent – if not the two most prominent – content externalists, and also because Chapters 5 and 6 respectively concern reductio arguments that apply to Burgean externalism – or rather Brown’s interpretation of what Burgean externalism is – and Kripkean externalism specifically.

The aim of this chapter is not to assess the positions of Kripke and Burge. The aim rather is to give an accurate characterisation of the respective provenances of Burge’s and Kripke’s content externalism.

I end the chapter characterising what (I think) is a key difference between Burge and Kripke, which is that Burge holds that the contents of *singular thoughts* are not individuated by their object whereas Kripke essentially does.

1.1: Definition of Content Externalism

What we can call *content externalism* holds that thought content does not supervene on intrinsic properties. A content externalist holds that there could be two subjects that are (qualitatively) identical, but who are nevertheless thinking thoughts with different contents.

Two such subjects will be identical in terms of their physical parameters. They will be atom-for-atom duplicates of each other, and they will have been throughout their entire respective life histories. Every chemical reaction that has taken place within the body of subject A (let's call her) will also have taken place in the body of her twin subject B (let's call her) and will have done so in a way that mimics exactly the way in which it has in A's body, and vice-versa. Every neural firing that has taken place in the brain of A, will also have taken place in the brain of B, and will have done so in a way that mimics exactly the way in which it has occurred in the brain of A, and vice-versa.

The two subjects will also have experienced all the same phenomenology. Every event that has taken place in A's conscious life that has a distinctive phenomenology will have taken place in B's conscious life also. Every itch, every pain, etc. that A has experienced in her entire life history, B will also have experienced, and all what B will have experienced will have had a phenomenological character that is indistinguishable from what A has experienced, and vice-versa.

There will be some difference in the subjects' respective external environments, which the subjects will not be cognizant of. That difference will be why the subjects will be thinking thoughts with different contents.

Content externalism holds that thought is *relational*. The content externalist contends that a subject's cognitive act of thinking certain thoughts is constituted by her being, during the moment of her cognitive act, in relations or complexes of relations with things that are in her external environment.

What I call *Kripkean externalism* holds that a subject's thinking certain thoughts about a particular external object or individual is constituted by her being in a 2-way relation with that external object or individual.

What I call *Burgean externalism* holds that a subject's thinking certain thoughts is constituted by her being in a complex of relations with both her environment and her socio-linguistic community.

The focus of this thesis is on the content of what one is presently thinking, and the epistemic relationship one has with that content. Thus, I couch my definition of content externalism in terms of thought content. It will, of course, apply in obvious ways to the content of propositional attitudes as well.

The reverse is also true. When talking of content externalism, although the points being made may be couched in terms of the content of propositional attitudes, those points often will apply, in obvious ways, to the content of occurrent thought. There will often, but (possibly) not always, be an obvious interchangeability between the two.

1.2: Burgean Externalism⁷

1.2.1: Ascription and Oblique Occurrence: Burge's Framework

Burge derives his brand of content externalism from an assumption that is practically tautological. This is what (I think) makes Burge's arguments compelling.

The assumption, basically, is that if it would be deemed correct to ascribe to someone a thought or attitude with some content, then it will be the case that that person harbours that attitude or is thinking that thought.

Suppose that it can be said truthfully that: 'John believes that Paris is the capital of France, and John also believes that Montevideo is the capital of Uruguay'. What this will entail, according to the assumption is that John harbours two beliefs with distinct contents. It will be the case that John believes that Paris is the capital of France – that he harbours a belief that has the content 'Paris is the capital of France' – and that he believes that Montevideo is the capital of Uruguay – he harbours a belief with the content 'Montevideo is the capital of Uruguay'.

More specifically and accurately, Burge's assumption is that the content of a subject's propositional attitudes corresponds to the *'that'-clauses* of what would be veridical propositional attitude ascriptions in which referring expressions *occur obliquely*:

...obliquely occurring expressions in content clauses are a primary means of identifying a person's intentional mental states or events.

(Burge, 1979, p.76)

An attitude *content* is the semantic value associated with oblique occurrences of expressions in attributions of propositional attitudes.

(Burge, 1982, p.81, fn. 2)⁸

...it is expressions in oblique occurrence that play the role of specifying a person's mental contents, what his thoughts are.

(Ibid., p.86)

⁷ The key works in which Burge develops his brand of externalism are 'Belief *De Re*', 'Individualism and the Mental', 'Other Bodies', and 'Intellectual Norms and the Foundations of Mind' (Burge, 1977; 1979; 1982; 1986).

⁸ Page numbers correspond to the reprint in Burge 2007.

A referring expression *occurs obliquely* in the ‘that’-clause of a propositional attitude ascription when that referring expression *cannot* be exchanged for co-referential referring expressions *salva veritate*.

Suppose that unbeknownst to Hank, the man who lives next door is an undercover master criminal who happens to be at the very top of the FBI’s most wanted list. This master criminal has been putting on an act of being a profoundly unremarkable suburbanite, which has totally fooled Hank. We can clearly say that: ‘Hank believes that his next-door neighbour is unremarkable’. Suppose now we were to say in addition that: ‘Hank believes that the most wanted man in America is unremarkable’. There is a context in which the latter ascription – where we substitute ‘his [Hank’s] next door neighbour’ for the co-referential ‘the most wanted man in America’ – would be acceptable, and a context in which it wouldn’t. It is in the latter context that the referring expressions are understood to occur obliquely. Burge’s contention is that it is only consideration of this latter context that provides us with the means of determining what – in the relevant philosophical sense – Hank actually believes.

Burge is not as interested in propositional attitude ascriptions as he is in propositional attitudes. He sees considerations of propositional attitude ascription as a means to an end. By considering discourse involving propositional attitude ascriptions (in which referring expressions occur obliquely) we can gain insight into propositional attitudes. In particular, this methodology should reveal facts about the individuation of propositional attitudes, and, by extension and more relevantly for the purposes of this thesis, the individuation of thoughts.

1.22: Content Externalism via “Incomplete Understanding”

In ‘Individualism and the Mental’ (Burge, 1979), Burge presents a famous three-step thought experiment that is meant to yield content externalism, involving a case in which a subject *incompletely understands* a certain term.

First Step

The first step requires that we concede that vocalised propositional attitude ascriptions (in which referring expressions occur obliquely – see above) can be veridical, even if they involve a term that the ascriber doesn’t fully understand.

Such veridical ascriptions should accurately characterise that ascriber’s propositional attitudes, given the near-tautological assumption – discussed in the previous subsection – that veridical attitude ascriptions (in which referring expressions occur obliquely) accurately characterise that subject’s propositional attitudes.

The example that Burge famously gives involves the term ‘arthritis’. Burge has us imagine a case in which someone, who we can call Alf, goes to the doctor complaining that the ‘arthritis’ in his knee has spread to his thigh. Alf does not realise that the term ‘arthritis’ only encompasses ailments of the joints.

Burge requires that we concede that ascriptions such as ‘Alf believes that arthritis is not as bad as cancer of the liver’ can be veridical, and thus accurately capture Alf’s belief about how arthritis is not as bad as cancer of the liver. Burge has us suppose that:

A given person [i.e. Alf] has a large number of attitudes commonly attributed with content clauses containing ‘arthritis’ in oblique occurrence. For example, he thinks (correctly) that he has had arthritis for years, that his arthritis in his wrist and fingers is more painful than his arthritis in his ankles, that it is better to have arthritis than cancer of the liver, that stiffening joints is a symptom of arthritis, that certain sorts of aches are characteristic of arthritis, that there are various kinds of arthritis and so forth. In short, he has a wide range of such attitudes. In addition to these unsurprising attitudes, he thinks falsely that he has developed arthritis in the thigh.

(Burge, 1979, p.77)

Second Step

The second step requires that we imagine a counterfactual scenario in which the term in question is used differently to how it actually is, but in which that change of usage does not impinge on the subject’s intrinsic properties in any way.

In the case of Alf the arthritis sufferer, we need to imagine a counterfactual scenario in which the term ‘arthritis’ – as used by Alf’s wider socio-linguistic community – encompasses a wider range of ailments than just ailments of the joints, but in which from Alf’s perspective everything is as it actually is. In this counterfactual scenario the term could potentially apply to ailments that can spread from the knee to the thigh. But that change in the definition of ‘arthritis’ will not impinge on Alf’s intrinsic properties in any way. Counterfactual Alf and Actual Alf will have been identical in terms of all their physical parameters throughout their entire respective life histories. Every neural firing that has taken place in the brain of Actual Alf, will also have taken place in the brain of Counterfactual Alf, and will have done so in a way that mimics exactly the way in which it has occurred in the brain of Actual Alf. They will have also experienced all the same phenomenology throughout their entire respective life histories. Every itch, every pain, etc. that Actual Alf has experienced, Counterfactual Alf will have experienced also.

In Burge’s own words:

We are to conceive of a situation in which the patient [Alf] 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 [that the arthritis has spread to his thigh] to his doctor. Precisely the same things (non-intentional ascribed) happen to him. He has the same physiological history, the same diseases, the same internal physical occurrences. He goes through the same motions, engages in the same behaviour, has the same sensory intake (physiologically described). His dispositions to respond to stimuli are explained in physical theory as the effects of the same proximate causes. All this extends to his interaction with linguistic expressions. He says and hears the same words (word forms) at the same times he actually does. He develops the disposition to assent to ‘Arthritis can occur in the thigh’ and ‘I have arthritis in the thigh’ as a result of the same physically describe approximate causes. Such dispositions might have arisen in a number of ways. But we can suppose that in both actual and counterfactual situations, he acquires the word ‘arthritis’ from casual conversation or reading, and never hearing anything to prejudice him for or against applying it in the way he does, he applies the word to an ailment in his thigh (or to ailments in the limbs or other) which seems to produce pains or other symptoms roughly similar to the disease in his hands and ankles. In both actual and counterfactual cases, the disposition is never reinforced or extinguished up until the time when he expresses himself to his doctor. We further imagine that the patient’s [Alf’s] non-intentional, phenomenal experience is the same. He has the same pains, visual fields, images, and internal verbal rehearsals. The *counterfactuality* in the supposition touches only the patient’s [Alf’s] social environment. In actual fact, ‘arthritis’, as used in his community, does not apply to ailments outside the joints. Indeed, it fails to do so by a standard, non-technical definition. But in our imagined case, physicians, lexicographers, and informed laymen apply ‘arthritis’ not only to arthritis but to various rheumatoid ailments. The standard use of the term is to be conceived to encompass the patient’s [Alf’s] actual misuse.

(*ibid.*, pp.77-78; Burge’s emphasis)

Third Step

The third and final step requires that we concede that, in the counterfactual scenario, the equivalent propositional attitude ascriptions that would involve the term in question would be veridical.

What follows directly from this is that the content of the subject's propositional attitudes in the counterfactual scenario will be different to the content of the subject's actual propositional attitudes, even though the subject's intrinsic properties remain constant.

In the counterfactual scenario in which 'arthritis' is defined differently it would be correct to say: 'Alf believes that arthritis is not as bad as cancer'. Thus, that ascription would accurately capture the content of what Counterfactual Alf believes. However, the ascription would mean something different to what it actually means – per the counterfactual definition of 'arthritis'. This forces the conclusion that Actual Alf and Counterfactual Alf believe different things, even though they share all the same intrinsic properties.

1.23: Content Externalism via "Ignorance of Expert Knowledge"

Suppose that somewhere in the very distant universe there is a planet that is Earth's exact doppelganger. Earth and this *Twin Earth* are for the sake of all appearances exactly alike. The only difference between Earth and Twin Earth is the chemical composition of what the planets' respective inhabitants call 'water'. On Earth there is a clear, odourless liquid that is found all over the place called 'water', and which is in fact H₂O, i.e. it is water. On Twin Earth there is also a clear, odourless liquid that is found all over the place called 'water', but which is not H₂O. It is some other natural kind that has some other chemical composition, and that is standardly called – by Earthling philosophers – *twater*. Now suppose that on Earth there is Oscar who is ignorant of chemical theory – i.e. he doesn't know that water is H₂O. On Twin Earth there will be Twin Oscar, who will be Oscar's atom-for-atom duplicate.⁹ Twin Oscar will be likewise ignorant of chemistry. He won't know what the chemical composition of *twater* is. What I am describing here is Putnam's *infamous* 'Twin Earth' thought experiment (Putnam, 1975). Content externalists ubiquitously hold that the content of Oscar's and Twin Oscar's thoughts and attitude will be different. When Oscar thinks the thought that he'd express 'water is wet', the equivalent thought that Twin Oscar would think would be different (even though it would be expressed, by Twin Oscar and by other twin earthlings, using phonetically identical words). Oscar and Twin Oscar are the classic case of two subjects who are qualitatively identical to one and other, but who nevertheless think thoughts with different contents.

⁹ We need to ignore any other differences between Earth and Twin Earth that might follow as a consequent of 'water' having a different chemical composition. Strictly speaking, Oscar and Twin Oscar should not be atom-for-atom duplicates. Earthlings are 97% water. So Twin Earthlings should be 97% *twater*. This is something, along with other things, we need to ignore. This, of course, can be remedied – if indeed it needs to be – by change of example – from 'water' to 'aluminium' (say).

Burge contends that Oscar and Twin Oscar would indeed think thoughts with different contents. His reasons for thinking this pertain, as in the case of Alf the Arthritis sufferer, to thought and attitude ascriptions in which the relevant terms occur obliquely:

When [Oscar] says or consciously thinks the words, ‘There is some water within twenty miles, I hope’, [Twin Oscar] says or consciously thinks the same word forms. But there are differences. As Putnam in effect points out, [Oscar’s] occurrences of ‘water’ apply to water and mean *water*, whereas [Twin Oscar’s] apply to twater and mean *twater*. And, as Putnam does not note, the differences affect *oblique* occurrences in ‘that’-clauses that provide the contents of their mental state and events. [Oscar] hopes that there is some water (oblique occurrence) within twenty miles. [Twin Oscar] hopes that there is some twater within twenty miles. That is, even as we suppose that ‘water’ and ‘twater’ are not logically exchangeable with coextensive expressions *salva veritate*, we have a difference between their thoughts (thought contents).

(Burge, 1982, p.86)¹⁰

Oscar’s and Twin Oscar’s thinking different thoughts is due to a difference in the meaning of the term ‘water’ on their respective planets. This in turn is due to the non-socio-linguistic physical facts that water and twater have the chemical compositions that they have, and how those facts effect (and have effected) the linguistic practices of Oscar’s and Twin Oscar’s respective socio-linguistic communities, which will include scientifically-literate individuals who are aware of the chemical composition of water/twater:

The difference in [Oscar’s and Twin Oscar’s respective] 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.

(*ibid.*, p.87)

1.24: Content Externalism via “Nonstandard Theory”

Along with the route that Burge first presented in ‘Individualism and the Mental’ (Burge, 1979), and along with the route to content externalism that consists of Burge’s reading of ‘Twin Earth’, which is found in ‘Other Bodies’ (Burge, 1982), Burge also presents a third dialectical route to content

¹⁰ Page numbers correspond to Burge 2007.

externalism, which is found in ‘Intellectual Norms and the Foundations of Mind’ (Burge, 1986). The first route centres around someone – Alf the Arthritis Sufferer – who has an *incomplete understanding* of a certain term – ‘arthritis’. The second centres around someone – Oscar – who is ignorant of the chemical composition of a natural kind – water. The third centres around someone who has a “nonstandard theory” with regards to the application of a certain term:

The argument of ‘Individualism and the Mental’ and ‘Other Bodies’ ascribe incomplete linguistic understanding and ignorance of expert knowledge (respectively) to the relevant protagonists...The [third] argument features not incomplete understanding or ignorance of specialised knowledge, but nonstandard theory.

(Burge, 1986, pp.708-9)

The third route consists a three-step thought experiment that echoes the one found in ‘Individualism and the Mental’. We start off with an individual who is wrong about the application of a term, but who we can nevertheless ascribe propositional attitudes to that involve that term (in oblique position). We then imagine a counterfactual scenario in which the person is right. Finally, we recognise that our counterfactual selves, in that counterfactual scenario, would be similarly able to ascribe propositional attitudes to the individual that involve the term (in oblique position), and because of that the individual would have attitudes that would have a different content to the attitudes that they actually have.

The key difference between the ‘Individualism and the Mental’ thought experiment, as already alluded to, pertains to the manner in which the respective individuals are wrong about a term’s application. Alf the Arthritis sufferer is wrong about the application of ‘arthritis’ due to his ignorance about how his wider socio-linguistic community define the term. Alf has an “incomplete understanding” of what ‘arthritis’ is. In the alternative thought experiment, we are to imagine someone who is fully aware of how their wider socio-linguistic community define a particular term, but who think their wider socio-linguistic community have got it wrong. In the alternative thought experiment, “incomplete understanding” is substituted for “nonstandard theory”.

Burge has us imagine a case in which someone – “person A” – subscribes to a nonstandard theory about ‘sofas’:

We begin by imagining a person A in our community who has a normal mastery of English. A’s early instruction in the use of ‘sofa’ is mostly ostensive, though he picks up normal truisms. A can use the term reliably. At some point, however, A doubts the truisms and hypothesises that sofa function not as furnishings to be sat on, but as works of art or religious artefacts. He believes that the usual remarks about the function of sofas conceal,

or represent a delusion about, an entirely different practice. *A* admits that some sofas have been sat upon, but thinks that most sofas would collapse under any considerable weight and denies that sitting is what sofas are pre-eminently for. *A* may attack the veridicality of many of our memories of sofas being sat upon, on the grounds that the memories are product of the delusion.

(Burge, 1986, p.707)

Burge then has us imagine a counterfactual scenario in which person *A*'s bizarre theory about sofas is right, in which 'sofas' are – and are recognised by the wider socio-linguistic community as being – iconographic artefacts that would normally collapse if sat upon, but in a way in which it would not impinge on *A*'s non-intentional intrinsic properties in any way:

As a second step, imagine a person *B* (or *A* in nonactual circumstances) who is, for all intents and purposes, physically identical to *A*. He has the same physical dispositions, receives substantially the same physical simulations, produces the same motions, utters the same sounds. Like *A*, *B* hears, though seldom, word forms that are counterparts to the truisms that *A* hears. But in *B*'s situation, these word forms are not taken as truisms; they are contextual appropriate remarks that do not purport to convey a general meaning. (They could be lies or jokes, but it is more natural to take them as not-completely-general contingent truths.) The objects that *B* is confronted with are objects that look like sofas, but are, and are widely known to be, works of art, or religious artefacts sold in showrooms and displayed in people's houses. Many of these objects would collapse under a person's weight. There are no sofas in *B*'s situation, and the word 'sofa' does not mean sofa. Call the relevant objects 'safos'. *B* assumes that most people would take these objects to function primarily as seats and that the remarks he hears are communally accepted truisms. But, like *A*, *B* develops doubts. At least by the time *B* expresses this scepticism and his theory, he is correctly doubting that safos function as furniture to be sat upon.

(ibid., pp.707-8)

Finally, Burge expects that we consequently draw the same conclusions *mutatis mutandis* about the contents of *A*'s and *B*'s propositional attitudes that we draw in the case of Actual Alf and Counterfactual Alf:

A mistakenly thinks that sofas do not function primarily to be sat upon. *B*'s counterpart thoughts do not involve the [actual] notion of sofa and could not correctly be ascribed with 'sofa' in oblique position [given its actual usage].

(*ibid.*, p.708)

1.3: Kripkean Externalism

1.31: Content Externalism via Millianism

Millianism is the thesis that the contribution that a name – a proper noun – makes to the meaning of a sentence in which it occurs, i.e. the content of the thought that that sentence expresses, is exhausted by the name's referent.¹¹ Millianism is something of a philosophical bogeyman. Something that according to received wisdom we must reject lest we face a host of inter-related issues pertaining to belief and desire ascription and the ways in which thoughts and attitudes feature in our cognitive economies. Kripke, nevertheless, argues in favour of Millianism, and against what he sees as the opposing 'Fregean'¹² view that holds that attached to a name there is something like a Fregean 'sense' that either makes the sole or an additional contribution to content determination. Kripke argues that a rejection of Millianism would be counter to common-sense intuitions about how names function in socio-linguistic practice, and that, crucially, rejecting it, and thereby accepting the opposing 'Fregean' view, would not in actual fact furnish us with the solutions that it is touted to furnish us with. Kripke does accept that Millianism poses serious philosophical challenges. He acknowledges the lure of Fregeanism. He argues however that Fregeanism fails to provide the solutions that it is touted to provide.

Kripke defines Millianism as the view that:

...a proper name is, so to speak, simply a name. It simply refers to its bearer, and has no other linguistic function. In particular, unlike a definite description, a name does not describe its bearer as possessing any special identifying properties.

(2011, p.126)

Immediately after, Kripke gives a loose working definition of what he sees as the opposing 'Fregean' view:

¹¹ So named after John Stuart Mill.

¹² The two most important Frege works are (arguably) 'On Sinn and Bedeutung' and 'Thought' (Frege, 1892; 1918). For an anthology of other key works by Frege see Beaney, 1997.

The opposing Fregean view holds that to each proper name, a speaker of the language associates some property (or conjunction of properties) which determines its referent as the unique thing fulfilling the associated property (or properties). This property(ies) constitutes the 'sense' of the name. Presumably, if '...' is a proper name, the associated properties are those that the speaker would supply, if asked, "Who is '...'" If he would answer "... is the man who ———," the properties filling the second blank are those that determine the reference of the name for the given speaker and constitute its 'sense.'
(ibid.)

Kripke then goes on to consider Quine's famous 'Cicero'/'Tully' example:

It has seemed...obvious that codesignative proper names are not interchangeable in belief contexts and epistemic contexts. Tom, a normal speaker of the language, may sincerely assent to "Tully denounced Catiline," but not to "Cicero denounced Catiline." He may even deny the latter. And his denial is compatible with his status as a normal English speaker who satisfies normal criteria for using both 'Cicero' and 'Tully' as names for the famed Roman (without knowing that 'Cicero' and 'Tully' name the same person). Given this, it seems obvious that Tom believes that: Tully denounced Catiline, but that he does not believe (lacks the belief) that: Cicero denounced Catiline. So it seems clear that codesignative proper names are not interchangeable in belief contexts. It also seems clear that there must be two distinct propositions or contents expressed by 'Cicero denounced Catiline' and 'Tully denounced Catiline.' How else can Tom believe one and deny the other? And the difference in propositions thus expressed can only come from a difference in sense between 'Tully' and 'Cicero.' Such a conclusion agrees with a Fregean theory and seems to be incompatible with a purely Millian view
(ibid., p.130)

According to the 'Fregean' explanation, in order to account for the phenomena so described – i.e. that 'Cicero' and 'Tully' *occur obliquely* (see previous section) – there must be attached to names such as 'Cicero' and 'Tully' something like a Fregean 'sense', i.e. something along the lines of what Kripke describes in the previous quotation. Now, such an attachment can either be something that is agreed upon by the socio-linguistic community as a whole, or it must pertain to the way or ways in which an individual utterer of the name thinks of its referent, i.e. what reference determining property (or properties) the speaker associates with the referent. Kripke argues that neither can be the case, and hence the 'Fregean' explanation cannot be right.

Kripke observes that in the case of many or most names – and particularly in the cases of the names ‘Cicero’ and ‘Tully’ – there is no socio-linguistic-community-wide-agreed-upon ‘sense’. Kripke compares the ‘Cicero’/‘Tully’ example to Frege’s paradigmatic ‘Hesperus’/‘Phosphorus’ example. Kripke initially considers how Fregeanism could be meshed with the most famous claim of *Naming and Necessity*¹³ that names are *rigid designators* to account for how the co-referential ‘Hesperus’ and ‘Phosphorus’ are non-interchangeable in intentional contexts:

In the case of ‘Hesperus’ and ‘Phosphorus,’ it is pretty clear what the different ‘modes of presentation’ [i.e. Fregean ‘senses’] would be: one mode determines a heavenly body by its typical position and appearance, in the appropriate season, in the evening; the other determines the same body by its position and appearance, in the appropriate season, in the morning. So it appears that even though, according to my view, proper names would be modally rigid—would have the same reference when we use them to speak of counterfactual situations as they do when used to describe the actual world— they would have a kind of Fregean ‘sense’ according to how that rigid reference is fixed. And the divergences of ‘sense’ (in this sense of ‘sense’) would lead to failures of interchangeability of co-designative names in contexts of propositional attitude, though not in modal contexts. Such a theory would agree with Mill regarding modal contexts but with Frege regarding belief contexts. The theory would not be purely Millian. (ibid, p.131)

Kripke argues, however, that this can’t be true in the case of names such as ‘Holland’ and ‘The Netherlands’, and – more to the point – names such as ‘Cicero’ and ‘Tully’:

After further thought, however, the Fregean conclusion appears less obvious. Just as people are said to have been unaware at one time of the fact that Hesperus is Phosphorus, so a normal speaker of English apparently may not know that Cicero is Tully, or that Holland is the Netherlands. For he may sincerely assent to ‘Cicero was lazy,’ while dissenting from ‘Tully was lazy,’ or he may sincerely assent to ‘Holland is a beautiful country,’ while dissenting from ‘The Netherlands is a beautiful country.’ In the case of ‘Hesperus’ and ‘Phosphorus,’ it seemed plausible to account for the parallel situation by supposing that ‘Hesperus’ and ‘Phosphorus’ fixed their (rigid) references to a single object in two conventionally different ways, one as the ‘evening star’ and one as the ‘morning star.’ But what corresponding conventional ‘senses,’ even taking ‘senses’ to be

¹³ Kripke, 1982.

‘modes of fixing the reference rigidly,’ can plausibly be supposed to exist for ‘Cicero’ and ‘Tully’ (or ‘Holland’ and ‘the Netherlands’)? Are not these just two names (in English) for the same man? Is there any special conventional, community-wide ‘connotation’ in the one lacking in the other? I am unaware of any.
(*ibid.*, p.132)

On the other horn of the dilemma, there is the view that ‘senses’ are individualistically attached to tokens of a name’s usage. Kripke observes that in the case of many, if not most, instances in which a name is uttered there could be no appropriate candidate ‘sense’ being individualistically assigned by the utterer:

...individuals who “define ‘Cicero’” by such phrases as “the Catiline denouncer,” “the author of *De Fato*,” etc., are relatively rare: their prevalence in the philosophical literature is the product of the excessive classical learning of some philosophers. Common men who clearly use ‘Cicero’ as a name for Cicero may be able to give no better answer to “Who was Cicero?” than “a famous Roman orator,” and they probably would say the same (if anything!) for ‘Tully.’ (Actually, most people probably have never heard the name ‘Tully.’) Similarly, many people who have heard of both Feynman and Gell-Mann, would identify each as ‘a leading contemporary theoretical physicist.’ Such people do not assign ‘senses’ of the usual type to the names that uniquely identify the referent (even though they use the names with a determinate reference). But to the extent that the indefinite descriptions attached or associated can be called ‘senses,’ the ‘senses’ assigned to ‘Cicero’ and ‘Tully,’ or to ‘Feynman’ and ‘Gell-Mann’ are identical.
(*ibid.*, pp.134-5)

What Kripke has presented us with is a simple but compelling argument for the Millian view that the contribution that a name makes to the content of a sentence is exhausted by the name’s referent. To deny the Millian view is to suppose that attached to the use of a name is something like a ‘Fregean’ sense. Such an attachment can either be agreed upon by the socio-linguistic community as a whole, or it must pertain to the way or ways in which an individual utterer of the name thinks of its referent. Neither option is tenable. In many, if not most, cases there can be no socio-linguistic-community-wide-agreed-upon ‘sense’. Similarly, in many, if not most cases, there will not be a ‘sense’ that is attached by an individual speaker. Therefore, the ‘Fregean’ view is false. Therefore, the Millian view is correct.

Supposing that attached to each name there is something like a Fregean ‘sense’ may seem very attractive, as a way of avoiding grievous philosophical pitfalls. But it simply isn’t viable.

Kripke also provides us with a second argument for Millianism. In *Naming and Necessity* Kripke presents a picture of how the referent of a name, but not anything like a ‘sense’, is passed down a socio-linguistic chain, which begins with the initial baptiser. If such a picture is correct, or if it is correct at least in some cases, then that would give strong support for Millianism:

...the real determinant of the reference of names of a former historical figure is a chain of communication, in which the reference of the name is passed from link to link. Now the legitimacy of such a chain accords much more with Millian views than with alternatives. For the view supposes that a learner acquires a name from the community by determining to use it with the same reference as does the community. We regard such a learner as using “Cicero is bald” to express the same thing the community expresses, regardless of variations in the properties different learners associate with ‘Cicero,’ as long as he determines that he will use the name with the referent current in the community. That a name can be transmitted in this way accords nicely with a Millian picture, according to which only the reference, not more specific properties associated with the name, is relevant to the semantics of sentences containing it.

(2011, p.136)

Now, in giving us arguments for Millianism, Kripke has provided us with additional routes to content externalism. Content externalism – as defined at the start of this chapter (§1.1) – follows trivially from Millianism. Indeed, if Millianism is correct then changing the reference of a name changes the content of the sentence in which it occurs, and thus the content of the thoughts that those sentences express. And, it is easy to conceive of actual/counterfactual pairs in which a name, as used by some individual, changes referent, but in which the intrinsic properties of that individual remain constant. For instance, suppose I think the thought that I’d express: ‘Aristotle was an alcoholic’. Imagine a counterfactual in which ‘Aristotle’ referred to someone else, who would be numerically non-identical to the actual Aristotle. Imagine also that within the counterfactual scenario all my intrinsic properties would be exactly as they actually are. Drawing on the Kripkean picture of names being passed down socio-linguistic chains may help in conceiving such a counterfactual, but it is not essential. We can imagine a counterfactual new-born baby, who we can call *Arisnotle*, being baptised – or given the name via the ancient Macedonian equivalent – ‘Aristotle’. We can imagine this Arisnotle growing up and doing all the things that (the actual) Aristotle did, and writing all off the works that (the actual) Aristotle wrote, and none of this effecting me and my intrinsic properties in any palpable way. In such a counterfactual, the reference of ‘Aristotle’ on my lips would refer to Arisnotle, according to the Kripkean picture. The reference in both the actual and counterfactual case is determined by the linguistic activities of the original baptisers of ‘Aristotle’ to which I am

connected to via a trans-millennia socio-linguistic chain. In the actual situation, such a socio-linguistic chain connects me to persons who baptised (the actual) Aristotle ‘Aristotle’. In the counterfactual scenario, the chain connects me to persons who baptised the numerically non-identical Aristotle ‘Aristotle’. In any case, regardless of how we suppose that the reference of ‘Aristotle’ is determined, Millianism forces the conclusion that my thought ‘Aristotle was an alcoholic’ would have a different content in a counterfactual scenario in which ‘Aristotle’ refers to someone else, even though *ex hypothesi* that change of reference would have no effect on my intrinsic properties.¹⁴

1.32: Millianism Extended to Natural Kind Terms

Kripke doesn’t seem to make it explicit whether he believes that something akin to Millianism would be true of natural kind terms, i.e. whether he believes that the contribution that a natural kind term makes to the content of a sentence in which it occurs is exhausted by the term’s referent. Kripke does seem to see natural kind terms has having many of the same semantic properties as proper nouns, however. (We can see this throughout the third lecture of *Naming and Necessity* in particular.) So, it does not seem too far-fetched to suppose that Kripke may endorse the view that the contribution that a natural kind term makes to the content of a sentence in which it occurs is exhausted by the term’s referent.

In any case, supposing that Millianism may be expandable to natural kind terms gives a different perspective on why it might be that Oscar and Twin Oscar, of Putnam’s ‘Twin Earth’ (see above, §1.23), will be thinking thoughts with different contents. The reference of the natural kind term ‘water’ on Oscar’s lips will be different to what it is on the lips of Twin Oscar. On Oscar’s lips it will refer to water. On Twin Oscar’s lips it will refer to twater. Thus, if the contribution that a natural kind term makes to the content of a sentence is exhausted by the term’s referent, then the content of Oscar’s and Twin Oscar’s respective sentences in which the term ‘water’ occurs will be different, and thus the content of the thoughts that those sentences express will also be different. In short, Oscar and Twin Oscar will be thinking thoughts with different contents.

1.4: Burge (and Kripke) on Singular Thought

On the face of it, it may seem that the Burgean and Kripkean routes to content externalism are similar, as both are derived via similar considerations of linguistic practice. They are nevertheless the products of opposing points of view.

To begin with Burge rejects Millianism. In ‘Belief *De Re*’, Burge writes:

¹⁴ Cf. Nuccetelli, 2003, pp.9-10.

Several philosophers have held that proper names do not express anything, but merely tag objects. I think this incorrect.

(Burge, 1977, p.344)

In the adjoining footnote (ibid, fn.7), Burge gives two reasons for his dismissal of what he calls “the ‘tag’ view”. The first is “Frege’s paradox of identity”, (which is one of the aforementioned issues that pertains to the ways in which thoughts and attitudes feature in our cognitive economies – see above, §1.31). The ‘tag’ view – i.e. Millianism – fails to explain why statements such as ‘Hesperus is Phosphorus’ are informative, and why statements such as ‘Hesperus is Hesperus’ are trivial; why the former has a cognitive value whilst the latter does not. The second reason is that Millianism is incompatible with Burge’s view that names are essentially disguised indexicals. Briefly, according to Burge’s view, names have a predicative use and thus also have a demonstrative use. Aristotle, the great philosopher of antiquity, was an Aristotle. It is also the case that Aristotle Onassis was an Aristotle. In either instance, we can make the move from the predicative use to the demonstrative use and refer to the Aristotle in question as ‘*that* Aristotle’. According to Burge, when I say: ‘Aristotle was an alcoholic’, what I really mean is: ‘*That* Aristotle [i.e. not the one who was married to Jackie Kennedy] was an alcoholic’.¹⁵

In addition, there is also a deeper reason why Millianism and Burge’s views are in conflict. For Burge what are known as *singular thoughts*¹⁶ – or alternatively *de re* thoughts – have contents that are not individuated by objects of which they are about. If Millianism is correct, then there will be singular thoughts that have contents that are individuated by the objects of which they are about.

Consider the opening two paragraphs of ‘Other Bodies’, which I quote in full:

It is fairly uncontroversial, I think, that we can conceive a person’s behaviour and behavioural dispositions, his physical acts and states, his qualitative feels and fields (all non-intentionally described) as remaining fixed, while his mental attitudes of a certain kind – his *de re* attitudes – vary. Thus we can imagine Alfred’s believing of apple 1 that it is wholesome, and holding a true belief. Without, altering Alfred’s dispositions, subjective experiences, and so forth, we can imagine having substituted an identically appearing but internally rotten apple 2. In such a case, Alfred’s belief differs, while his behavioural dispositions, inner causal states, and qualitative experience remain constant.

¹⁵ See Burge, 1973. Also see Segal, 2001.

¹⁶ For an anthology of papers on singular thought see Jeshion, 2010.

This sort of point is important for understanding mentalistic notions and their role in our cognitive lives. But, taken by itself, it tells us nothing very interesting about mental states. For it is easy (and I think appropriate) to phrase the point so as to strip it of immediate philosophical excitement. We may say that Alfred has the same belief content in both situations. It is just that he would be making contextually different applications of that content to different entities. His belief is true of apple 1 and false of apple 2. The *nature* of his mental state is the same. He simply bears different relations to his environment. We do say in ordinary language that one belief is true and the other is false. But this is just another way of saying that what he believes is true of the first apple and would be false of the second. We may call these relational beliefs different beliefs if we want. *But differences among such relational beliefs do not entail differences among mental states or contents, as these have traditionally been viewed.*

(Burge, 1982, pp.82-3, my emphasis)¹⁷

Compare now Burge's reading¹⁸ of his 'Alfred' case with the Millian/Kripkean reading of my 'Aristotle/Arisnotle' case (see above, §1.31). Burge presents his 'Alfred' case in the hope that it excites the intuition that any singular thought or propositional attitude entertained or held by Alfred will have the *same* content in the counterfactual situation in which the wholesome apple 1 is replaced with the rotten apple 2, and he expects that we draw the conclusion that the same will be true *mutatis mutandis* in the case of *all* singular thoughts and attitudes. In contrast, if Millianism is correct, then it will be the case that my singular thought 'Aristotle was an alcoholic' will have a *different* content in

¹⁷ Page numbers correspond to Burge 2007.

¹⁸ Segal holds essentially the same view on singular thought as Burge – see Segal 2000, pp.3-4. Everyday Abraracourcix has the belief that tomorrow the sky will fall on his head. For Segal there is a looser sense of the word 'content' under which the belief that Abraracourcix has on Tuesday will have a 'content' that is different to the belief he had on Monday in that the beliefs will have different truth conditions. But under the philosophically proper sense of *content* the Monday and Tuesday beliefs will have the same content. The *re* changes but the content – as the term properly should be understood – does not. In the Segal case, the relevant *res* are days, in the Burge case the *res* are pieces of fruit, but the principle is the same. In the Segal case, the *re* changes from Tuesday to Wednesday but the content remains the same. In the Burge case, the *re* changes from wholesome apple 1 to rotten apple 2 but the content remains the same. A consequence of the Burge/Segal view on singular thought is that in the case of singular thought *content* – as it should be properly understood – is divorced from truth conditions – this is why Abraracourcix's beliefs have a different 'content' in the aforementioned looser sense. Abraracourcix's belief on Monday will be true if and only if the sky falls on his head on Tuesday. Whereas, Abraracourcix's belief on Tuesday will be true if and only if the sky falls on his head on Wednesday. Yet, both beliefs have the same *content*. Likewise, Actual Alfred's belief or thought that *that apple is delicious* (say) will be true if and only if apple 1 is delicious. Whereas, the equivalent belief or thought of Counterfactual Alfred would be true if and only if apple 2 is delicious. Yet, there is no change of *content* from the actual to the counterfactual.

the counterfactual situation in which (the actual) Aristotle is replaced with (the counterfactual non-numerically-identical) Arisnotle. Burge seems to offer a way one can be a content externalism, whilst still agreeing adamantly with Frege that Mont Blanc *with all its snow fields* cannot be a propositional constituent. In contrast, Millians are Millians and hence are content externalists essentially because they are sympathetic to Russell's¹⁹ position that Mont Blanc with all its snow fields *can* be a propositional constituent. For Millians content is inextricably tied to object. Thoughts such as 'Aristotle was an alcoholic' owe their *very* content to Aristotle. Changing the object of the thought from Aristotle to Arisnotle changes the content of the thought. According to the Burgean view, the content of thoughts can be and is affected by things that are external to the thinker. For instance, if we – reversing a famous example of Putnam²⁰ – place Alfred in a world where what are known as 'apples' are not grown but are in fact a sort of man-made confectionery, then that might change the content of Alfred's thoughts. However, placing Alfred in a world where the wholesome apple 1 is replaced by the rotten apple 2 would not.²¹

¹⁹ In this thesis, as seems to be common throughout the philosophical literature, when we talk of Russell we'll be talking exclusively of the *earlier* Russell, i.e. the Russell of works such as 'On Denoting', 'Knowledge by Acquaintance and Knowledge by Description' and *The Problems of Philosophy* (Russell, 1905; 1910-11; Russell 1912), i.e. Russell before his adoption of *neutral monism* (see for example Russell, 1921). The famous exchange between Frege and Russell can be found in Beaney, 1997, pp.290-3.

²⁰ Putnam actually attributes the example to Rogers Albritton. It's the one that involves considering a counterfactual in which pencils are organisms, and are therefore grown and not made (Putnam, 1975, p.242).

²¹ This Chapter has focused on the content externalism of Burge and Kripke, as will the rest of the thesis. Other prominent content externalists, which aren't to feature in this present thesis, include Evans and McDowell (Evans, 1982; McDowell, 1986), and also Davidson (Davidson, 1987; 2001a; 2001b; – see also Bridges, 2006), and *maybe* Putnam (Putnam, 1975).

The versions of content externalism offered by Evans and McDowell can be viewed as being versions of the Russellian account of singular propositions that allow for a broader set of possible propositional constituents than the narrow set permitted in Russell's original framework. Evans and McDowell allow for there to be *illusion of thought*, and that broadens the set of available propositional constituents. (See in particular the first two sections of McDowell 1986 (pp.137-141).) Russell, in the case of singular propositions, saw content has being tied to object. There are thoughts that require the object of which they are about for their very content. No object, no thought. In other words, objects can be propositional constituents. For Russell, an object can be such a propositional constituent as long as the thinker can 'know which' object the thought concerns – the thinker must be able to differentiate the object from the rest of creation. But that was not the only restriction Russell placed on *propositional-constituent-hood*. Russell did not countenance that there could be *illusions of thought*. This in turn severely limited the scope of the what could be a propositional constituent to only include things that cannot fail to exist if they are contained within one's thoughts. It meant, in particular, that external world objects cannot be propositional constituents, because if it so turned out that the external world object in question did not exist – if the thinker was suffering from a perceptual illusion (say) – then the thinker would be suffering from illusions of thought, and Russell did not countenance that possibility. In one sense Mont Blanc *with all of its snowfields* is the sort of thing that could be a propositional constituent. In another sense Mont Blanc (with all of its snowfields) is precisely the sort of thing that cannot be a propositional constituent, because it might not exist. Evans and McDowell *do* countenance the possibility of illusions of

thought. This allows for external world objects that we have had perceptual contact with to enter into our thoughts as propositional constituents. Content externalism – as I define it in this chapter (§1.1) – follows directly from this.

The content externalism of Evans and McDowell is similar to Kripke's content externalism – as I characterise it in §1.3 – in that both involve *singular thoughts* viewed as having (external-world)-object-dependent contents. (They are two forms of what Brown calls "singular anti-individualism" – see Brown, 2004, Ch.1; see also my Chapter 3, §3.22.) A key difference is that Evans and McDowell reject the Millianism that drives Kripkean content externalism and essentially embrace Fregeanism. (See in particular the third section of McDowell 1986 (pp.141-3).) Allowing for illusions of thought enables a perfect marriage of Frege and Russell. (Or at the very least it enables Russell to live in Sinn with Frege!) Indeed, if we, on the hand, disallow illusion of thought, then that means occurrent sense impressions of Venus would need to be propositional constituents instead of Venus itself. Thus, a thought about Venus during the early morning will automatically have a different content to a thought about Venus in the early evening, as they will involve different propositional constituents. Allowing illusion of thought, on the other hand, means that Venus (itself), which can enter one's perceptual field both during early morning and during early evening, may enter into one's thoughts as a propositional constituent. If Venus (itself) is a propositional constituent, then it will be a propositional constituent of both the morning thought and the evening thought. Thus, if the two thoughts are to have differing contents as the standard arguments demand that they should, then there must be something else that differentiates them. This differentiating something will be Venus's respective *modes of presentation* (which are not to be confused with Venus-picking-out definite descriptions but are something else entirely.) Or so the Fregean-cum-neo-Russellian/Evansian/McDowellian story goes.

Evans and McDowell's crucial endorsement of the view that there is such a thing as illusion of thought possibly gives their brand of content externalism a level of immunity against the sort of reductio arguments considered in Chapters 4-6. If we were/are happy to say that what may seem to me to be the thought that I'd express 'water is wet' (say) may and may quite easily be an illusion of thought – if (say) it turned out water did not exist – then there is possibly less pressure to suppose that I must know in any sort of special way that I am thinking what I believe I am thinking.

Davidson is committed to content externalism – as I define it in this present chapter – but he is so because of a commitment to a more radical and overarching thesis, which is that the very possibility of thought requires the right sort of environmental causal relations, and which is given its most striking explication in Davidson's famous 'The Swampman' thought experiment:

Suppose lightning strikes a dead tree in a swamp; I am standing nearby. My body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into my physical replica. My replica, The Swampman, moves exactly as I did; according to its nature it departs the swamp, encounters and seems to recognize my friends, and appears to return their greetings in English. It moves into my house and seems to write articles on radical interpretation. No one can tell the difference.

But there *is* a difference. My replica can't recognize my friends; it can't *recognize* anything, since it never cognized anything in the first place. It can't know my friends' names (though of course it seems to), it can't remember my house. It can't mean what I do by the word 'house', for example, since the sound 'house' it makes was not learned in a context that would give it the right meaning – or any meaning at all. Indeed, I don't see how my replica can be said to mean anything by the sounds it makes, nor to have any thoughts. (Davidson, 1987, p.443-4 (p.91 in Ludlow and Martin, 1998))

The ideas put forth in this thesis may be brought to bear on such other versions of content externalism, but the focus of this present thesis will remain mostly on the versions of content externalism advanced by Burge and by

Kripke. Having said that, anything in the thesis that is applicable with respect to content externalism *simpliciter* – which as I define is the thesis that two qualitatively identical subjects could be thinking thoughts with different contents – should automatically be applicable with respect to *all* forms of content externalism. In looser words, there may be times in the thesis in which the version of content externalism doesn't matter.

The thesis refers frequently to Putnam's famous 'Twin Earth' thought externalism. It is not clear, however, whether Putnam himself at least at around the time he wrote the 'The Meaning of 'Meaning'' (Putnam, 1975), which is the paper from which the thought experiment originates, was in fact a content externalist – as I define the term. The thought experiment was originally intended to show the inconsistency of two assumptions. The first assumption is "[t]hat knowing the meaning of a term is just a matter of being in a certain psychological state" (ibid., p.219), where "psychological state" is understood to be a *narrow* state, i.e. as denoting an intrinsic property of a subject. The second assumption is "[t]hat the meaning of a term (in the sense of 'intension') determines its extension" (ibid.). If the 'Twin Earth' thought experiment does indeed demonstrate that one of these two assumptions must be rejected, then arguably it would only be a rejection of the first assumption that could lead to content externalism. Also – and relatedly – Putnam's contention that the word 'water' is indexical, or is like an indexical, possibly precludes him being a content externalist (see §II of Burge, 1982.)

Chapter 2: Slow Switching I: The Redeployment Thesis

2.0: Introduction to Chapter 2

This present chapter serves a dual purpose. It characterises what in the literature is the principle positive case for *compatibilism* – the thesis that content externalism is compatible with traditional notions of privileged access. The four subsequent chapters all consider negative cases against compatibilism. This chapter also sets up something of a *cliff hanger* that is not to be resolved until the final chapter of this thesis.

2.1: The Redeployment Thesis and the Incomplete Argument for Compatibilism

What we can call the *Redeployment Thesis*, after Peacocke (Peacocke, 1996 – also see Sawyer, 2001, pp.121-2)²², states that we should be able to form second-order beliefs about our first-order thoughts such that the content of such second-order beliefs should be *reflective* of the content of the first-order thoughts of which they are about in such a way that guarantees truth. (The concepts that are involved in the first-order thoughts are *redeployed* in the corresponding second-order beliefs about those thoughts.)

Consider Oscar and his atom-for-atom Twin-Earthling duplicate, Twin Oscar. Whilst Oscar is on Earth thinking that water is wet, Twin Oscar will be on Twin Earth thinking that twater is wet. Should Oscar reflect upon what he is currently thinking he should be able to form the belief that he is thinking the thought that water is wet. Twin Oscar will at the same time also form a belief about what he is thinking. The redeployment thesis states that the content of Oscar's and Twin Oscar's respective

²² We are restricting 'the redeployment thesis' to the special case of occurrent thought. Peacocke and Sawyer characterise it as applying to a larger range of mental states that includes propositional attitudes.

beliefs will both have content that is reflective of the content of their respective thoughts, in a way that ensures the truth of both beliefs. Oscar's belief will be that he is thinking that water is wet. It will indeed be the case that Oscar is thinking that water is wet. Thus, Oscar's belief will be true.

Symmetrically, Twin Oscar's belief will be that *he* is thinking that twater is wet. It will indeed be the case that Twin Oscar is thinking that twater is wet.

If it is indeed the case that Oscar's and Twin Oscar's first-order thoughts differ in content, as the content externalist maintains, then Oscar's and Twin Oscar's respective second-order beliefs about those first-order thoughts will also differ in content but they will both be true.

Twin Oscar would, in expressing *his* belief, use the same words as Oscar would to express his belief. Twin Oscar would say: "I believe that I am thinking that water is wet". But, those words on Twin Oscar's lips will mean something different to what they would on the lips of Oscar.

The redeployment thesis is neutral with respect to explanations of why Oscar and Twin Oscar will be thinking different thoughts. In other words, it should apply with respect to both *Kripkean externalism* and *Burgean externalism* (see previous chapter).

Consider also the case of *Slow-Switching Sally*, which is to feature heavily in the next chapter. Slow-Switching Sally was born here on Earth but, completely unbeknownst to her, has spent a significant portion – roughly half (say) – of her lifetime on the *Twin Earth* of Putnam's famous thought experiment. Sally has for some mysterious reason been subject to switches between the two planets. After spending days, months or years (even) on one planet she will undergo a switch and will be transported across the vastness of space to the other planet. She may spend seven months on Earth (say) and then be transported to Twin Earth, and then after spending thirteen months (say) on Twin Earth will return to Earth, and so on and so on. Slow-Switching Sally is totally unaware that she periodically takes such incredible universe-spanning journeys. As far as Slow-Switching Sally is concerned it is as if she spends the entirety of her time on Earth.

Both the Kripkean and Burgean externalist should concede that sometime not long after a switch the content of the thoughts (and propositional attitudes) that Sally would express using the word 'water' will undergo a change. The thought that Sally would express 'water is wet' whilst on Earth will be different in content to the thought that she'd express 'water is wet' whilst on Twin Earth.

For the Burgean externalist the thoughts will be different in content because Sally has switched between a socio-linguistic community in which 'water' means one thing to a socio-linguistic community in which 'water' means something else. Sometime after a switch, Sally will cease to be a member of one of the socio-linguistic communities and become again a member of the other socio-linguistic community.

The Kripkean externalist should concede 'water' will *name* different natural kinds depending upon whether Sally is on Earth or Twin Earth. Sometime after a switch from Earth to Twin Earth 'water' on Sally's lips will cease to name the natural kind water and start to name the Twin Earthly

equivalent natural kind twater. Thus, it should follow that Sally will be thinking thoughts with different contents depending upon which planet she is on.

(There may be complexities with respect to how Kripkean views of reference mesh with the case of Slow-Switching Sally that we need not here fully go into. Indeed, if we were to suppose that ‘water’ on the lips of Sally acquires its reference via Sally being at the end of a Kripkean socio-linguistic chain that has at its other end some mythic original baptisers of water, then the reference of ‘water’, on the lips of Sally, is in danger of being the same on Twin Earth as it is on Earth. If so, then the contents of Sally’s ‘water’ thoughts will be the same on Twin Earth as they are on Earth. This is implausible, however. The Kripkean externalist cannot be committed to such a story of how ‘water’ gets its reference given that it has that consequence.)

According to the redeployment thesis when Sally is on Earth and forms a second-order belief about her thinking that water is wet, the content of that belief will be reflective of her first-order thought in such a way that guarantees the truth of her belief. Symmetrically, when Sally is on Twin Earth, what’ll be the equivalent second-order belief – that she is thinking that twater is wet – is guaranteed to be true, because *its* content will be reflective of the content of the first-order thought – that twater is wet.

For the redeployment thesis to be true it must be the case that there is a *linking belief-forming process* by which a subject comes to form a belief about what she is presently thinking, and that links that second-order belief to the first-order thought that that second-order belief will be about.

If such a process can be successfully identified, and if the process can be said to be *a priori*, then that would – prima facie at least (see subsequent chapters) – present a strong case for supposing that we’d be able to know the contents of our thoughts a priori even if content externalism is true – that the two things could well be compatible.

Likewise, if there is such a belief-forming process, and that process is a process by which a subject can acquire guaranteed-to-be-true second-order beliefs about her own thoughts, but would not also be a process by which a third-party can acquire guaranteed-to-be-true beliefs about what a subject is thinking, then that would – prima facie at least – present a strong case for supposing that we’d be able to know what the contents of our own thoughts are better than anyone else. We’d have a *privileged access* to the contents of our own thoughts, even if content externalism is true

In the next section we’ll be considering a candidate for what such a process could consist of that is presented by Burge (Burge, 1988). Burge’s candidate process could potentially be a priori. It would most definitely constitute a privileged method by which a subject can ascertain the content of her own thoughts.

The next chapter (Chapter 3) considers the *discrimination argument*. Proponents of the discrimination argument contend that slow-switch cases such as that of Slow-Switching Sally mean that the guaranteed-to-be-true beliefs about our thought contents that we would have per the

redeployment thesis won't constitute knowledge – even though such beliefs would be guaranteed to be true. In this present chapter we put that worry to one side.

Subsequent chapters (Chapters 4-6) consider *reductio* arguments that if successful would seem to demonstrate that if content externalism were true, then there being an *a priori* process by which we can ascertain the contents of our thoughts would seem to mean that we'd also be able to ascertain things about the external world in a way that we ought not to be able to. Again, we put such worries to one side for the moment.

2.2: The Burgean Candidate

If I judge that: *I am thinking that writing requires concentration*, then it will indeed be the case that I am thinking that writing requires concentration. By the very act of making the judgement I am thinking the thought. The second-order belief that I will have formed by judging that I am thinking that writing requires concentration is guaranteed to be true, because the very process by which the belief has been formed has made the belief true. The judgement is *self-verifying*. It is through such *cogito-like* judgements that Burge proposes that we have an ultra-reliable way of ascertaining the content of our thoughts (Burge, 1988 – see also Boghossian, 1989, §III; Peacocke, 1996; MacDonald, 1998; Sawyer, 2001; Spicer, 2009; 2011; Brueckner, 2011; 2015)²³. A subject can judge that she is

²³ None of the cited authors seem to fully countenance what (I think) is the most pressing problem for Burge, which is the 'no provenance' objection – see next section.

That is, apart from Boghossian – see fn. 24.

Burge 1996, Peacocke 1996, and MacDonald 1998 all attempt to place the notion of *cogito-like* judgements in the wider context of the Burgean notion of 'entitlement to self-knowledge'. This aspect is something else that this present thesis omits. Discussing it properly would require that we go beyond what is the very narrow focus of this present thesis. This does not necessarily mean that such considerations cannot be brought to bear on the specific lines of argument of this thesis. It just means that that'll have to wait until potential sequels.

Sawyer 2001 is in a similar vein. In order to better elucidate the – actually very limited – role *cogito-like* judgements play in the bigger Burgean picture of self-knowledge Sawyer demonstrates that three putative counterexamples *aren't* counterexamples. The first of which comes from Gallois:

The...thesis that thinking that one thinks that *p* entails thinking that *p*...seems implausible. Surely I can mistakenly ascribe a thought to myself. Convinced by a dogmatic psychoanalyst I think that I think that my mother does not love me even though I have no such first-order thought.

(Gallois, 1996, p. 173; quoted in Sawyer, 2001, p.111)

Sawyer (correctly) identifies the mistake that Gallois is making which is that he doesn't understand the very restricted application of the Burgean method:

thinking a thought with a certain content, because she is making that judgement it will indeed be the case that she is thinking a thought with that content, and thus the subject judges correctly.

If Slow-Switching Sally, whilst on Earth, were to judge that she was thinking that water is wet it will be the case that she is thinking that water is wet, and thus she judges correctly. Her judging that she is thinking that water is wet is the very thing that makes it the case that she is thinking the thought that water is wet. Now, both Burgean and Kripkean externalists should contend – although for slightly different reasons – that when Sally is on Twin Earth and makes the equivalent judgement, that judgement will have a different content to what it would be were Sally on Earth. Whereas on Earth Sally would judge that she is thinking that water is wet, when on Twin Earth she instead judges that she is thinking that twater is wet. And, it will indeed be the case that she is thinking that twater is wet, as her judging that she is thinking that twater is wet makes it the case that she is thinking that twater is wet.

Slow-Switching Sally would not be able to discriminate between a situation in which she is on Earth and judges that she is thinking that water is wet, and an equivalent situation in which she is on Twin Earth and judges that she is thinking that twater is wet. She would not be able to tell the difference. Thus, Sally cannot discriminate between the situation in which she is thinking that water is wet from the situation in which she is thinking that twater is wet. Proponents of the discrimination argument – which is to be the topic of the next chapter – maintain that because of this Sally cannot be said to *know* what she is thinking. It is worth emphasising, nevertheless, that in either situation there would be no way in which Sally's judgement would be false. Proponents of the discrimination argument in effect contend that infallible true belief is not sufficient for knowledge.

It needs emphasising that Burge's proposed process for thought self-ascription would have a very limited application. However, that limited application is the entire focus of this thesis. The process would only apply with respect to what a subject is *presently* thinking. Burge's proposed

...we need to distinguish between the thought *I am thinking my mother does not love me* and the thought *I think my mother does not love me*...the former is self-referential and self-verifying in virtue of being so, whereas the latter is an attribution of a standing state with which it can at best be contingently related.
(Sawyer, 2001, p.119)

The two other supposed counterexamples, which come from Goldberg and Bernecker (Goldberg, 2000; Bernecker, 1996), essentially make the same mistake.

The Spicer and Brueckner exchange (Spicer, 2009; 2011; Brueckner, 2011; 2015) is very technical. It concerns an argument, advanced by Spicer, that aims to demonstrate that cogito-like judgements are not in fact self-verifying (essentially because cogito-like judgements do not share the same truth conditions as the first-order thoughts of which they are about). We are, in this present thesis, making a concession to Burge, and his defenders, in assuming that cogito-like judgements would indeed be self-verifying.

method of thought self-ascription could not be used by a subject to ascertain what she was thinking at *any* earlier moment in time. She would not be able to use the method to determine what she was thinking when she stood taking in the *Mona Lisa* during a recent Paris visit (say). Nor could the subject use the method to determine what she was thinking a few moments ago.

Burge's proposal would bestow a subject with a privileged method by which she can ascertain the content of her own thoughts, however. Indeed, this is (hopefully) very easy to see. The necessary asymmetry is obvious. If I judge that *you* are thinking that writing requires concentration, then it may or may not – most probably not – be the case that you are thinking that writing requires concentration. But, if I judge that *I* am thinking that writing requires concentration, then it will be the case that I am thinking that writing requires concentration, because my making the judgement that I am thinking that writing requires concentration makes it the case that I am thinking that writing requires concentration. My making the judgement that you are thinking that writing requires concentration would not make it the case that you are thinking that writing requires concentration.

Since Burge does not seem to explain what the provenance of cogito-like judgments could be – see next section – it is not entirely clear whether they'd actually be a priori. Thus, all we can say is that they would potentially be a priori.

2.3: The 'No Provenance' Objection and the 'No Phenomenology' Problem: A Cliff Hanger

If I judge that I am thinking that writing requires concentration, then it will indeed be the case that I am thinking that writing requires concentration. But what is it that prompts me to make that judgement in the first place? That is something that Burge does not explain.²⁴

Now, surely the reason why I would judge that writing requires concentration should be based on the fact I am thinking that writing requires concentration. I shouldn't just judge that I am thinking that writing requires concentration apropos of nothing. The appropriate event should precede the judgement. That event should presumptively be my thinking that writing requires concentration.

²⁴ C.f. Boghossian:

We are struck by our ability to know, non-inferentially and authoritatively, that a certain mental event has occurred, immediately on its having occurred. We think: Writing requires concentration. And then we know, directly and unproblematically, that that is what we thought. A first-order thought occurs. And we are then able, without benefit of inference, to form a correct judgement about what the thought was. The second-order judgement in these central cases is not self-verifying... How does [Burge's] proposal help explain how they are possible? The fact that had the thought been part of a second-order judgment, then that judgement would have been self-verifying, does not help explain how we are able to know what thought it was, given that it *wasn't* part of such a judgement.

(Boghossian, 1989, p.22)

However, it is not clear how it is that such an event would prompt me to make the judgement that I am thinking that writing requires concentration. As was noted in the opening paragraph of this thesis thoughts produce no phenomenology (§0.0 – see also §0.3). Thus, it is not clear how I would be made aware of the event describable as ‘my thinking that writing requires concentration’ that accordingly causes me to make the judgement that I am thinking that writing requires concentration.

What we have here is a problem for Burge’s account of thought self-ascription, and a problem that we encounter when we try to remedy the situation, which is a problem that *any* account of thought self-ascription would need to overcome. We can call these two problems the ‘*no provenance*’ objection and the ‘*no phenomenology*’ problem. Burge’s proposal is missing an explanation of why I’d make the judgement that I am thinking that writing requires concentration in the first place. This is the ‘no provenance’ objection. When we consider how we might complete Burge, or might construct an alternative, by considering what might cause me to make the judgment that I am thinking that writing requires concentration, our efforts are stymied by the realisation that thought produces no phenomenology. This is the ‘no phenomenology’ problem. We have thus arrived at the aforementioned cliff hanger.

It is worth emphasising that even if we were to suppose that the content of our thoughts is decided by factors that are entirely *internal* – i.e. if we were to reject content externalism – we’d still face the same problems, if, or when, we turn to the question of how thought self-ascription works.

I leave the resolution of this cliff hanger to the final chapter of this thesis. In the interim chapters, we will be in effect taking seriously the possibility that the redeployment thesis is true, and that crucially the linking belief-forming process so required by the redeployment thesis, whether or not it would be as Burge describes, will be a priori. In the next chapter we will be in effect making the concession to compatibilists that the redeployment thesis would be true, even if content externalism is true, and that the requisite linking belief-forming process is a priori. We will be considering whether ‘slow-switching’ cases such as that of Slow-Switching Sally, that would be possible if content externalism were true, would mean that the guaranteed-to-be-true beliefs about our thought contents that we should have per the redeployment thesis won’t always constitute knowledge, if content externalism were indeed true. In the subsequent chapters, we will in effect be making an ex hypothesi assumption that something like the redeployment thesis is true and that the requisite linking process is a priori, along with the ex hypothesi assumption that content externalism is true, to consider whether those two ex hypothesi assumptions lead to absurdities.

Chapter 3: Slow Switching II: The Discrimination Argument

3.0: Introduction to Chapter 3

Consider (again) the case of *Slow-Switching Sally*. Slow-Switching Sally was born here on Earth but at various points in her life history has been routinely transported between Earth and the *Twin Earth* of Putnam's thought experiment (see Chapter 1, §1.23 and §1.32). (We need not concern ourselves with why or how Sally is subject to such incredible universe spanning journeys.) By such switches, Sally has spent a significant amount of her life on both Earth and Twin Earth. She could spend (say) seven months on Earth before being transported to Twin Earth. She could then spend the next eighteen months (say) on Twin Earth, after which she is then transported back to Earth – and so on and so forth. Sally is totally unaware that such an extraordinary thing keeps on happening to her. As far as she's concerned it's as if all her life takes place on only the one planet. If content externalism is true, then it should follow that after Slow-Switching Sally has spent a sufficient amount of time on Twin Earth after a switch, the thought that she'd express as 'water is wet' should change its content. And, the same should happen when Sally returns to Earth, that is, the content should change back to what it originally was on Earth. Whereas when on Earth Sally may be thinking that water is wet, if she happened to be on Twin Earth she would instead be thinking that twater is wet.

It is argued that because Slow-Switching Sally would be unable to *discriminate through a priori means* her being on Earth and thinking that water is wet from her being on Twin Earth and thinking that twater is wet she cannot be said to know a priori that she is thinking that water is wet, when she is on Earth and thinking that water is wet. Thus, if content externalism were true there would be cases in which someone does not know a priori what they are thinking. This essentially is what has become known as the *discrimination argument* (Brown, 2004), and is attributable to Boghossian (Boghossian, 1989).

3.1: The Discrimination Argument

The dialectical impetus of the discrimination argument is provided by Goldman's famous 'fake barn' case (Goldman, 1977).

Bert is driving through the countryside when he happens to see a barn. Bert forms the true belief that what he is looking at is a barn.

There will be innumerable possible worlds within which it won't be the case that Bert is looking at a barn. For instance, there will be worlds in which a crop silo stands where the barn stands in actuality. There will also be more distant possible worlds in which what Bert is looking at is not in fact a barn but is a hollow film-set-like structure that has been purposefully made to look exactly like a barn. There will also be even more distant worlds in which Bert is a BIV (a brain in a vat) and the grand computer simulation to which he is plugged into is leading him to falsely believe that he is looking at a barn.

In a world in which Bert was looking at a crop silo, as opposed to a barn, things would seem different to Bert to how they actually seem. If Bert were looking at a crop silo and not a barn, he'd be very much aware of it. He can *discriminate* his looking at a barn from his looking at a crop silo. Thus, Bert is able to rule out the possibility that he is looking at a crop silo.

In a world in which Bert was looking at a *fake barn*, i.e. a hollow film-set-like structure that has been purposefully made to look exactly like a barn, things would *not* seem any different to Bert to how they actually seem. Bert cannot rule out the possibility that he is looking at a fake barn.

In a world in which Bert was a BIV, things, again, would not seem any different to Bert to how they actually seem. Bert cannot rule out the possibility that he is a BIV.

Consider now the case of Russ. Like Bert, Russ is driving through the countryside and happens to see a barn, and forms the true belief that he is looking at a barn. Unlike in the case of Bert however, there is something peculiar about the region that Russ is driving through. Russ is driving through a region where, unbeknownst to Russ, there is, for some reason or other, a large number of hollow film-set-like structures that have purposefully been made to look exactly like barns, i.e. there is a large number of fake barns.

Russ would, like Bert, be able to rule out, based on how things seem to him, the possibility that he is looking at a crop silo. He, also like Bert, would not be able to rule out the possibility that he is looking at a fake barn, nor would he be able to rule out the possibility that he is a BIV.

Standardly it is held that Bert would know that he was looking at a barn, whereas Russ could not know that he is looking at a barn.

A necessary condition for knowledge is that a putative knower must be able to rule out all and any *relevant alternatives*. Both Russ and Bert must be able to rule out any and all possibilities in

which it is not the case that they are looking at a barn that are *relevant*, if it is to be said that they are to know that they are looking at a barn.

Neither Russ nor Bert can rule out the possibility that they are a BIV. But, in neither the case of Russ nor Bert are such possibilities relevant. The possibility that Bert is looking at a crop silo constitutes a relevant alternative, as does the possibility that Russ is looking at a crop silo. But both Bert and Russ are able to rule out such possibilities.

Neither Bert nor Russ can rule out the possibility that they are looking at a fake barn. They would not be able to discriminate between their looking at a real barn from their looking at a fake. In the case of Bert, this will not in itself prevent his true belief that he is looking at a barn from qualifying as knowledge. But it does in the case of Russ. Russ does not know that he is looking at a barn *because* he cannot discriminate his looking at a barn from his looking at a fake. Of course, neither can Bert. But in the case of Bert that possibility does not constitute a relevant alternative. It does in the case of Russ because Russ is driving through ‘fake barn’ country, and it could very easily have been the case that he was looking at a fake barn as opposed to a real one.

The *redemption thesis* (see Chapter 2, §2.1) states that when Oscar – one half of the dramatis personae of Putnam’s original ‘Twin Earth’ (see Chapter 1, §1.23 and §1.32) – reflects upon what he is thinking and forms, presumptively through a priori means (see Chapter 2), the belief that he is thinking that water is wet, that belief is guaranteed to be true (because its content will be reflective of the thought of which it is about).

Now, in the case of Oscar, there will be counterfactual scenarios in which that guaranteed-to-be-true belief would be false. If sometime before the thought, Oscar had been the unwitting victim of the sort of switches Slow-Switching Sally periodically undergoes, and had been transported to Twin Earth, then Oscar would not have been thinking what he was actually thinking. He would be thinking that twater is wet as opposed to thinking that water is wet. In such a counterfactual scenario, Oscar would not form the belief that he is thinking that water is wet. He’d form what would be the true belief that he is thinking that twater is wet. But it would still, nevertheless, be the case that he is thinking something other to what he is actually thinking, and in which his actual belief would be false. Oscar is not able to discriminate between the actual and such a counterfactual scenario. Oscar is not able to discriminate through a priori means between his being on Earth and thinking that water is wet from his being on Twin Earth and thinking that twater is wet. This, following Goldmanian logic, would disqualify Oscar’s belief from counting as knowledge *if* such a counterfactual scenario constitutes a relevant alternative. But it does not. Symmetrically, all this will, of course, be true, *mutatis mutandis*, of Twin Oscar – the other half of the dramatis personae of Putnam’s ‘Twin Earth’ – as well.

Similarly, when Arthritic Alf – the star of Burge’s famous ‘arthritis’ thought experiment (see Chapter 1, §1.22) – reflects upon what he is thinking, and forms a belief with respect to his ‘arthritis’

thought, via similar *ex hypothesi a priori* means, there will be counterfactual scenarios in which that belief is false. In counterfactual scenarios in which ‘arthritis’ is defined by medical practitioners and the like to encompass a wider range of ailments, it will not be the case that Alf is thinking what he is actually thinking. In such counterfactual scenarios everything would seem to Alf exactly as it actually does. But such counterfactual scenarios would not constitute relevant alternatives. Thus, such counterfactual scenarios do not preclude Alf’s belief from qualifying as knowledge.

Consider now the case of Slow-Switching Sally. The redeployment thesis states that Slow-Switching Sally, regardless of what planet she is on, should be able to form, presumptively via *a priori* means, guaranteed-to-be-true beliefs about what she is thinking. Suppose she is on Earth and forms the belief that she is thinking that water is wet. Like in the case of Oscar, there will be counterfactual scenarios in which *her* belief would be false. If she had undergone a switch sometime before her thought, then she would not have been thinking what she was actually thinking. She would be thinking that twater is wet as opposed to thinking that water is wet. In such a counterfactual scenario Sally would not form the belief that she is thinking that water is wet. Sally would instead form what would be the true belief that she is thinking that twater is wet. But it would still, nevertheless, be the case that she is thinking something other to what she is actually thinking. Sally is not able to discriminate *a priori* between the actual and such a counterfactual scenario. She is not able to discriminate between her being on Earth and thinking that water is wet from her being on Twin Earth and thinking that twater is wet. This, following Goldmanian logic, should disqualify Sally’s belief from counting as knowledge, *if* such a counterfactual scenario constitutes a relevant alternative. In the case of Oscar such a counterfactual scenario does not constitute a relevant alternative, as Oscar is not being routinely switched between Earth and Twin Earth like Sally. Since Sally is being periodically switched from Earth to Twin Earth, a counterfactual scenario in which Sally is on Twin Earth thinking that twater is wet does constitute a relevant alternative. Hence, following Goldmanian logic, Sally’s belief that she is thinking that water is wet cannot count as knowledge.

The cases of Oscar and Slow-Switching Sally and the cases of Bert and Russ are exactly analogous. Both Bert and Russ are unable to discriminate between the actual and a counterfactual scenario in which they are looking at a fake barn. Both Oscar and Sally are unable to discriminate the actual from a counterfactual scenario in which they are on Twin Earth thinking that twater is wet. For Bert a counterfactual scenario in which he is looking at a fake barn does not constitute a relevant alternative. But, since Russ is driving through ‘fake barn’ country, a counterfactual scenario in which *he* is looking at a fake barn does constitute a relevant alternative. Thus, Russ’s belief that he is looking at a barn is precluded from qualifying as knowledge, though Bert’s belief is not. For Oscar a counterfactual scenario in which he is on Twin Earth thinking that twater is wet does not constitute a relevant alternative. For Sally, on the other hand, a counterfactual scenario in which *she* is on Twin Earth thinking that twater is wet *does* constitute a relevant alternative. Since Sally is routinely

switched between the two planets, it could very easily have been the case that she were on Twin Earth thinking that twater is wet. Thus, Sally's belief is precluded from qualifying as knowledge, whereas Oscar's is not. This is the discrimination argument.

Proponents of the discrimination argument concede that Slow-Switching Sally per the redeployment thesis is able to, either through the Burgean mechanism characterised in the previous chapter (§2.2) or otherwise, form guaranteed-to-be-true beliefs about what she is thinking. Proponents of the discrimination argue that, nevertheless, such guaranteed-to-be-true beliefs cannot constitute knowledge, because of the knowledge-precluding relevant alternatives that there are in the case of Slow-Switching Sally.

3.2: Responses to the Discrimination Argument

3.2.1: Warfield's Response

Warfield (1992) argues that the discrimination argument would only have any teeth if slow-switching cases were actual real-world occurrences. They are not. In the case of Slow-Switching Sally, when she thinks that water is wet, her thinking that twater is wet is a relevant alternative. But when *we* think that water is wet, although we may not be able to distinguish this from our thinking that twater is wet, our thinking that twater is wet does not constitute a relevant alternative. We live in the real world, and not in some contrived philosophical thought experiment.

In response to Warfield, Ludlow (1995) has argued that, *Burgean externalism* (see my Chapter 1) entails a "prevalence" of real-world slow-switch cases. We, almost certainly, are not routinely switched from Earth to Twin Earth. We are all potentially victims of slow-switch cases of another sort, however. Consider the case of Biff. Biff, due to his job, spends a significant proportion of his life in either America or Britain. He may whilst nominally living in Britain keep on having to spend two-to-four-month stints in America. Unbeknownst to Biff, in America 'chicory' has a different meaning to what it has in Britain. It refers to a similar but different species of edible plant. Biff concerned about his health may have thoughts to the effect that he should eat more vegetables. He may think to himself specifically that *eating some chicory might be a good idea*. Ludlow supposes that the relationship between Biff and the term 'chicory' is analogous to the relationship that Alf the Arthritis sufferer has with the term 'arthritis'. Biff will have a moderate degree of linguistic competence with respect to the term 'chicory', even though he won't know the term's precise definition. Following Burgean logic, this should mean that the content of thoughts that Biff would express using the term 'chicory' should be determined by his wider socio-linguistic community in the same way that Alf's wider socio-linguistic community determine what thought Alf would express when using the term 'arthritis'. An adherent to Burgean externalism should maintain that the thoughts

that Biff would express using the term would have a different content depending upon whether Biff is in America or in Britain.

Biff would be able to discriminate between his being in America from his being in Britain. However, the way in which Biff does so will be via a posteriori judgments about his external environment – about which side of the Atlantic Ocean he is on. If, however, we suppose that Biff is able to make a priori judgements with respect to what he is thinking (see my Chapter 2), then Goldmanian logic dictates that such judgements cannot constitute knowledge.

Ludlow suggests a number of other instances of *everyday slow-switching cases*. One such is the case of Griff. Griff spends some of his time in philosophical circles where he often hears ‘pragmatists’ being spoken of. The word ‘pragmatists’ has a different usage in such philosophical circles to what it has in *non*-philosophical circles. In the former the word is used to refer to the likes of Pierce and James. In the latter it is used to refer to those who are guided more by practical considerations than by ideals. The Burgean externalist should, similarly, maintain that depending upon whether Griff is cavorting with his philosopher pals, or whether he is chatting with his other more pedestrian acquaintances, Griff could entertain thoughts with different contents about ‘pragmatists’.

In response to Ludlow, Brown (2004, pp.138-142) has argued that the sort of slow switch cases invoked by Ludlow cannot be as prevalent as Ludlow supposes.²⁵

Brown delineates three criteria that a bona-fide Ludlowian slow-switch case would need to meet. There would need, firstly, to be a subject who “is slowly switched between two linguistic communities that share a single word but define it differently” (ibid., p.139). Secondly, that subject

²⁵ Warfield has given what is frankly a bizarre counter-response to Ludlow (Warfield, 1997). Warfield notes that at best what Ludlow’s everyday slow-switching cases demonstrate is that in some possible worlds including the actual world content externalism being true would be incompatible with our having a priori knowledge of the content of our own thoughts. What it doesn’t show is that in *every* possible world externalism is incompatible with our having a priori knowledge of our thought content. The sort of everyday slow-switch cases described by Ludlow won’t occur in every possible world:

The problem that Ludlow and Boghossian face is not merely a problem of detail. The kind of argument they offer couldn’t possibly show that externalism is incompatible with privileged [a priori – see my Chapter 7, §7.2] self-knowledge. To show that these doctrines are incompatible one needs to show that *every* possible world in which externalism is true is a world in which individuals do not have privileged [a priori] self-knowledge. Boghossian shows *at most* that *some* possible worlds are worlds in which externalism is true and individuals lack privileged self-knowledge and Ludlow shows *at most* that *one* world, the actual world, is a world in which externalism is true and (some) individuals lack privileged self-knowledge.
(ibid., p.284, Warfield’s emphases)

This seems a considerable *retreat* on the part of Warfield. Surely compatibilists should want it shown to be the case that in the *actual* world if externalism is true then we would (all) still be able to know a priori what the content of our thoughts are.

would need to be “ignorant of this linguistic difference” (ibid.). Thirdly and finally, the subject would need to be “a competent speaker of both languages” (ibid.). Specifically, the subject must be ‘competent’ with respect to the word that changes from socio-linguistic community to socio-linguistic community.

Brown then argues that these three criteria “are not normally jointly met” (ibid., p.141). She contends that if there were cases in which all three criteria were met, then the differences between the socio-linguistic communities involved would breakdown. They’d become one. Hence it would cease to be the case that the subjects who would be switching between the socio-linguistic communities would be switching between *different* socio-linguistic communities. Suppose there is a term ‘ ϕ ’ that has different meanings within two linguistic communities. Suppose that it is common for subjects to be switched between the two linguistic communities. Suppose also that it is common for those subjects to be ignorant of there being a difference in the meaning of ‘ ϕ ’, and yet otherwise be considered competent with respect to the use of ‘ ϕ ’. Brown contends that if that were to happen, then the two linguistic communities would eventually “settle on a single meaning” (ibid.) for ‘ ϕ ’.

Even if this is true, it is unclear why it would entail that Brown’s three criteria “are not normally jointly met”. It may be that because of persons such as Biff switching between Britain and America it will *eventually* happen that ‘chicory’ will come to mean the same in America as it does in Britain. However, that hasn’t happened *yet*. Thus, it may still be that Biff and persons similar to Biff – who switch between Britain and America, and who are ignorant of the difference in the meaning of ‘chicory’ – could *presently* meet Brown’s three criteria – even though they may cease to do so in the future. And thus, it may still be that the content of Biff’s and others’ ‘chicory’ thoughts do indeed change in the way that Ludlow supposes, even though that may cease to be the case at some future date.

3.22: *The Reliabilist Response*

What we can call the *reliabilist response* to the discrimination argument (Falvey and Owens, 1994; McLaughlin and Tye, 1998b)²⁶ proposes a reinterpretation of Goldman’s ‘fake barn’. It suggests that

²⁶ There are differences in the details of Falvey and Owens’s response and McLaughlin and Tye’s response. But those differences are ultimately superficial.

Falvey and Owens characterise the discrimination argument as resting on the following assumption:

- (FO1) If (i) q is a relevant alternative to p , and (ii) S ’s belief that p is based on evidence that is compatible with it being the case that q , then S does not know that p .

Slow-Switching Sally's belief that she is thinking that water is wet is based on introspective evidence that would seem the same to Sally were she were on Twin Earth thinking that twater is wet. It would have the same *phenomenological character*. Thus, the evidence available to Sally would be the same if she were on Twin Earth thinking that twater is wet. That is, it would if the 'sameness of evidence' is to be identified with sameness of phenomenological character. Thus, the evidence available will be compatible with it being the case that she is thinking that water is wet. Thus, Sally and her belief that she is thinking that water is wet satisfies the antecedent of (FO1). And thus, if (FO1) is correct, then Sally cannot know that she is thinking that water is wet.

Falvey and Owens characterise the impetus for (FO1) as coming from 'fake barn'. They suggest, however, that (FO1) might be the wrong conclusion to draw from 'fake barn'. They suggest an alternative conclusion. They suggest that rather than it being a tacit commitment to (FO1) that prejudices us against ascribing Russ knowledge, it is instead a commitment to:

(FO2) If (i) q is a relevant alternative to p and (ii) S's justification for their belief that p is such that, if q were true, then S would still believe that p, then S does not know that p.

If Russ – who is driving through 'fake barn' country – sees a barn, and hence forms the true belief that what he is looking at is a barn, then (i) Russ looking at a fake barn constitutes a relevant alternative, and (ii) Russ would still believe that what he is looking at is a barn if what he was looking at were a fake. *That* is why Russ doesn't know that what he is looking at is a barn.

Although (FO2) disqualifies Russ's belief that he is looking at a barn from counting as knowledge, it does not disqualify Slow-Switching Sally's belief that she is thinking that water is wet from counting as knowledge. The case of Slow-Switching Sally does not meet both of the criteria stated in the antecedent of (FO2). It meets the first, but not the second.

McLaughlin and Tye suggest that a rejection of (FO1) in favour of (FO2) may not be the correct conclusion to draw from 'fake barn'. Consider the case of *Dennis the Diviner*. There is a wide-spread rumour that Dennis the Diviner is all too aware of, which is that an evil wizard has been going around turning peoples' brains into sawdust. Dennis worried that he's a victim removes the entrails of a chicken. By sheer coincidence, the reading from the chicken entrails suggests that Dennis has nothing to worry about – that he does have a brain after all. Based on that reading Dennis forms the true belief that he has a brain. Now, it is highly unlikely that Dennis has a head full of sawdust. However, given the nature of Dennis's enquiry, it would still be a relevant alternative that Dennis's head is full of sawdust. (In a similar way, my being a brain in a vat would be a relevant alternative with respect to the question of whether I know that I am not a brain in a vat.) If Dennis's brain were made of sawdust, then the chicken entrails would still give Dennis the reading that he has a brain. Thus, Dennis's evidence that he has a brain would be compatible with him *not* having a brain. Thus (FO1) applied to Dennis gives us the correct result that Dennis doesn't know that he has a brain. Dennis and his belief that he has a brain does not however satisfy the antecedent of (FO2). If Oscar's brain were full of sawdust, then he could not form any belief at all. Let alone the false belief that he has a brain and his head isn't full of sawdust. Thus (FO2) is "silent" where (FO1) gives us the correct result. McLaughlin and Tye contend that that gives us some grounds for not jettisoning (FO1) in favour of (FO2). (McLaughlin and Tye, 1998b, p.356, fn.15).

McLaughlin and Tye (*ibid.*) characterise the discrimination argument as consisting of an incompatible quadruple of assumptions. The first of which is the principle that Falvey and Owens reject:

(FO1) If (i) q is a relevant alternative to p, and (ii) S's belief that p is based on evidence that is compatible with it being the case that q, then S does not know that p.

Secondly there is:

The Introspective Evidence Thesis. Introspective [i.e. a priori] knowledge of what we are occurrently thinking is based on evidence that we can introspect.

(*ibid.*, p.358)

Thirdly, there is the assumption that we are able to know a priori what the contents of our own thoughts are – which McLaughlin and Tye label “privileged access”. Fourth and finally, we have:

The Alternative Thought Thesis. For at least one type of occurrent thought that P, there is some possible circumstance in which one is occurrently thinking that P, one’s faculty of introspection is functioning properly, and the evidence that one can introspect is compatible with it being the case that one is thinking some relevant alternative thought that Q.

(*ibid.*, p.358)

McLaughlin and Tye concede that these four assumptions do indeed form an incompatible quadruple. They suggest that our way out of that incompatible quadruple is to deny the *alternative thought thesis*. They argue that externalists are free to do so, and are free to do so without rejecting the other three assumptions – so without rejecting (FO1). McLaughlin and Tye maintain that externalism would only imply the alternative thought thesis if the following were true:

The Underdetermination Thesis. For at least one type of occurrent thought that P, whether one is occurrently thinking that P (at t) fails to supervene on the evidence introspectively available to one (at t).

(*ibid.*, p.360)

McLaughlin and Tye contend that externalists are free to reject this *underdetermination thesis*. It is not entailed by externalism. Hence externalists are free to reject the *alternative thought thesis*.

McLaughlin and Tye contend that externalists can reject the *underdetermination thesis* if they understand the ‘evidence’ that per the *introspective evidence thesis* should ground our knowledge of our thought contents, in terms of its *causal powers* as opposed to its phenomenological character. Suppose that slow-switching Sally is (on Earth) thinking that water is wet. There should be some sort of mental item(s) that Sally can introspect, and on which Sally can take as ‘evidence’ that she is thinking that water is wet. That mental item should have a certain phenomenological character. If counterfactually Sally were on Twin Earth thinking that twater is wet then there’d be a mental item that corresponds to that thought. That mental item would have the very same *phenomenological character* as the mental item that corresponds to Sally’s actual thought that water is wet. However, the mental item would have different *causal powers*. It would *cause* Sally to believe that she is thinking that twater is wet, as opposed to causing her to think that water is wet.

We have here two criteria for the sameness of evidence. Under one criterion the ‘evidence’ introspectively available to Sally would be the same if she were thinking that twater is wet. Thus, Sally’s evidence that she is thinking that water is wet will be compatible with her thinking that twater is wet. This would validate McLaughlin and Tye’s *underdetermination thesis*. That in turn would imply the *alternative thoughts thesis*, and we would be faced with the incompatibility quadruple.

the conclusion that is standardly drawn from ‘fake barn’ may in fact be the wrong one. The reason why Russ does not know that he is looking at a barn is *not* because he is unable to discriminate between his looking at a barn and a relevant alternative in which he is looking at a fake barn. Russ does not know that he is looking at a barn because if those relevant alternatives in which Russ is looking at a fake barn were the case, then the method by which Russ attains his actual belief would cause Russ to form the *false* belief that he is looking at a barn.

If this is the *real* reason why Russ does not know that he is looking at a barn, then the analogy between Russ and Slow-Switching Sally, which is what drives the discrimination argument, collapses. Slow-Switching Sally is on Earth and thinking that water is wet. She forms the a priori belief that she is thinking that water is wet. A relevant alternative may be that Sally is on Twin Earth and thinking that twater is wet. However, if that relevant alternative was the case, then Sally would not form the false belief that she is thinking that water is wet.

That is, she wouldn’t if the redeployment thesis is true (see Chapter 2). The redeployment thesis states that the (presumptively) a priori second-order beliefs that a subject forms with respect to what she is thinking will have content that is reflective of the first-order thoughts of which they are about, in such a way that guarantees their truth. When Slow-Switching Sally is on Earth she would form the belief that she is thinking that water is wet, which will be true because on Earth she will be thinking that water is wet. Whereas when Sally is on Twin Earth, she would not form the belief that she is thinking that water is wet. She would instead form the belief that she is thinking that twater is wet, which will be true because it will be the case that when she is on Twin Earth she will be thinking that twater is wet.

Similarly, when Biff and Griff form (presumptively a priori) beliefs with respect to their respective ‘chicory’ and ‘pragmatist’ thoughts, there will be relevant alternatives in which they are not thinking what they are actually thinking, and in which their respective beliefs are false. But, according to the redeployment thesis, within such relevant alternatives, Biff and Griff would not form what

However, under the other criterion, the evidence introspectively available to Sally *wouldn’t* be the same if she were (on Twin Earth) thinking that twater is wet – as it is individuated in terms of causal powers – and thus wouldn’t be compatible with her thinking that twater is wet. If this latter view of ‘evidence’ is correct, then it undermines the underdetermination thesis. Thus, we could reject the underdetermination thesis. Hence, we could reject the alternative thought thesis – as opposed to (FO1).

Falvey and Owen and Mclaughlin and Tye give accounts of why Slow-Switching Sally’s belief that she is thinking that water is wet counts as knowledge that, although differ in some of their details, both ultimately and essentially argue that Sally’s belief counts as knowledge even though she is not able to discriminate between her being on Earth and thinking that water is wet, from her being on Twin Earth and thinking that twater is wet.

Both accounts invoke phenomenological evidence that Sally supposedly uses to self-ascribe her thought that water is wet. This shared detail is problematic. Thoughts (at least by assumption – see Introduction, §0.3) produce no phenomenology. It is, therefore, difficult to see what could constitute such ‘phenomenological evidence’.

would be false beliefs about what they would be thinking. They'd form, instead, true beliefs that have content that is reflective of what they would counterfactually be thinking.

Now, it may be thought that if this re-interpretation of 'fake barn' is accepted then that should allow us to conclude that *any* instance in which a subject reflects upon her thoughts and forms, via an a priori means, a belief that she is thinking that <p> is an instance in which that subject knows that she is thinking that <p>. Such beliefs would be such that a subject could only form them if it is true. A subject could only form the belief that she is thinking that <p> if she is thinking that <p>. When a subject does form the belief that she is thinking that <p>, there may be relevant alternatives in which things would seem to the subject exactly as they actually are, but in which the subject is not thinking that <p>, as there is in the case of Slow-Switching Sally, and as there are in the cases of Biff and Griff. But, within those relevant alternatives the subject would not form the false belief that she is thinking that <p>. Brown, however, does present a putative case in which a subject does form the sort of belief being described, but in which putatively that belief does not count as knowledge.

Brown presents the case of *Wasp-Phobic Wendy* (Brown, 2004, p.115). Wasp-Phobic Wendy has a phobia of wasps. She is prone to imagine that she is hearing a wasp when there isn't one there. She hears a wasp. That is, she actually hears an actual wasp. She doesn't just imagine that she is hearing a wasp. Wendy thinks a *perceptually demonstrative singular thought* that has as its object the wasp that she hears, and that she'd express '*that wasp is near*'. Wendy forms the a priori belief that she is thinking that thought.

Brown characterises three broad forms of content externalism – see below (§3.23-7). One of which is what she calls "singular anti-individualism"²⁷, which is the thesis that "a subject's thought contents are individuated partly by the particular objects that are in her environment" (ibid., p.13-4). More specifically, singular anti-individualism holds that *singular thoughts* have contents that are in part individuated by the object of which they are about. Changing the object of a singular thought changes its content. Kripkean externalism is a form of singular anti-individualism.²⁸ Burgean externalism is not (see Chapter 1, §1.4). A singular thought, according to the singular anti-individualist, is a thought that owes its very content to, and thus requires the existence of, the object of which it is about. A *perceptual demonstrative singular thought* is a singular thought that has as its object something that is presently within the thinker's perceptual field, and thus requires that object to exist. Wendy's perceptual demonstrative singular thought '*that wasp is near*' requires "*that wasp*" to exist

²⁷ Brown prefers the term 'anti-individualism' to 'content externalism'. The two terms can be and are used interchangeably.

²⁸ The philosophers that Brown deems to be singular anti-individualists along with Kripke are Perry (1979), Evans (1982), Peacocke (1983), McDowell (1986), Salmon (1986), Soames (1987), and Kaplan (1989).

Consider, a counterfactual scenario in which things seems to Wendy exactly has they actually are, but in which there is no wasp. There is only a wasp in Wendy’s imagination. Due to Wendy being prone to hearing imaginary wasps, such a counterfactual scenario in which she imagines hearing a wasp should be a relevant alternative. Within the counterfactual scenario, Wendy cannot be thinking a perceptual demonstrative singular thought that has a wasp as its object, as there is no wasp.

There are several options with respect to what we could suppose is going on in this counterfactual – in which ‘*that wasp*’ does not exist. One option is to suppose that Wendy’s thought could be *descriptive*. Wendy could be thinking the thought that: *the F is near*, where ‘the F’ would be some definite description that Wendy believes her imaginary wasp to uniquely satisfy. A second option is that Wendy is not thinking a thought at all. She is suffering from what is known as an *illusion of thought* (Evans, 1982; McDowell, 1986). Although it would seem to Wendy that she was thinking a thought, she in fact wouldn’t be. A third option is to suppose that Wendy is thinking a singular thought. That singular thought, however, does not have as its object an actual flesh-and-blood wasp. It has as its object an illusory wasp, which only Wendy can hear – ‘*that wasp*’ does in fact exist in some sense, but not in the sense that Wendy believes it to exist.

The first option is problematic. It requires assent to a sort of *second order externalism*, whereby it is certain external factors that determine the logical form of a thought. This, as we’ll be seeing in Chapter 6 (§6.2) of this thesis, is problematic²⁹. For reasons that will be intimated elsewhere in the thesis (e.g. Chapter 4, §4.211), I believe that the third option may be viable however. In any case, let us suppose *anyway* that only the second option is open to us. Wendy, let us suppose, would suffer an illusion of thought.

If Wendy would be suffering an illusion of thought, then she would not be able to form the belief that she actually does form. Similar to how she would be suffering an illusion of thought, she would also be suffering an *illusion of belief*. Thus, the method by which Wendy forms her actual belief would not cause her, within the counterfactual scenario in which there is no wasp, to form what would be the false belief that she is thinking the thought ‘*that wasp is near*’. This, *prima facie*, should mean that her actual belief can still count as knowledge, as our relevant alternative is not one in which Wendy forms a belief that is false.

Suppose now that when Wendy formed her belief, she also, based on her belief, formed the belief that she is thinking a bona-fide waspy thought – i.e. she is not suffering an illusion of thought. This belief in the actual case would be true. In our counterfactual scenario in which there is no wasp,

²⁹ To be clear, it is me who is saying that it is problematic, not Brown. Brown, in fact distinguishes between two forms of “singular anti-individualism”. The first of which – what Brown calls “the *descriptive version*” (Brown, 2004, p.16) – holds that in the counterfactual scenario in which there is no wasp, Wendy would think a descriptive thought. According to the second form – what Brown calls “the *illusion version*” (ibid.) – Wendy would suffer an illusion of thought.

Wendy would still form this belief, and it would be false, as Wendy would be suffering an illusion of thought.

Brown argues that if it is maintained that *global reliability* is required for knowledge, and not just *local reliability*, then because Wendy would form the false belief that she is thinking a bona-fide waspy thought in our counterfactual scenario, Wendy's belief that she is thinking the thought 'that wasp is near' cannot count as knowledge.

Suppose a subject S forms a true belief that $\langle p \rangle$ is the case via a method of belief formation M. That method of belief formation M will be *locally reliable* just in case there are no relevant alternatives in which $\langle p \rangle$ is not the case, but M would still cause S to form the false belief that $\langle p \rangle$ is the case. The method of belief formation M will be *globally reliable* just in case M is and would be locally reliable with respect to a range of propositions that are similar to $\langle p \rangle$. M is globally reliable just in case, for any true proposition $\langle q \rangle$ that M might cause S to believe, and that is similar to $\langle p \rangle$, there will be no relevant alternatives in which $\langle q \rangle$ is not the case, but M would still cause S to form the false belief that $\langle q \rangle$ is the case.

Brown argues that the proposition \langle Wendy is thinking a bona-fide waspy thought \rangle is the same sort of proposition that forms the content of Wendy's other belief, which is \langle Wendy is thinking the thought 'that wasp is near' \rangle , because both propositions concern what Wendy is currently thinking. Thus, the method by which Wendy has acquired her belief that she is thinking the thought 'that wasp is near' won't be globally reliable, as it would lead her to believe falsely that she is thinking a bona-fide waspy thought in the counterfactual scenario in which there is no wasp. And thus, if we assume that global – and not just local – reliability is required for knowledge, then we must concede that Wendy's guaranteed-to-be-true belief that she is thinking the thought 'that wasp is near' cannot count as knowledge.

For Brown's argument to go through it must be the case that knowledge requires global – and not just local – reliability.

Even if that is assumed, Brown's argument still faces a couple of further inter-related problems.

Firstly, there is what Brown herself identifies as “the generality problem” (2004, p.132). A method of belief formation M will be globally reliable just in case M is and would be locally reliable with respect to a range of propositions that are similar to $\langle p \rangle$. What Brown calls “the generality problem” is the problem of what determines which propositions are similar to $\langle p \rangle$. If a subject S has come to believe that $\langle p \rangle$ is the case via a method M, then M won't be globally reliable if there is a proposition $\langle q \rangle$ that is *sufficiently similar* to $\langle p \rangle$, and for which there will be a relevant alternative in which $\langle q \rangle$ is not the case, but M would still cause S to believe that $\langle q \rangle$ is the case. The generality problem is the problem of what would be required of such a global reliability denying proposition $\langle q \rangle$

to be sufficiently similar to <p>. It is the problem of what *criteria* such a proposition <q> must meet. Brown's argument requires a convincing solution to this problem.

Secondly, assuming this generality problem can be solved, and appropriate criteria can be found for what it takes for a proposition <q> to be knowledge denying, then it must also be the case that the proposition that Wendy is thinking a thought meets those criteria. It is not immediately clear that the proposition that Wendy is thinking a thought about a wasp is indeed the same sort of proposition that constitutes the content of Wendy's other belief. The latter proposition is a proposition that ascribes to Wendy a thought with a *specific* content, whereas the former is the proposition that Wendy is thinking a bona fide thought. It could be maintained that the method by which Wendy acquires her belief that she is thinking the thought 'that wasp is near' is globally reliable if it is locally reliable with respect to any proposition that ascribes to Wendy a thought with a specific content, which it will be. If Wendy's method were to lead Wendy to form beliefs with respect to *any* such proposition, then they'd be sort of guaranteed-to-be-true beliefs that are envisioned by the redeployment thesis. Thus, the method by which Wendy acquires the belief is globally reliable, even though, it may not be locally reliable with respect to all of the beliefs that the method produces, and in particular with respect to Wendy's belief that she is thinking a bona-fide waspy thought.

In sum, *if* (i) in the counterfactual scenario in which there is no wasp Wendy would be suffering an illusion of thought, and *if* (ii) global, and not just local, reliability is required for knowledge, and *if* (iii) the generality problem can be solved, i.e. if criteria can be successfully and correctly identified for what it takes for a proposition <q> to be knowledge denying, and *if* (iv) it can be shown that the proposition that Wendy is thinking a bona fide waspy thought constitutes a knowledge-denying proposition, i.e. if it can be shown that it meets those criteria, *then* Brown may have indeed identified a case in which a subject forms a guaranteed-to-be-true belief about what she is presently thinking – of the sort envisioned by the redeployment thesis – but in which that belief does not count as knowledge.

3.23: Brown's Response I: Brown's Overall Strategy

In her (2004) book *Anti-Individualism and Knowledge*, Brown mounts a comprehensive and anticipatory offensive against the discrimination argument. The approach she takes assumes that 'fake barn' should be interpreted as it standardly is. That is, Brown's approach rejects the reinterpretation discussed in the previous subsection. It is "a response to the discrimination argument that takes seriously the incompatibilist's claim that knowledge requires discriminative abilities" (Brown, 2004, p.75).

Brown identifies three forms of externalism. For each of those three forms, she argues cases that may be touted by an incompatibilist, in which there are knowledge-denying relevant alternatives,

either involve counterfactual scenarios that could never constitute relevant alternatives in the case of normal subjects, and/or there are *deeper* reasons why they could not occur.

3.24: Brown's Response II: Natural Kind Anti-Individualism and Recognitional Capacities

The case of Slow-Switching Sally is too outlandish to be taken seriously. Brown suggests, however, that there is also a *deeper* reason why cases such as that of Slow-Switching Sally cannot occur, and why it is that there cannot be cases in which a subject forms a true belief that they are thinking a kind-directed thought, but in which there are knowledge-denying relevant alternatives.

Brown defines "*natural kind anti-individualism*" as the thesis that "a subject's thought contents are individuated partly by the natural kinds in her environment" (ibid., p.9). Essentially, natural kind anti-individualism is singular anti-individualism (see above, §3.22), but in which objects are replaced with natural kinds. Singular anti-individualism holds that we can think singular thoughts that owe their very content to the objects of which they are about. Natural kind anti-individualism holds that we can think what we can call *kind-directed thoughts* that owe their very content to the *natural kind* of which they are about. Oscar's thought that 'water is wet' will be such a kind-directed thought. That is why it will be different in content to Twin Oscar's equivalent thought. Oscar's thought is directed at water. Whereas, Twin Oscar's thought is directed at twater.

Brown (2004, pp.142-5)³⁰ proposes that in order for a subject to think the sort of kind-directed thought envisioned by natural kind anti-individualism, that subject must have a *recognitional capacity* with respect to the kind that the thought is directed at. The subject must have encountered the kind in her environment, and she must be able to distinguish it from the other kinds that she finds in her environment. On the occasions in which she encounters the kind, the subject must *know* that what she's encountered is the *same* kind that she's encountered on previous occasions in which she's encountered the kind. This, Brown contends, should ensure that when or if the subject forms a belief with respect to a kind-directed thought that there cannot be any relevant alternatives. In short, in order for a subject to think a kind-directed thought she must meet a certain condition – she must have a recognitional capacity with respect to the relevant kind. The fact that the subject meets that requirement ensures that if she were to form a second-order belief that she is thinking a particular kind-directed, then that belief would constitute knowledge.

Consider René. René is thinking the thought that he'd express 'water is wet' – where 'water' would refer to water directly. René forms the belief that he is thinking that thought. That belief could not count as knowledge if there were a relevant alternative in which it is not true that René is thinking

³⁰ See also Brown, 1998.

that water is wet, and in which things would seem to René exactly as they actually do. But, Brown contends, there cannot be any such relevant alternatives. Such a relevant alternative would be one in which René has had encounters with a substance that is superficially identical with water, that as far as René is concerned is no different to water, but is not water. If such a possibility was relevant, then it could not be the case that René is thinking what he is thinking. For René to be thinking what he is thinking he must have a recognitional capacity for water. That is, when he encounters samples of water he must know that he is encountering the same substance that he has encountered on previous occasions. For that to be the case there can be no danger of René encountering a substance that is superficially identical to water, which he'd mistake for water.

Implicit in Brown's proposal is the contention that what seems to Slow-Switching Sally to be the thought that water is wet, is not in fact a thought at all. When it seems to Sally that she is thinking such a thought, she is in fact suffering an *illusion of thought*, similar to how Wasp-Phobic Wendy would suffer an illusion of thought in the counterfactual in which there is no wasp (See above, §3.22). The reasons why Sally and Wasp-Phobic Wendy would suffer an illusion of thought diverge. In the case of Wasp-Phobic Wendy it is because there is no wasp. Sally suffers an illusion of thought not because there is not water, but because she does not have a recognitional capacity for water. When Sally is on Earth and encounters a sample of water, a relevant alternative would be that she is on Twin Earth encountering a sample of twater. Thus, Sally cannot know that she is encountering a sample of the same stuff that she encountered on previous occasions. And thus, she does not have a recognitional capacity for water. And thus, Sally is *not* able to think the sort of water directed thoughts that we are supposing she can think.

This is tantamount to saying that when Sally says 'water is wet', her utterance is meaningless, which seems bizarre.

There are in fact a second and a third way of interpreting the case of Slow-Switching Sally that respect Brown's invocation of recognitional categories – i.e. that grant that Sally must have a recognitional capacity for water if she is to think the sort of kind-directed thoughts that we were supposing that she can think.

The second way supposes that Sally's thoughts would be *descriptive*. Sally is *not* able to think the sort of water directed thoughts that we are supposing she can think. Since she is the victim of a slow-switch case, she does not have a recognitional capacity for water. *However*, that does not mean that Sally suffers illusions of thought. If Sally weren't a victim of a slow-switch case, then she'd be thinking a kind-directed thought. But, by virtue of the fact that Sally is a victim of a slow-switch case, the thought that Sally would express by 'water is wet' would be something along the lines of: *the clear, odourless liquid that I (Sally) drank a lot of yesterday is wet*.

If this were the case, then it should allow that the beliefs that Sally forms with respect to what she is thinking could count as knowledge. When Sally is on Earth thinking that thought that she'd

express ‘water is wet’, she’d be thinking the exact same thought that she’d be thinking if counterfactually she were on Twin Earth. Thus, there will be no relevant alternatives in which Sally is not thinking what she is thinking, but in which things would seem to Sally exactly as they actually are.

This second option is problematic. It requires assent to the sort of *second order externalism* mentioned previously (§3.22), whereby external factors determine whether a term refers directly or is descriptive, which, as we’ll be seeing in Chapter 6 of this thesis, is problematic.

The third interpretation supposes that Sally *is* able to think the sort of water/twater directed thoughts that we are supposing she can. Therefore, she *does* have a recognitional capacity for water/twater. Thus, when Sally is on Earth and encounters a sample of water, her being on Twin Earth encountering a sample of twater is *somehow* not a relevant alternative. And equivalently, when Sally is on Twin Earth and encounters a sample of twater, her being on Earth and encountering a sample of water is *somehow* not a relevant alternative.

It may seem ad hoc to deny that counterfactual scenarios in which Sally is on Earth/Twin Earth encountering a sample of water/twater constitute relevant alternatives. We could maintain, however, that we do have a sort of *quasi-transcendental* argument for denying that Sally being on Earth/Twin Earth and thinking that water/twater is wet constitutes a relevant alternative. Sally *is* thinking the sort of kind directed thought that we were supposing she can think. It cannot be the case that Sally is suffering an illusion of thought – because it would be odd to suppose that she is. Nor can it be the case that she is thinking a descriptive thought – for the reasons we’ll be considering in Chapter 6. If Sally is thinking the sort of kind-directed thought that we were supposing she can think, then she must have a recognitional capacity for water. Sally *is* thinking the sort of kind-directed thought that we were supposing she can think. She therefore *must* have a recognitional capacity for water. Thus, when Sally encounters a sample of water she knows it is the same substance that she has encountered on previous occasions. *Ipsa facto*, when Sally encounters a sample of water and judges that she is encountering a substance that she has encountered on previous occasions, her being on Twin Earth and encountering a sample of twater cannot constitute a relevant alternative.

Sally having a recognitional capacity for water, in itself, does not *necessarily* entail that Sally’s belief that she is thinking that water is wet will constitute knowledge. A *further* thing is required. It must be the case that there is a *transference of irrelevance* from the question of whether Sally has a recognitional capacity for water to the question of whether Sally’s belief that she is thinking that water/twater is wet counts as knowledge. It must be that – i.e. it must be argued that – because the counterfactual scenarios that involve Sally being on Twin Earth that are ruled out from constituting relevant alternatives when it comes to the question of whether Sally has a recognitional capacity for water, those very same counterfactual scenarios are ruled out as constituting relevant

alternatives when it comes to the separate question of whether her belief that she is thinking that water is wet counts as knowledge.

If there is no such transference of irrelevance, then Sally's belief the water/twater is wet won't be knowledge constituting, which means that Brown's proposal is void.

We can, of course, follow the Warfield line and discount the case of Slow-Switching Sally on the grounds that it is too outlandish to be taken seriously. When you or I think that water is wet, we may not be able to discriminate the actual from a counterfactual scenario in which we have been transported to Twin Earth and are thinking that twater is wet. But, for us at least, such a far-fetched counterfactual scenario cannot constitute a relevant alternative. Thus, it can still be claimed that both you and I know what we are thinking when we think the thought: *water is wet*. If we go down this route. i.e. if we are to dismiss the case of Slow-Switching Sally on the grounds that it is too outlandish to be taken seriously, then that begs the question of why, within the present context, we need to invoke Brown's recognitional capacity requirement at all.³¹

3.25: Brown's Response III: Perceptual Demonstrative Singular Thought: The Problem Case

The singular anti-individualist holds that we are able to think thoughts that have contents that are inextricably tied to the objects of which the thoughts are about. Brown proposes that such thoughts fall into either of two categories. There are perceptual demonstrative singular thoughts of the sort that Wasp-Phobic Wendy entertains towards a nearby wasp (see above, §3.22). There are also *recognition-based singular thoughts* (see next subsection).

Brown concedes something that she finds to be "disappointing" (ibid., p.152). There *will* be cases, everyday cases – which don't involve subjects being transported an unthinkable number of light years – in which a subject thinks a *perceptual demonstrative singular thought* and forms a guaranteed-to-be-true belief with respect to that perceptual demonstrative singular thought, but in which that belief won't count as knowledge.

In Sandra's neighbourhood there live identical twins Anja and Tanja. One day Sandra is sitting in the park. She sees a child, who happens to be Anya, playing in the children's play area. Sandra thinks a *perceptual demonstrative singular* thought that she'd express, using a demonstrative and possibly pointing to Anya, as '*that* child is happy'. Sandra forms a true belief with respect to that thought. It could very easily be the case that it is Tanja that Sandra is seeing frolicking in the children's play area. If that were the case, then Sandra's thought would be different in content to what it actually is, as it would be Tanja, and not Anya, who is the object of the thought. Such a

³¹ Brown does make a separate case – which has nothing to do with what is presently being discussed – for kind-directed thought requiring recognitional capacities in Brown 1998.

counterfactual scenario should constitute a relevant alternative. It will also be the case that, within such a counterfactual scenario, things would seem to Sandra exactly as they actually are. Thus, Sandra's belief about what she is thinking cannot count as knowledge (Brown, 2004, pp.86-88).

Of course, if we go down the reliabilist response route of denying that discriminative abilities are required for knowledge, we could say that Sandra's belief does count as knowledge. Within the counterfactual scenario in which it is Tanya not Anja that Sandra is seeing, the belief that Sandra would form would have a different content and would be true. But that is not the route that Brown is trying to defend.

3.26: Brown's Response IV: Recognition-Based Singular Thought

Recognition-based singular thoughts are singular thoughts that have as their object something that we perceived in the past, but that is no longer within our perceptual field. We are able to think recognition-based singular thoughts as long as a certain condition holds. A subject is able to think a recognition-based singular thought about some object *o*, which they've perceived but no longer perceive, on the condition that if, at the moment of the thought, the subject were to re-encounter *o* then they'd know that what they are encountering is the same object that they had perceived previously.³²

Brown assumes that all thinkable singular thoughts are either perceptual demonstrative singular thoughts or recognition-based singular thoughts. In order for us to think a singular thought about a particular object, that object must either currently be in our perceptual field, or the object must have been within our perceptual field in the past and if we were to re-encounter that object we'd know it is the same object.

Brown finds it to be "disappointing" that when a subject thinks a *perceptual demonstrative singular thought* and forms a belief with respect to that perceptual demonstrative singular thought, there may be relevant alternatives in which the subject thinks something other to what she is actually thinking, but in which thing seems to the subject exactly as they actually are. Brown argues however that the same won't be true with respect to *recognition-based singular thought*. If a subject thinks a recognition-based singular thought and forms a belief with respect to that recognition-based singular thought, then – Brown argues – there *won't* be any relevant alternatives in which the subject thinks something other to what she is actually thinking, but in which thing seems to the subject exactly as they actually are.

We have seen (§3.24) how Brown identifies a condition for natural kind directed thoughts, and argues that if a subject is indeed thinking such a thought – and hence meets the condition that

³² Brown draws heavily from Evans's account of recognition-based singular thought (Evans, 1982; Brown, 2004, pp.77-83).

Brown identifies – then that precludes there being relevant alternatives in which things seems to the subject exactly as they actually do, but in which the subject is thinking something other to what she is actually thinking. Brown argues for a similar thing with respect to recognition-based singular thoughts. If a subject is thinking a recognition-based singular thought, then she will meet the condition that Brown identifies as necessary for recognition-based singular thought and that, Brown argues, will entail that, if the subject forms a belief with respect to that thought, there will be no danger of knowledge-disqualifying relevant alternatives.

Nick is looking at a cat, Tabby, when whoosh all of a sudden Tabby is nowhere to be seen. Whilst Tabby was still in Nick's perceptual field, Nick *was* able to think perceptually demonstrative singular thoughts that have Tabby as their object. But, now that Tabby is gone, Nick has ceased to be able to think such perceptually demonstrative singular thoughts. It may, however, still be possible for Nick to think a Tabby-centred singular thought of another sort. Nick will be able to think a Tabby-centred *recognition-based* singular thought on the condition that if he were to re-encounter Tabby he would *know* that he was re-encountering the same cat that was previously within his perceptual field. In order, for Nick to think a Tabby-centred recognition-based singular thought, it must be the case that, if Nick were to re-encounter Tabby, there would be no relevant alternative in which the cat that Nick encounters looks exactly like Tabby, but is in fact another cat. That could only be the case – Brown contends – if there is no nearby doppelganger cat that could have wandered into a certain area that can be called the *area of search*³³. The area of search is defined by how far Tabby could have moved since Tabby was in Nick's perceptual field – i.e. is defined by how fast Tabby can move. If there were a doppelganger cat that could have moved into the area of search, then Nick would not be able to think a Tabby-centred recognition-based singular thought.

Brown's contention is that if Nick were to form a second-order belief about his Tabby-centred recognition-based thought, then there would be no relevant alternative in which things would seem to Nick exactly as they actually would, but in which Nick is thinking a different thought to what he would actually be thinking. That is, there will be no such relevant alternatives so long as there is no

³³ By 'area of search' I mean what Brown calls the 'actual area of search' – i.e. what Brown proposes should be regarded as the 'area of search' – as opposed to the 'estimated area of search' – which is what Evans regarded as the 'area of search'. Under Evans's original account of recognition-based singular thought (Evans, 1982) the 'area of search' comprised the 'estimated area of search', which, in our example, would be defined by Nick's *estimation* of how fast Tabby can move. Under Evans's account, Nick would only be able to think a Tabby-centred recognition-based singular thought if there is no Tabby-look-alike cat within this 'estimated area of search'. Brown argues that this is wrong (Brown, 2004, pp.96-106). Nick would only be able to think a recognition based singular thought if, were he to re-encounter Tabby, he'd *know* he was reencountering Tabby. That means – Brown argues – that Nick would only be able to think a Tabby-centred recognition-based singular thought if there is no Tabby-look-alike cat in the 'actual area of search', which is defined by how fast Tabby can *actually* move.

doppelganger cat that could have moved within the area of search since Tabby was last in Nick's perceptual field – i.e. so long as Nick is able to think a recognition-based singular thought.

Brown, however, fails to make it clear why this should be the case. It is not clear why there being no duplicate within the area of search entails – as Brown contends that it does – that there are no knowledge denying relevant alternatives in which Nick would be thinking a recognition-based singular thought that is centred on a Tabby-look-alike cat.

Consider Nicola. Nicola encounters a cat, Tibby. Nicola is then momentarily distracted and whoosh Tibby is nowhere to be seen. Nearby there is another cat, Kitty, who is an exact doppelganger for Tibby. Although Kitty is nearby, Kitty is of a sufficient distance away from the boundary of the area of search so as to render it impossible – given how fast Kitty and Tibby can travel – for Kitty to have moved within that area. That is, Kitty could not have gotten to somewhere where Tibby could have gotten to since Tibby was last within Nicola's perceptual field. There are no other doppelgangers for Tibby nearby. Crucially, there is not a doppelganger for Tibby that could have gotten within the area of search. Thus, Nicola should be able to think a recognition-based thought with respect to Tibby. She should consequently be able to form a guaranteed-to-be-true belief that she is thinking that thought.

Consider, now, a counterfactual scenario in which it is Kitty, and not Tibby that Nicola encountered, and in which Tibby is exactly where Kitty actually is. In such a counterfactual scenario, Nicola should be able to think a recognition-based thought. That recognition-based thought will be different to what Nicola will actually be thinking. It will be a *Kitty-centred* recognition-based singular thought, as opposed to a *Tibby-centred* recognition-based singular thought. Due to the proximity of Kitty, such a counterfactual scenario would arguably constitute a relevant alternative with respect to the question of whether Nicola's second-order belief counts as knowledge. Furthermore, it would be a relevant alternative in which things would seem to Nicola exactly as they actually are. Thus, we possibly have a relevant alternative that disqualifies Nicola's belief about what she is thinking from counting as knowledge.

It is intuitively plausible that counterfactual scenarios in which it is Kitty that Nicola re-encounters should *not* be considered relevant alternatives with respect to the question of whether Nicola would know that she is encountering the same cat as she encountered previously, were she to re-encounter Tibby. It is, however, *equally* intuitively plausible that a counterfactual scenario in which it was Kitty and not Tibby that Nicola initially encountered, and in which it is Tibby and not Kitty that is outside of the area of search, should be considered a relevant alternative with respect to the question of whether Nicola could know what she was thinking.

There are several options here.

One option is that we try to argue – in a sort of quasi-transcendental way – that because certain counterfactual scenarios should not be considered relevant alternatives with respect to one

question of knowledge, similar counterfactual scenarios should not be considered relevant with respect to another question of knowledge. That is, we may try to argue that because scenarios that involve Nicola encountering Kitty³⁴ won't constitute relevant alternatives when it comes to the question of whether Nicola would, were she to re-encounter Tibby, know that she is re-encountering the same cat – i.e. because Nicola is able to think a Tibby-centred recognition-based singular thought – counterfactual scenarios in which it is Kitty and not Tibby that Nicola initially encounters should not be considered relevant when it comes to the question of whether Nicola knows what she is thinking. There is a *transference of irrelevance* from one of the questions of knowledge to the other.

This first option should, if successful, leave Brown's proposal intact.

However, a second and just as plausible option is that we try to go the other way. We could try to argue that because of the presence of Kitty, Nicola does *not* know what she is thinking, and because of a *transference of relevance* she cannot think a Tibby-centred recognition-based singular thought. If Nicola were able to form a Tibby-centred recognition-based singular thought Nicola initially encountering Kitty would constitute a relevant alternative when it comes to the question of whether Nicola knows what she is thinking. Thus, scenarios in which Nicola encounters a different cat – namely Kitty – should be considered relevant alternatives when it comes to the question of whether Nicola would know that she is re-encountering Tibby, were she to re-encounter Tibby. And thus, if Nicola were to re-encounter Tibby she would not know that she is re-encountering Tibby. And thus, Nicola won't be able to think a Tibby-centred recognition-based singular thought.

If this is the case, then it is possible that being able to think a recognition-based singular thought could still preclude knowledge denying relevant alternatives in the way that Brown contends that it would. Brown, however, would need to provide a new account of what is required to think a recognition-based thought that would preclude Nicola's putative Tibby-centred recognition-based singular thought from counting as a bona-fide Tibby-centred recognition-based singular thought.

There is, finally, a third option that is just as plausible as the other two. Under this third option we maintain that Nicola *is* able to think a Tibby-centred recognition-based singular thought, but the belief that that thought would give rise to *would not* constitute knowledge. Were Nicola to re-encounter Tibby, a counterfactual scenario in which it is Kitty that she encounters *should not* be considered a relevant alternative with respect to the question of whether Nicola would know that she is re-encountering Tibby. However, a counterfactual scenario in which it is Kitty that Nicola initially encounters *should* be considered a relevant alternative when it comes to the question of whether Nicola

³⁴ These will be *second order counterfactuals*, that is, counterfactuals that are within counterfactuals. We are considering counterfactuals in which Nicola reencounters Tibby, and we are considering what would be a counterfactual within those counterfactuals in which it is Kitty and not Tibby that Nicola encounters.

knows what she is thinking. There is no transference of irrelevance nor relevance between the two questions of knowledge.

If this third option were to succeed, then it would demonstrate that Brown's strategy has failed. A subject being in a position to think a recognition-based singular thought would not guarantee that there are no relevant alternatives in which the subject is thinking something other to what she is actually thinking, but in which things seem to the subject exactly as they actually are. Thus, if discriminative abilities are required for knowledge, a subject being able to think a recognition-based singular thought would not guarantee that the subject would be able to know a priori what she is thinking.

3.27: Brown's Response V: Social Anti-Individualism (a.k.a. Burgean Externalism) and Illusion Cases

As well as singular anti-individualism and natural kind anti-individualism, Brown also characterises a third form of anti-individualism(/content externalism), which she calls "social anti-individualism", and which is essentially what I've been calling Burgean externalism.

We have already seen how Brown deals with the sort of cases in which a subject forms a belief about what she is thinking in which there are knowledge-denying relevant alternatives that pop up under Burgean externalism/social anti-individualism. Cases such as that of Biff and his 'chicory' thoughts and that of Griff and his 'pragmatist' thoughts aren't so outlandish that they are not to be taken seriously. But, Brown suggests, there is a deeper reason why such cases could not occur. They'd need to involve a subject-word pairing that meets three criteria. There could not be, Brown contends, a subject-word pairing that meets those three criteria. Hence, cases such as those of Biff and Griff cannot actually exist. The problem with this, as we've seen (§3.21), is that there *could* be a subject-word pairing that meets those three criteria – even though they may do so only for a limited period of time.

There is however still one final sort of putative case in which a subject forms a belief about what she is thinking, and in which there are knowledge-denying relevant alternatives. These are cases in which a subject is somehow prone to hallucinate the existence of something that doesn't in fact exist, and in which it is supposed that because of that the subject suffers an illusion of thought.

The case of Wasp-Phobic Wendy is such a case. When Wendy thinks her wasp-directed (perceptual demonstrative) singular thought, she is unable to discriminate what is actual from a counterfactual scenario in which she is imagining a wasp, and in which she is therefore suffering an illusion of thought. For Wendy, due to her proneness to aurally hallucinate non-existent wasps, such a counterfactual scenario does constitute a relevant alternative. Thus, the second-order guaranteed-to-be-true belief that Wendy would form with respect to her wasp-directed thought cannot constitute knowledge. Brown contends, however, that the case of Wendy is exceptional. We are not normally

prone to the sort of aural hallucinations that poor Wendy is subject too. If a normal subject were to hear a wasp and think a wasp-directed singular thought, then she, like Wendy, would not be able to discriminate what is actual from a counterfactual scenario in which she is imagining a wasp and is therefore suffering an illusion of thought. However, for a normal subject, such a counterfactual scenario cannot constitute a relevant alternative.

This may at first glance seem wrong. We can all recall times in which we thought we heard something that wasn't in fact there. If, however, we are to suppose that Wendy is able to think a wasp-directed perceptual demonstrative singular thought when she hears the wasp, it can't just be the case that Wendy thinks she might be hearing a wasp. Wendy must experience the vivid phenomenology of hearing a wasp. It must also be the case that Wendy would experience that same vivid phenomenology in the counterfactual scenario in which there is no wasp. And, for such a counterfactual scenario to constitute a relevant alternative, it must be that Wendy is prone to such vivid aural illusions. Wendy, it turns out, must suffer from some form of psychosis.

This ends our discussion of Brown's response to the discrimination argument. Brown's strategy is, firstly, to concede to the advocate of the discrimination argument that discriminative abilities are required for knowledge. The strategy is to then argue that despite that under all of the three forms of content externalism – what Brown labels “natural kind anti-individualism”, “singular anti-individualism”, and “social anti-individualism” – there will still be no cases in which a subject thinks a thought, but in which there are the knowledge-denying relevant alternatives that an advocate of the discrimination argument contends that there would be. Brown herself admits that her strategy is less than successful in the case of perceptual demonstrative singular thought – i.e. with one half of singular anti-individualism. Brown's strategy also seems to be unsuccessful in the case of social anti-individualism/Burgean externalism. It is also – as we've seen – questionable whether it is successful in the case of natural kind anti-individualism and in the case of recognition-based singular thought – i.e. in the case of the other half of singular anti-individualism. There is, nevertheless, something very intriguing about Brown's idea that requirements that are baked into the thinking of a thought could preclude knowledge-denying relevant alternatives. It just that it seems that if indeed there is a kernel of a good idea here, then more work is needed to bring it to fruition.

Chapter 4: Inferring from Mind to World I: Boghossian's Reductio

4.0: Introduction to Chapter 4

Water exists. This is something I know. It is not, however, something that I could know a priori. I know that water exists because of my repeated encounters with – through seeing, touching, tasting, etc. – a substance that has certain perceptual qualities. I have a true belief that water exists. I have a justificatory basis for my belief, which consists of my repeated encounters with a substance that has certain perceptual qualities. By virtue of my belief having that justificatory basis, my belief qualifies as knowledge. Suppose that a subject could have a knowledge-granting justificatory basis for her belief that water exists that does not rest with her senses in this way or in *any* way. That is, suppose that a subject could come to know *a priori* that water exists. Such a prospect is absurd. Suppose now that there were two philosophical theses such that those theses both being true would entail that a subject could know a priori that water exists. If the two theses both being true would entail something that is so absurd, then it must be the case that at least one of those theses *must* be false. Boghossian (1998) ventures that content externalism along with the contention that a subject should be able to know a priori what she is thinking constitute two such theses. Boghossian ventures that both being true would have the absurd consequence that a subject could know a priori that water exists. Thus at least one must be false.

4.1: Boghossian's Reductio: Preliminaries and the Basic Argument

4.11: Preliminaries I: Content and 'Concept'

Boghossian's *reductio* argument requires a number of commitments with respect to the notion 'concept'.

The first of these commitments is that concepts are key contributing factors in what determines the content of a subject's thought. When a subject thinks a thought about water (say), thinks that water is wet (say), she does so by in some sense *deploying* a concept of water. It is such a cognitive deployment that is in part responsible for her thinking that water is wet – i.e. for her thinking a thought with the content that it has. Had the subject not been deploying that concept she would not have been thinking that water is wet – she wouldn't have been thinking a thought with the content that she was thinking. Had she been deploying a different concept *instead* of that concept, she would have been thinking a different thought – she would have been thinking a thought with a content that is non-identical to the content of the thought she was in actual fact thinking.

The second commitment is that concepts are in part individuated by their *extension*. Two tokens of a concept that have different extensions will always be tokens of two different concepts. The extension of a concept is all the things that *fall under* it. The extension of our concept of water encompasses all and every instance of water. Lake Geneva, rain falling from the sky, what would come out if I were to turn on a tap, ...all fall under the concept of water. They are all part of its extension. Such things are not part of the extension of the concept of *twater* – the substance that is on *Twin Earth* in lieu of water (see Chapter 1, §1.23, and below). Rather it's the equivalent things – Twin Lake Geneva of Twin Earth, the 'rain' that falls from the skies of Twin Earth, what would come out if a Twin Earthling were to turn on a tap – that constitute the extension of the concept of *twater*. When Oscar thinks that water is wet, he deploys a concept that has belonging to its extension things such as Lake Geneva, rain that falls from the sky, etc.. When Twin Oscar – Oscar's atom-for-atom duplicate – thinks that *twater* is wet he deploys a concept that has its extension not *those* things but their Twin Earthly equivalents. Hence Oscar and Twin Oscar deploy different concepts. This is why the contents of Oscar's and Twin Oscar's respective thoughts are different.

The third commitment is that a concept is either *atomic* or *non-atomic*. If a concept is not atomic then it is non-atomic. If a concept is not non-atomic then it is atomic. A non-atomic concept – or alternatively a *compound* concept – will be *breakdownable* into constituent parts that will themselves be concepts. An atomic concept, on the other hand, won't be *breakdownable* in that same way.

Suppose that Oscar was to think a thought that'd be expressed as *the tasteless, odourless liquid that flows in rivers and from taps is wet*. Oscar would be thinking a thought that will be different in content to what it would have been had he been thinking a thought that'd be expressed as *water is wet*. In the former instance Oscar would be deploying a non-atomic concept of water, one that *breaks down* into the concept of tastelessness, the concept of lacking an odour, etc.. However, in the

latter instance it is conceivable that the concept being deployed in Oscar's thought processes could not be broken down into other concepts, and thus Oscar's concept is atomic.

The fourth commitment is that a concept must have determinate *application conditions*. For any given concept there must be a *fact of the matter* with respect to what does, what doesn't, what would, and what wouldn't belong to its extension. For any given thing, actual or counterfactual, and for any given concept, it must either be the case that the thing belongs or does not belong, or would or would not belong, to the concept's extension.

For the concept water, it is water itself that provides the sole application condition. Something belongs to the extension if and only if it is water – i.e. if and only if it is made up of H₂O molecules. Something is water if and only if it is water. The liquid found in Lake Geneva falls under the concept's extension because it is water. (The liquid found in Lake Geneva is water because it is water.) The liquid methane found in the methane lakes of Titan (Saturn's largest moon) won't belong to the extension of the concept water, as that liquid will be liquid methane and not water. If I had in my hand a glass of water, which I don't, then the counterfactual liquid found within that counterfactual glass would belong to the extension, as it would be water. If I had in my hand a glass of beer, which I also don't, then the counterfactual liquid found within *that* counterfactual glass would not belong to the concept's extension, as it would be beer and not water.

The same will be true mutatis mutandis in the case of Twin Oscar's twater concept. It will be the natural kind twater, as opposed to water, that acts as the sole application condition for Twin Oscar's concept. Something will belong, or would belong, to the extension if and only if it is, or it would be, an instance of twater.

A concept's extension can in the actual world be *empty*. That is, it can be the case that nothing in actuality belongs to a particular concept's extension. But for such a concept it will still need to be a fact of the matter over what items *would* belong to the concept's extension. There is nothing in actuality that falls under the concept *unicorn*. However, something could arguably fall under the concept in certain possible worlds. If counterfactually there were an animal that was in appearance and in physiology almost identical to a horse, except it had a spiracle horn protruding from just above its nose, and if that animal had certain magical characteristics, then *perhaps* that animal would fall under the concept 'unicorn'. (This is supposing that the concept 'unicorn' is not, as some may contend, *necessarily vacuous* – see below (§4.29).)

The fifth and final commitment is that individual concepts are associated with particular words. When Oscar utters 'water is wet' he expresses a thought. As he utters the word 'water' he also *expresses a concept*. This concept will be different to the concept that Twin Oscar expresses when he utters the word 'water'.

4.12: Preliminaries II: Content and 'Concept' and Twin Earth

Boghossian's reductio also requires a *couple* of further commitments that concerns the provenance of the famous Putnamian result that Oscar and Twin Oscar will be thinking thoughts with different contents. Boghossian views Putnam's Twin Earth case as an *intuition pump* that generates the result that Oscar and Twin Oscar will be thinking thoughts with different content:

...let us remind ourselves how Putnam's thought experiment is supposed to work. Whereas Oscar, an ordinary English Speaker, lives on Earth, his molecular and functional duplicate, Toscar, lives on Twin Earth, a planet just like Earth except that the liquid that fills its lakes and oceans, while indistinguishable from Earthly water in all ordinary circumstances, is not H₂O but some other substance with a different chemical composition – call it XYZ. Going by whatever criteria are relevant to such matters, water and twin water are distinct kinds of substance, even though a chemically ignorant person wouldn't be able to tell them apart. Now widespread intuition appears to have it that, whereas Oscar's tokens of 'water' apply exclusively to H₂O, Toscar's tokens of 'water' apply exclusively to XYZ. Widespread intuition appears to have it, in other words, that Oscar's and Toscar's 'water' tokens have distinct extensions.
(Boghossian, 1997, p.164)

Boghossian claims that this "widespread intuition" is only generated if it is supposed that the concepts that are respectively expressed by 'water' on the respective lips of Oscar and Twin Oscar are atomic:

...the concepts expressed by the Earthly and Twearthly tokens of 'water' have to be thought of as atomic concepts, not compound concepts that are compositionally built up out of other concepts in well defined ways. For example, the experiment presupposes that water [i.e. the concept expressed by 'water'] can't be thought of as capable of being defined as: A tasteless, odourless liquid that flows in the rivers and faucets. For if it were a compositional concept of that sort, its extension would be determined by the extension of its ingredient parts. Hence, a conclusion to the effect that water and twater have different extensions would have to proceed differently than it does in Putnam's original experiment, by showing that one of the *ingredients* of water – the concept expressed by 'liquid', for example—has a different extension from that expressed by its Twin counterpart.
(ibid., p165)

Boghossian also claims that the intuition is only generated if it is supposed that, not only do Oscar and Twin Oscar succeed in expressing an atomic natural kind concept, Oscar and Twin Oscar also “aim” to express an atomic natural kind concept when they utter the word ‘water’:

...the word ‘water’ – whether on Earth or on Twin Earth – must be thought of as aiming to express a natural kind concept; otherwise, the fact that water and *twater* are distinct natural kinds will not be semantically relevant.

(*ibid.*)

4.13: *The Basic Argument*

Boghossian’s *reductio* involves an a priori philosophical argument that repurposes Burge’s famous *Dry Earth* case (Burge, 1982). Dry Earth is for all appearances identical to our own Earth – and, of course, to Twin Earth also. On Dry Earth, however, there is no water. Neither is there any twater. There is *no* liquid whatsoever that flows through Dry Earth’s equivalent of rivers and from Dry Earth’s equivalent of taps and fills Dry Earth’s equivalent of lakes and seas. The inhabitants of Dry Earth, nevertheless, are all under the mass illusion that there is.

Whilst Oscar is thinking that water is wet on Earth, and Twin Oscar is thinking that twater is wet on Twin Earth, there also will be on Dry Earth *Thirsty Oscar*, who will be qualitatively identical to both Oscar and Twin Oscar. If this Thirsty Oscar is thinking a thought at all, then he’ll be doing so by deploying some sort of concept that would play the same cognitive role as Oscar’s water concept does in Oscar’s mental life, and the same cognitive role as Twin Oscar’s twater concept does in Twin Oscar’s mental life.

Boghossian contends that such a concept can be neither atomic nor non-atomic. Thus, it cannot be any kind of concept at all. Thus, poor Thirsty Oscar cannot be thinking any kind of thought at all.

Boghossian contends that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting an identical brain state is deploying an atomic concept. Thus, if Oscar and Twin Oscar’s respective concepts are atomic, then it cannot be the case that Thirsty Oscar’s is non-atomic. Thirsty Oscar’s concept cannot be non-atomic whilst Oscar and Twin Oscar’s respective concepts aren’t, as all three will be exhibiting identical brain states. Oscar and Twin Oscar’s respective concepts will both be atomic. Thus, Thirsty Oscar’s concept cannot be non-atomic.

Boghossian contends that Thirsty Oscar’s concept cannot be atomic because if it were atomic then there’d be nothing that could act as its application conditions. It is water itself that acts as the

sole application condition for Oscar's concept. Likewise, it is twater that acts as the sole application condition for Twin Oscar's concept. There is no clear, odourless liquid on Dry Earth to provide similar membership conditions for the extension of Thirsty Oscar's concept.

What we have here is a conclusion and a philosophical argument that allows us to arrive at that conclusion through armchair a priori reasoning. We have the conclusion, $\sim F$, that Thirsty Oscar cannot be thinking a thought. This conclusion, $\sim F$, is derived via our reasoning that if it were the case that, F , Thirsty Oscar were thinking a thought, then a contradiction would be true. We would have the contradiction, $\sim G \& \sim G$, that Thirsty Oscar's 'water' concept would be neither non-atomic nor atomic. Therefore, it must be the case that, $\sim F$, Thirsty Oscar cannot be thinking a thought. Crucially, this is something that we know a priori. We know a priori that, $\sim F$, Thirsty Oscar could not be thinking a thought.

Supposing that content externalism is indeed true, content externalism being true will also be something that we know a priori. We know that content externalism is true through a priori consideration of philosophical argument. (According to Boghossian, we allegedly know through consideration of Putnam's 'Twin Earth' and the intuitions it allegedly generates – see above and below, §4.12; §4.212)

By an ex hypothesi assumption a subject should be able to know a priori what she is thinking. If am thinking that water is wet, I should be able to know a priori that I am thinking that water is wet.

Boghossian ventures that these three pieces of a priori knowledge put together could confer a fourth piece of a priori knowledge. That fourth piece of knowledge concerns something that we ought not to know a priori.

Suppose I am thinking the thought that I'd express 'water is wet'. I should, Boghossian ventures, be able to know a priori that *if* I am in a world in which there exist a water-like natural kind, then I will be deploying an atomic concept that has as its extension all and only instances of that kind. A priori knowledge of this conditional should be bequeathed to me via two things. The first of those two things is my a priori knowledge of the arguments for natural kind externalism. The second thing is my knowledge, which should also be a priori, of what my semantic intentions would be were I to utter aloud 'water is wet'. I should be able to know, and know a priori, that I, like Oscar and Twin Oscar, 'aim' to express an atomic natural kind concept by my use of the term 'water'. If I, like either Oscar or Twin Oscar, am on a planet where there is a natural kind that has the properties that we believe water to have – that is clear, odourless etc. – then I, like Oscar and Twin Oscar, will be deploying an atomic concept that has as its extension all and only instances of that kind. This conditional is something that I should be able to know a priori.

If, conversely, I am not an inhabitant of such a world, if there is no such kind 'water', then I'll be in a predicament that will essentially be the same as that of Thirsty Oscar. If Boghossian's 'Dry Earth' argument is sound, this should mean that I won't be thinking a thought. If there is no such kind

‘water’ then I won’t be thinking a thought. This second conditional is something that I should be able to know a priori through following Boghossian’s ‘Dry Earth’ argument.

I, *ex hypothesi*, should be able to know a priori what I am thinking. By extension, I should know a priori that I am thinking a thought. I also know a priori that if I am thinking a thought I cannot be a victim of a predicament similar to that of Thirsty Oscar. I know a priori that I am thinking a thought, therefore I know a priori that I am not the victim of a predicament like that of Thirsty Oscar. I thus know a priori that my term ‘water’ denotes an existent natural kind. It is here we arrive at the absurdity. Hence, we have our *reductio*. It must be the case that content externalism is false, or I don’t have a priori knowledge of what I am thinking. Otherwise, I’d be able to know a priori that water exists, which would be absurd.

4.2: Three Problems for Boghossian’s *Reductio*

4.2.1: *The Three Problems*

Boghossian’s *reductio* faces a number of potential problems that may be exploitable by a would-be refuter.

Firstly, Boghossian’s *reductio* requires that it be the case that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting a (qualitatively) identical brain state is deploying an atomic concept. This, as we will see, is something that is potentially contestable.

Secondly, Boghossian’s *reductio* also requires that it be the case that Thirsty Oscar’s ‘water’ concept cannot be atomic. This is also something that is potentially contestable.

Thirdly and finally, Boghossian’s *reductio* requires it to be the case that Oscar’s concept will be atomic. Boghossian’s case for this rests on what is in fact an idiosyncratic account of the provenance of content externalism.

4.2.2: *First Problem I: Korman’s Counterexamples*

Boghossian’s *reductio* requires that it be the case that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting a (qualitatively) identical brain state is deploying an atomic concept. Korman (2006, pp.514-517) presents two alleged counterexamples to this thesis.

The first involves *Burgean externalism* (see my Chapter 1). It consists of “a variant of Tyler Burge’s arthritis example” (ibid., p.514 – see also my Chapter 1, §1.22).³⁵ The concept that is expressed by ‘H₂O’ is paradigmatically non-atomic. Suppose there is a subject who is unaware that this is the case. Let’s call him *Ignorant Iain*. This Ignorant Iain hears about a substance “aichtuwo” by listening to his socio-linguistic peers. He is extremely scientifically ignorant – but is otherwise intelligent – and is unaware that what he hears as “aichtuwo” is being used with the intention that it refer to the chemical compound ‘H₂O’. Korman contends that according to the Burgean externalist, a subject such as Ignorant Iain would possess the very ‘H₂O’ concept that his socio-linguistic peers possess. Korman draws comparisons with Alf the arthritis sufferer. Alf has an *incomplete understanding* of what ‘arthritis’ is. He doesn’t understand that the term ‘arthritis’ applies exclusively to ailments that affect the joints. He thinks that arthritis is something that can spread to the thigh. Alf nevertheless possesses the concept ‘arthritis’, that is, the concept that Alf’s socio-linguistic peers – i.e. other English speakers – express by their use of the term ‘arthritis’, or so the Burgean story goes. Korman contends that the situation that Alf is in is analogous to the situation of a subject such as Ignorant Iain. Ignorant Iain will have an incomplete understanding of what he hears as “aichtuwo”. He will nevertheless still possess the concept that his linguistic peers possess. Imagine now a counterfactual scenario in which what Ignorant Iain hears as “aichtuwo” is not being used by his socio-linguistic peers to refer to a chemical compound. Suppose that for whatever reason the phrase is being used to express an atomic concept. Like in Burge’s original arthritis case, this need not impinge on Ignorant Iain’s intrinsic properties, including his brain states. Ignorant Iain can in this imagined counterfactual scenario be exhibiting all the same brain states that he is actually exhibiting. And like Alf in Burge’s original thought experiment, Ignorant Iain will be in possession of the concept that his peers are in possession of. Thus, Ignorant Iain’s concept would be atomic. Thus, we have a case in which a subject – actual Ignorant Iain – is deploying a non-atomic concept, whilst another subject – counterfactual Ignorant Iain – is exhibiting an identical brain state but deploying an atomic concept.

There are two responses available to a defender of Boghossian’s reductio with respect to this first alleged counterexample. Firstly, it could be maintained that the counterexample shouldn’t be considered a counterexample to Boghossian’s thesis. Rather it should be considered a reductio that is specific to Burge’s particular brand of content externalism. If Burgean externalism has the consequence that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting an identical brain state is deploying an atomic concept, then ipso facto Burgean externalism must be false.

The second response does not require any such wholesale rejection of the ideas of one of content externalism’s most important pioneers. The second response is to maintain that the cases of

³⁵ Korman “owes the example” to Derek Ball (Korman, 2006, p.514, fn.17).

Ignorant Iain and Arthritic Alf are disanalogous. The reason that Alf can be said to have the concept ‘arthritis’ that his socio-linguistic peers possess, is that Alf has *some* understanding of the concept even though that understanding is *incomplete*. (He has some sort of *grasp* of the concept – see below §2.29.) There is a *standard* of understanding that Alf meets, and it is because of that that he is in possession of the concept. For Korman’s counterexample to work it must be supposed that Ignorant Iain has a very dim understanding of his peers’ concept ‘H₂O’. Thus, it could be maintained that Ignorant Iain’s understanding is insufficient for him to be in possession of the concept in the first place.

Korman’s second counterexample invokes Kripkean baptism:

Suppose that my duplicate and I each introduce a word ‘*F*’ with the following stipulation: Let ‘*F*’ express the atomic concept *two* iff there are aliens and the complex concept the *even prime* iff there are no aliens. In the actual world (let us suppose) there are aliens. My duplicate inhabits a world otherwise identical to the actual world except that there are no aliens. It would be a mistake to think that we now express one and the same concept when we say ‘*F*’ – to think, for instance, that we both express the concept *two iff there are aliens and the even prime iff there are not*. On the contrary, the nature of the stipulation guarantees that different concepts are expressed: ‘*F*’ expresses the atomic concept *two* in my mouth and the complex concept the *even prime* in the mouth of my duplicate.

(*ibid.*, p.515)

Korman admits that the idea that he is free to stipulate what concept his term ‘*F*’ expresses with the aid of conditionals in the way he describes may “raise some eyebrows” (*ibid.* p.516). But he protests that it should follow from Kripkean ideas about reference:

...if Kripke is right that I can successfully fix the semantic properties of a name for some unfamiliar mineral simply by saying “Let ‘*G*’ refer to stuff like this,” I should likewise be able to fix its semantic properties with a more explicit baptism: “If this stuff is all SiO₂ then let ‘*G*’ express a concept that applies only to SiO₂, and if this stuff is all Fe₂O₃ then let ‘*G*’ express a concept that applies only to Fe₂O₃, and if I see no reason for taking the baptism of *F* to be any less legitimate than this explicit baptism of *G*? if anything, the former is more secure, for more is known about the nature of *F* at the time of the baptism than about the nature of *G*.”

(*ibid.*, p.516)

There are three claims that Korman is making here, which Korman's counterexample requires. The first of which is correct. The second is plausible. The third is just plain wrong. Thus, Korman has not succeeded in giving a counterexample to Boghossian's thesis that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting a (qualitatively) identical brain state is deploying an atomic concept.

Korman's first claim is that that according to Kripke's views on reference it is possible to "fix the semantic properties of a name for some unfamiliar mineral simply by saying "Let 'G' refer to stuff like this,"". According to the Kripkean story about reference, it is indeed the case that a subject can fix the reference of a term 'G' by such a stipulation.

The second claim is that from this it should follow that a subject should be able to fix the "semantic properties" of the term 'G' with "a more explicit baptism: "If this stuff is all SiO_2 then let 'G' express a concept that applies only to SiO_2 , and if this stuff is all Fe_2O_3 then let 'G' express a concept that applies only to Fe_2O_3 ". Nowhere in Kripke's published material does he explicitly state that it is possible to fix the reference of a term via the aid of conditionals in the way Korman describes. Nevertheless, it is a plausible extension of Kripkean doctrine.

Thirdly, Korman is claiming that given the Kripkean story of reference, it will be possible for him to control which concept his term 'F' expresses, and in particular it is possible for him to control whether it's atomic or non-atomic. This is simply incorrect. Given the Kripkean story of reference Korman will have control of the *reference* of his term 'F'. What this means is that Korman has *some* control over the concept that the term expresses. He can control what it is that the concept is a concept *of* – by controlling what 'F' refers to – but nothing more than that. It doesn't follow that he has control over whether the concept is atomic or non-atomic. It may be the case that because of Korman's stipulation that, in both the supposed actual case in which aliens exist and in the counterfactual case in which they don't, 'F' refers to the number two. Thus, it is because of Korman's stipulation that the concept that is expressed by 'F', in both cases, is a concept *of* the number two. However, nothing in the Kripkean story of reference permits that Korman's stipulations could have any effect on whether the concept is atomic or non-atomic. Furthermore, it is not at all clear why a concept that is acquired through the sort of deployment of *other* concepts that Korman describes could be labelled as 'atomic'. The concept that 'F' expresses would arguably be non-atomic in *both* cases, as it would be a composite *of sorts* of the concepts 'the number two', 'evenness', 'aliens', etc..

4.23: First Problem II: Boghossian on the Language of Thought Hypothesis

For Boghossian's reductio to work it needs to be true that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting an identical brain state is deploying an atomic concept. We have thus far considered and (hopefully) debunked two supposed

counterexamples to that thesis, which are the two that are alleged by Korman. But, if two alleged counterexamples to a thesis are debunked, then that does not automatically mean that that thesis is therefore true. A positive case needs to be made for Boghossian's atomicity thesis.

Boghossian argues that his thesis is true because it is "hard to see how the *compositionality* of a concept could be a function of its external circumstances" (Boghossian, 1998, p.173, Boghossian's emphasis) in the way supposing that Thirsty Oscar's concept could be non-atomic requires. He cites the role of "internal syntax" (ibid, p.173) in determining whether a concept is atomic or non-atomic. He claims that the truth of his thesis "is especially clear on a 'language of thought' picture of mental representation", but states very cryptically that such an adoption of the language of thought picture is inessential to his thesis.³⁶

If Boghossian's thesis is to be argued for on the basis of the *Language of Thought Hypothesis* (the LOTH), then it needs to be demonstrated how it is that the intuitive notion of conceptual atomicity could have a basis in the operations of Mentalese. It would need to be argued that there will be some feature of the relevant bits of Mentalese that could be said to correspond to any given concept's atomicity or non-atomicity. That feature would need to be such that it would necessarily be the same in identical subjects – such as in the Oscar/Twin Oscar/Thirsty Oscar trio. There would need to be a *match up* with our intuitive notion of conceptual atomicity and what would be a more theoretical notion of conceptual atomicity that would pertain to the constituents of Mentalese strings.³⁷

If, on the other hand, Boghossian's thesis is not to be argued for on the basis of the LOTH, then Boghossian does not make it clear how else he intends it to be argued for.

4.24: First Problem III: An Alternative Argument for Why Thirsty Oscar's Concept Could not be Non-Atomic

³⁶ Here is the relevant passage quoted in full:

The compound option requires the externalist to say that one and the same word, with one and the same functional role, may express an atomic concept under one set of external circumstances and a compound decompositional concept under another set of external circumstances. But it is hard to see how the *compositionality* of a concept could be a function of its external circumstances in this way. Compositionality, as I understand it, can only be a function of the internal syntax of concept; it can't supervene on external circumstances in the way that the compound proposal would require. (This is especially clear on a 'language of thought' picture of mental representation, but is independent of it.)

Boghossian, 1997, p.172-3,

This all Boghossian gives us with respect to why he thinks that Thirsty Oscar's concept could not be non-atomic.

³⁷ There is also the issue of whether such knowledge regarding the language of thought could be a priori, which it needs to be if Boghossian's reductio is to work.

There is (I believe) a much simpler reason that can be given for why Thirsty Oscar's concept may not be non-atomic, which doesn't invoke anything as theoretically weighty as the LOTH. If Thirsty Oscar is to be in possession of his putative concept, then he must have the requisite *grasp* of his concept. In order to grasp a non-atomic concept, it is surely necessarily to have a grasp of its constituent concepts, and to have an understanding of how the non-atomic concept is, as it were, built up from those constituent concepts. I grasp the concept *bachelor*, because I grasp the concepts *marriage*, *negation*, and *human male*, and I understand how these concepts fit together to give us the concept *bachelor*. If Thirsty Oscar's concept were non-atomic, then it is very hard to see how it could be said that he grasps his concept. As far as Thirsty Oscar will be concerned his concept will be atomic. He won't even recognise it has a concept that is made up of simpler constituent concepts.

4.25: Second Problem I: Two Ways in Which Thirsty Oscar's Concept Might be Atomic

For Boghossian's reductio to work, I need to be able to rule out a priori my being in a similar predicament that Thirsty Oscar is unwittingly the victim of. The way I am meant to be able to do that is by establishing that Thirsty Oscar's 'water' concept can be neither atomic nor non-atomic. (If it can't be either, then it can't be any sort of concept at all. In which case, Thirsty Oscar can't really be thinking any sort of thought at all.) Supposing that Thirsty Oscar's concept cannot be non-atomic, and I am able to convince myself of the truth of that via a priori means, then what would remain for me to do is to establish that it can't be atomic either via similarly a priori means.

Boghossian's contention is that Thirsty Oscar's concept cannot be atomic because there would be nothing to determine what belongs to the concept's extension. There would be no *bouncer* at the door of the concept's extension calling the shots on what is and what isn't getting in. In the case of *our* concept of water, what acts as the bouncer for that concept's extension is water itself. Anything that's a sample of a substance that is the same as the substance as what flows in rivers and from taps is happily let in. Anything that isn't is barred.

Boghossian's line of argument here requires the fourth of the five commitments regarding the notion 'concept' that were outlined at the start of this chapter (§4.11), which is that there must be a *fact of the matter* with respect to what belongs to a concept's extension. A second thing that Boghossian's line of argument requires is that Thirsty Oscar's putative concept could not have application conditions. This presents two options for the would-be refuter of Boghossian's reductio. Firstly, she could challenge the commitment that there must be a fact of the matter what belongs to a concept's extension. Secondly, she could try to point to something that could constitute Thirsty Oscar's concept's application conditions.

2.26: *Second Problem II: Thirsty Oscar's Concept Does Not Need Application Conditions: McLaughlin and Tye's "Concrete Proposal"*

McLaughlin and Tye (1998a) challenge the commitment that it must be determinable what belongs to a concept's extension and argue that Thirsty Oscar's putative concept would indeed be a concept, despite the fact that it is bereft of application conditions. They present the following "concrete proposal":

Suppose [...] that the concept of water has conceptual role CR. Since [Thirsty Oscar's] concept has the same conceptual role as Oscar's concept (which is, of course, the concept of water), [Thirsty Oscar's] concept has role CR....What concept is [Thirsty Oscar] actually expressing when he uses 'water'? He is expressing the concept one possesses if and only if one has an atomic concept with conceptual role CR, but nothing satisfies the reference-fixing description associated with it, and there is no kind, natural or motley, at the end of the relevant causal chains leading to one's use of it.

(*ibid.*, p.307)

What McLaughlin and Tye are describing here is a mental entity E that can feature in a subject S's cognitive processes, that features in Thirsty Oscar's cognitive processes, and that is identified by the following two properties. Firstly, E is such that it could only feature in a subject S who is a denizen of either Dry Earth, or a similar planet that is superficially identical to Earth but contains no water. Secondly, E is such that it *would* have constituted S being in possession of an atomic natural kind concept – i.e. an atomic concept of which only instances of a particular natural kind are permitted within its extension – *had* the subject been a denizen of either Earth, Twin Earth or a similar planet on which there exists a water-like natural kind. McLaughlin and Tye contend that because we can describe, in this way, the conditions under which E would feature in a subject S's cognitive processes, it is therefore the case that if E features in a subject S's cognitive processes then that constitutes her being in possession of a concept, which is the very concept that Thirsty Oscar is in possession of. That concept would be of a different character to the concept of water, and to the concept of twater. There wouldn't be a natural kind such that something falls under it if and only if that something is an instance of that kind. But it would be a concept nonetheless.

McLaughlin and Tye do not state what such a concept's application conditions would be. They do not state what it would take for something to fall under it. They offer no argument as to why such a concept, or any concept, need not have application conditions. Boghossian's question is – if Thirsty Oscar's putative concept is indeed a concept, then what could constitute that concept's application conditions? McLaughlin and Tye response to that question is – yes, Thirsty Oscar's

putative concept is indeed a concept, its *possession conditions* are such and such. To which the defender of Boghossian will respond by angrily again asking – *if Thirsty Oscar's putative concept is indeed a concept, then what could constitute that concept's application conditions?!*

4.27: *Second Problem III: Three Ways in which Thirsty Oscar's Concept Could have Application Conditions*

There may be other ways the would-be refuter of Boghossian's reductio may endeavour to challenge the commitment that there must always be a fact of the matter with respect to what belongs to a concept's extension.³⁸ If, however, it is to be maintained that the concept must have application conditions, then the would-be refuter of Boghossian's reductio could opt for a different route by arguing that it *would be* determinable what belongs to Thirsty Oscar's concept's extension.

There are at least *three* potential ways in which she might be able to do this.

Firstly, she could argue that Thirsty Oscar's concept could be a *motley concept*. She could argue that the concept could be such that it applies to all and any liquid that has the superficial properties that water has.

Secondly, she could argue that Thirsty Oscar's concept could be *necessarily vacuous*. She could argue that the concept could be such that nothing could possibly fall under it.

Thirdly and finally, the would-be refuter could argue that although what Twin Oscar calls 'water' is illusory, it still in some sense exists, and it is still able to function as the thing that

³⁸ It could possibly be argued that the phenomenon of vagueness challenges Boghossian's contention that it must be determinable what belongs to a concept's extension. Take for instance the concept tall for an adult male. Anything that is an adult male that is taller than 6'2" will most definitely fall under that concept. Adult males that aren't as tall as 5'7" most definitely will not fall under that concept. Arguably, there will be a range of heights somewhere in between 5'7" and 6'2" such that for any adult male whose height is within that range there will be no fact of the matter with respect to whether that adult male falls under that concept. If that is indeed the case, then the poor men whose height falls within that range will have in a sense been failed by the concept. The concept has failed to sort out which of those poor men fall under it, and which don't. Nevertheless, tall for an adult male is most definitely a bona-fide concept.

There are a couple of things to note with such a proposal. Firstly, the problem with Thirsty Oscar's putative concept is not that it would be vague. The problem with Thirsty Oscar's concept would not be that it divides the world up into things that fall under it and things that don't, apart from a problematic but restricted range of instances. The problem with Thirsty Oscar's concept would be that it wouldn't divide the world up into things that do and do not fall under it at all. The malaise afflicting vague concepts is of a different character than that that would afflict Thirsty Oscar's concept. Secondly, vague concepts such as tall for an adult male are regarded as a serious philosophical problem. This is partly due to concepts being regarded as things that should divide the world up into things that do and do not fall under them, and vague concepts apparently fail to do so. Rather than contradict the thesis, the phenomenon of vagueness, and our reaction to the phenomenon of vagueness, possibly confirms it.

determines what belongs to Thirsty Oscar's concept. Something falls under the concept if and only if it is the same as the illusory liquid that Thirsty Oscar finds all around him.

4.28: *Second Problem IV: Thirsty Oscar's Concept is 'Motley'*

The first way in which it may be possible that Thirsty Oscar's 'water' concept could in fact have application conditions is suggested by Korman. Korman suggests that it could be that Thirsty Oscar's concept is such that a sample of any liquid that has the superficial properties that water has, whether it be water, twater, or something else, will fall under it. It could be that:

... 'water' expresses a concept that applies, with respect to all counterfactual situations, to all and only samples with the superficial features that water was believed to have.

(Korman, 2006, p.508)³⁹

The problem with this suggestion is that it is hard to see how a concept could be such that it would apply to any liquid that has the superficial properties that water has, and yet be atomic. It is easy to conceive of a non-atomic concept that has such application conditions. Such a non-atomic concept would be one that can be *broken down* into the concepts of 'odourlessness', 'clearness', etc..

4.29: *Second Problem V: Thirsty Oscar's Concept is Necessarily Vacuous*

To solve what I am referring to has the second problem it needs to be demonstrated emphatically that if Thirsty Oscar is indeed in possession of a 'water' concept then it cannot be atomic. Otherwise the would-be refuter is afforded a possible line of attack. Boghossian's case for it not being atomic is that if it were atomic it could not have application conditions.

We have so far considered and discounted one suggestion for how it could be that Thirsty Oscar's concept could be both atomic and have application conditions, which is that the concept could be *motley*. There is, however, a second way in which it may be possible that Thirsty Oscar's 'water' concept could in fact have application conditions. It could be that Thirsty Oscar's concept is *necessarily vacuous*.

³⁹ What I'm quoting here is from the third of what Korman describes as "a series of "default conditionals" which he proposes govern the semantics of the word 'water' given certain contingencies regarding the nature of 'water' (Korman, 2006, pp.507-8).

This second suggestion is also mooted by Korman (Korman, 2006), and by others (e.g. Goldberg, 2005) as well.

A *necessarily vacuous* concept is a concept that has an empty extension not because there isn't anything in the actual world that meets its application conditions, but because there is nothing that could possibly belong to its extension. If Thirsty Oscar's concept was necessarily vacuous then in no possible world will there be any substance that falls under his concept.

This second suggestion is also problematic. Integral to our notion of what a concept is is the idea that concepts are *grasped*. To possess a concept is to grasp that concept. In a similar way that it is hard to see how Thirsty Oscar's could be said to grasp his concept if it were non-atomic – see above (§4.24) – it is also hard to see how he could be said to grasp his concept if it were a necessarily vacuous concept. As far as Thirsty Oscar is concerned his concept applies to an everyday substance that he believes is all around him. Yet according to what is proposed here, Thirsty Oscar's concept would in fact be such that nothing falls under it, and nothing *could* fall under it.

Possibly, if we consider other putative instances of persons being, like Thirsty Oscar, unwittingly in possession of necessarily vacuous concepts, we could draw parallels between those instances and the situation of Thirsty Oscar. That is, if we are able to identify what constitutes the subjects of such instances' grasp of their respective concepts, then that might give us clues as to what could constitute Thirsty Oscar's grasp of *his* concept.

Firstly, consider the concept *largest prime*. This concept although necessarily vacuous, will be non-atomic. What will constitute a subject's grasp of the concepts will be an understanding or grasp of their constituent concepts. In order to grasp the concept 'largest prime' it is necessary and sufficient to grasp what it is for something to be the largest of a given set of numbers and to grasp what a prime number is. A subject can grasp what it is for something to be the largest of a given set of numbers, and to grasp what a prime number is, but be unaware that there are an infinite number of primes. Thus, a subject can grasp the concept 'largest prime' without being cognizant of its necessary vacuity. This, however, offers us little help in our present endeavour of trying to find something that could constitute Thirsty Oscar's grasp of his supposedly necessarily vacuous concept. The whole point of supposing that Thirsty Oscar's concept is necessarily vacuous is to allow it to be atomic. It can't be the case that Thirsty Oscar is able to grasp his concept by grasping its constituent concepts.

Consider another (putative – at least) necessarily vacuous concept. Consider the concept that is expressed by the proper name 'Sherlock Holmes'. According to Kripkean theory, since Sherlock Holmes is a fictional character, the concept 'Sherlock Holmes' will be necessarily vacuous. The majority of the populace, however, won't recognise it as such. The majority of the populous won't have read *Naming and Necessity* (nor will they have read the belatedly published *Reference and Existence: The John Locke Lectures* (Kripke, 2013)). Thus, they won't have had the opportunity to have been convinced by Kripke's arguments that fictional names are necessarily vacuous. And thus,

they won't be in a position to recognise the concept 'Sherlock Holmes' as being necessarily vacuous. Nevertheless, the majority of the populace will still most definitely be in possession of the concept 'Sherlock Holmes'. *Ipsa facto*, they must all have the requisite grasp of the concept.

The reason why the concept 'Sherlock Holmes' will be necessarily vacuous will be because something falls under it if and only if it is numerically identical to the fictional creation Sherlock Holmes. Since there could be nothing in the real world that meets that application condition – there is nothing in the real world that could be identical to a fictional creation – there could not possibly be anything in the real world that falls under the concept.

Grasping the concept 'Sherlock Holmes' should be a matter of having an understanding of who Sherlock Holmes is. There are at least three things that could be said to constitute the majority of the populace's understanding of who Sherlock Holmes is, and thus constitute their grasp of the concept 'Sherlock Holmes'. Firstly, there is the significant body of knowledge about Sherlock Holmes that the majority of the populace will be in possession of. They'll know he is a fictional detective. They'll know, or most of them will know that he was created by Sir Arthur Conan Doyle. They'll know (probably) that he is played in an acclaimed TV adaptation by Benedict Cumberbatch – and so on. Secondly, the majority of the populace will have an understanding of how their socio-linguistic peers use proper nouns. In particular, they will have an understanding of how the names of persons – such as 'Sherlock Holmes' – that they hear being used all around them function. They'll understand how names of persons are passed on from user to user. And, they'll understand that at the beginnings of such chains of communication usually resides a new-born baby who has been baptised such and such, or, as is sometimes the case, a fictional character whose creator has named such and such. Thirdly, knowing all that about the functioning of proper names together with the significant body of knowledge about Sherlock Holmes that the majority of the populace will be privy to could be said to what jointly constitutes the majority of the populace's grasp of the concept 'Sherlock Holmes'.

It could be argued that the majority of the populace's grasp of the concept 'Sherlock Holmes' is predicated on them knowing that Sherlock Holmes is a fictional character. It could be argued that someone who does not know that Sherlock Holmes is fictional, who (say) believes Sherlock Holmes was a real-life Victorian detective, couldn't be said to grasp the concept, and thus couldn't be in possession of the concept. If this is so, then any chance for us drawing parallels between this second example and the case of Thirsty Oscar's concept is obviously scuppered. There will be a marked difference between the possessors of the 'Sherlock Holmes' concept and Thirsty Oscar. The former will know that their concept concerns something that doesn't exist, whereas Thirsty Oscar does not.

Consider a third example of a necessarily vacuous concept, which is similar, but which involves subjects not knowing that they are referring to a fictional character. Suppose that Napoleon never really existed. Suppose that Napoleon was invented as part of a very elaborate hoax executed with great competency by late nineteenth century propagandists (say). Contrary to wide-spread belief,

Napoleon is in fact a fictional character. If we believe we are using ‘Napoleon’ to refer to a real-life-flesh-and-blood person when in fact we are not, then it is questionable whether we could be said to grasp the concept that ‘Napoleon’ on our lips purportedly expresses. As I’ve already intimated, taking either side with respect to this matter could prove controversial. Nevertheless, if it is conceded that we would grasp the concept ‘Napoleon’, then there’ll be the same three candidates, *mutatis mutandis*, for what could constitute that grasp, as there was in the case of the concept ‘Sherlock Holmes’. Firstly, there’ll be the body of knowledge concerning Napoleon that we’d be privy to. Yes, a lot of what we think we know about Napoleon would be wrong. We’d, for instance, believe falsely that Napoleon was a real person who declared himself Holy Roman Emperor. However, there would still be a great body of stuff that we’d still know about Napoleon. We’d know about the things that are *alleged* with respect to Napoleon. We’d know that it is *alleged* that he declared himself Holy Roman Emperor, etc.. Our being privy to this body of things that Napoleon is alleged to have done, might be what would constitute our grasp of the concept ‘Napoleon’. Secondly, what would constitute our grasp of the concept ‘Napoleon’ could be our understanding of how there’ll be a socio-linguistic chain determining the reference of the proper noun ‘Napoleon’. It’s just that we wouldn’t know that it so happens that that socio-linguistic chain began with nineteenth century propagandists inventing a fictional character called ‘Napoleon’. Thirdly, our grasp of the concept ‘Napoleon’ could be constituted jointly by our being privy to things that are alleged with respect to Napoleon, and by our understanding of how the term ‘Napoleon’ will have had its reference fixed.

Is this third example any more helpful with respect to our present purposes? Well, suppose that Thirsty Oscar’s distant ancestors decided to make up a fictional natural kind that they called ‘water’. Suppose that they characterised this non-existent ‘water’ as having certain properties. They said that it was a clear, odourless liquid. They said it was wet – and so on. Finally, suppose that it has now been forgotten that the original intention was that ‘water’ was to name a fictional kind. It is now believed that there is in fact a natural kind ‘water’ and it has all – or at least most – of the properties that the ancients described it as having. If this were the case, then Thirsty Oscar would have read about ‘water’ and its alleged properties in the relevant ancient texts. Thus, he could be said to possess a fictional, and therefore necessarily vacuous concept, even though he doesn’t recognise it as such. His grasp of the concept could be said to be grounded in knowing about water’s alleged properties, or in him knowing about how the term ‘water’ will have received its reference, or in both of those things. However, what is being described here is very far from what we are in fact supposing is the case when it comes to Thirsty Oscar’s use of the term ‘water’. Thirsty Oscar intends that his term ‘water’ refer to the clear, odourless liquid that he believes is all around him. It is not something he has read about in ancient texts.

Finally, there is a fourth sort of case that might involve a subject being in possession of a necessary vacuous concept and not recognising it as such. J. J. Becher may have stipulated that

‘phlogiston’ refer to *the substance that is present in all combustible substances and that is released during combustion* with the sort of reference-fixing intentions that are described throughout *Naming and Necessity*. That is, in doing so Becher won’t have been stipulating that ‘phlogiston’ be *shorthand* for ‘the substance that is present in all combustible substances and that is released during combustion’. He would have been doing something else. He would have been attempting to *fix the reference* of ‘phlogiston’ – à la Kripke. An arguable consequence of that would be is that because there is no substance that was picked out by Becher’s stipulation – because Becher *failed* in his attempt to fix the reference – the concept that Becher expressed with ‘phlogiston’ will have been necessarily vacuous. Becher could have nevertheless believed that there really was a substance present in combustible substances and was released during combustion, and that his concept was not empty, let alone necessarily so.

If it is to be maintained that Becher is here indeed in possession of the necessarily vacuous concept that is here being supposed he expressed with the term ‘phlogiston’, then there must be something that can be said to have constituted his grasp of that concept. What could this be? Well, Becher will have been ignorant with respect to the true nature of the concept’s application conditions. However, he will have been cognizant of the *mechanism* that determines the application conditions of his concept. This consisted of him stipulating that ‘phlogiston’ refer to the substance that is present in all combustible substances and that is released during combustion. It is just that he won’t have known that that will have led to his concept being necessarily vacuous. He will have had a sort of *proxy* epistemic access to the concept’s application conditions. This seems to be the best that can be done with respect to presenting a candidate for what could have constituted Becher’s grasp of his concept.

Does any of this help us in our quest to try and work out what could constitute Thirsty Oscar’s grasp of his concept? Well, if Thirsty Oscar had acquired his concept by virtue of him stipulating that ‘water’ refer to ‘the clear, odourless liquid that flows in rivers, and taps’, then by the same lights, Thirsty Oscar’s concept would be necessarily vacuous. And, by the same lights, Thirsty Oscar could be said to grasp his concept – by having the same sort of proxy epistemic access to the concept’s application conditions. However, this – again – is very far from what we are in fact supposing is the case when it comes to Thirsty Oscar’s use of the term ‘water’. We are not supposing that Thirsty Oscar would have acquired his concept in this way.

Also, if Thirsty Oscar had acquired his concept in this way, then it is doubtful that it could be described as being ‘atomic’. As is the case in the Korman example (§4.22), Thirsty Oscar’s concept would similarly be a *composite of sorts* of other concepts.

In summary, what we’ve been considering in this subsection is a suggestion for how it could be that Thirsty Oscar’s concept could have application conditions, which is that Thirsty Oscar’s concept could be *necessarily vacuous*. If Thirsty Oscar’s concept could be shown to have application conditions, then that would derail Boghossian’s reductio. Thus, if it could be shown that Thirsty

Oscar's concept would be necessarily vacuous, then that would derail Boghossian's reductio. The suggestion is problematic, because it is hard to see how Thirsty Oscar could be said to *grasp* his concept if it were necessarily vacuous. We considered putative instances of persons being unwittingly in possession of necessarily vacuous concepts and attempted to draw parallels between those instances and the situation of Thirsty Oscar. We adopted, on behalf of the would-be refuter of Boghossian's reductio, the hope that if we could identify what constitutes such person's grasp of their respective concepts, then that might give us clues as to what could constitute Thirsty Oscar's grasp of his concept, if it were indeed necessarily vacuous. Unfortunately for the would-be refuter, this hope, at least in the four examples we've considered, is yet to come to fruition.

4.210: Second Problem VI: Necessary Vacuity and the Three Burgean Paradigms

Partial conceptual ignorance is essentially what Burgean externalism is built upon. The three Burgean routes to content externalism all involve subjects who lack a complete grasp of a certain concept that Burge maintains that they nevertheless possess. Alf the Arthritis Sufferer (see my Chapter 1, §1.22) possesses the same 'arthritis' concept that his socio-linguistic peers possess, even though he has an *incomplete understanding* of that concept. Alf is unaware that the term 'arthritis' is only applicable to ailments of the joints. Oscar, in Burge's characterisation of 'Twin Earth', possesses the concept 'water' even though he is *ignorant of expert knowledge* in that he does not know that water is H₂O (see Chapter 1, §1.23). Person A possess the concept 'sofa' even though he subscribes to a *non-standard theory* with regards to what sofas are (see Chapter 1, §1.24).

Although none of these three cases involve necessary vacuity, a question that could still be asked is would any of these three examples be sufficiently analogous to the case of Thirsty Oscar and his ex hypothesi necessarily vacuous 'water' concept to warrant meaningful comparison? Remember that the aim of the would-be refuter of Boghossian's reductio, if she is to go down the route of supposing that Thirsty Oscar's concept could be necessarily vacuous, is to demonstrate that Thirsty Oscar could be said to grasp his concept even though he is totally unaware that it is necessarily vacuous. In the previous subsection, we considered ways in which the would-be refuter might be able to do this. Comparisons with the Burgean cases might afford the would-be refuter further ways of pursuing this aim. Since we have already spent what is probably too much time on what is a tangential topic, I won't here consider whether that line of enquiry could bear fruit.

The assumption that underpins the three examples can be viewed as a *sufficient condition* for conceptual grasp. If an ascriber can be ascribed thoughts with a term 'φ' in oblique position by his or her socio-linguistic cohorts, then that should mean that the ascriber can be said to grasp the concept that the term 'φ' expresses. If an ascriber can be ascribed thoughts with the term 'water' in oblique position by his or her socio-linguistic cohorts, then that should mean that the ascriber can be said to

grasp the very concept ‘water’ that his or her socio-linguistic cohorts possess. This sufficient condition, however, is not applicable in the case of Thirsty Oscar. When we ask whether Thirsty Oscar could be in possession of a necessarily vacuous concept, we are really asking whether Dry Earthlings as a whole – i.e. Thirsty Oscar’s socio-linguistic cohort – can be in possession of such a concept.

In lieu of a substantial account of what it is to grasp a concept – i.e. one consisting of a list of jointly-sufficient necessary conditions – to measure the case of Thirsty Oscar and his putative necessarily vacuous concept against, we can only measure the case of Thirsty Oscar and his putative necessarily vacuous concept against other relevant cases of conceptual grasp, which may or may not involve necessary vacuity. Alas, such comparisons with the four examples of the previous subsection have not proven fruitful. I am leaving it as an open question whether comparing the case of Thirsty Oscar and his putative necessarily vacuous concept with the examples that Burge provides could prove fruitful.

To reiterate the summation that ended the previous subsection, our enquiry has not of yet yielded a strong case for supposing that Thirsty Oscar’s concept could be necessarily vacuous, which would lead directly to a refutation of Boghossian’s reductio. However, we are not necessarily therefore left with a strong case for the opposite conclusion – that Thirsty Oscar’s concept could not be necessarily vacuous. There could still be hope for the would-be refuter of Boghossian’s reductio if she were to pursue this line.

4.211: *Second Problem VII: Illusory Bouncers*

Boghossian’s reductio requires that it must be the case that Thirsty Oscar’s ‘water’ concept cannot be a bona-fide concept. Boghossian argues that if Thirsty Oscar were in possession of a concept then that concept could neither be non-atomic nor atomic. Thus, it can’t be the case that Thirsty Oscar is in possession of a concept. A would-be refuter of Boghossian’s reductio could argue that Thirsty Oscar could be in possession of a non-atomic concept. Failing that, she could instead try to argue that Thirsty Oscar could be in possession of an atomic concept. Boghossian argues that Thirsty Oscar cannot be in possession of an atomic concept because there could be nothing that could constitute the application conditions of his concept. The would-be refuter could argue that Thirsty Oscar’s concept would not need application conditions. Failing that, the would-be refuter could instead argue that Thirsty Oscar’s putative concept would in fact have application conditions. To do so successfully the would-be refuter must stipulate what those application conditions could be. The would-be refuter could argue that Thirsty Oscar’s concept could be a *motley concept*, in that it applies to all and any liquid that has the superficial properties that water has. A second option would be to try to argue that Thirsty Oscar’s concept could be *necessarily vacuous*, in that nothing could possibly fall under it.

There is also a *third* option that might be open to the would-be refuter of Boghossian's reductio, which is what I turn to in this present subsection.

A would-be refuter of Boghossian's reductio could argue that, although it may seem a contradiction in terms, illusions are things that do in fact have an *existence of sorts*. Illusions are things that have spatio-temporal coordinates. They have perceptual properties, albeit ones that only the sufferer of the illusion is privy to. Illusions are things that are caused by other things – by psycho-illusion-agentic drugs and by psychosis – and, through the actions of those who suffer them, they in turn cause further things to happen.

A would-be refuter of Boghossian's reductio could argue further that illusions could act as the things that determine what belongs to a concept's extension – could act as the bouncer at the door of a concept's extension – in the same way that something that is 'real' could.

Suppose that the denizens of Dry Earth suffer the illusion that there is this watery substance that is all around them because disseminated throughout the atmosphere of Dry Earth there is a psycho-illusion-agentic substance that has been very cleverly chemically engineered to produce in people the illusion that there is this watery substance that is all around them. (Such a substance would be along the lines of what the psycho-chemists of the chemocracy depicted in Stanislaw Lem's dystopian novel *The Futurological Congress* might create.) We could say that something falls under Thirsty Oscar's 'water' concept if and only if it is a sample of the illusory watery substance that that cleverly chemically engineered substance causes people to see (or otherwise sense).

Imagine now a *second* Dry Earth. On this *Dry Earth 2* there is similarly no water or water-like substance, and it is similarly the case that its denizens all are under the mass illusion that there is some sort of water-like substance that is all around them. On Dry Earth 2, however, the illusion is not caused by the cleverly chemically engineered substance that causes the illusion on Dry Earth 1. The illusion of 'water' on Dry Earth 2 is caused by a *different* cleverly chemically engineered substance. It could be argued that the concept that 'water' expresses on the lips of the denizens of Dry Earth 2 would therefore be different to the concept that 'water' expresses on the lips of the denizens of Dry Earth 1. We have two different illusory substances constituting the sole application conditions of the two respective concepts.

We can also imagine a *third* and a *fourth* Dry Earth. On *Dry Earth 3* and on *Dry Earth 4* the mass illusion of a water-like substance is not caused by a cleverly chemically engineered substance. On Dry Earth 3 the illusion is instead caused by a weirdly coincidental mass psychosis. On Dry Earth 4 the illusion is similarly caused by a weirdly coincidental mass psychosis, but there are important physiological differences between what is happening in the brains of the denizens of Dry Earth 3 to what is happening in the brains of the denizens of Dry Earth 4 that mark their respective psychoses out as being different illnesses. We could say that the concept expressed by 'water' on the lips of the denizens of Dry Earth 3 would thus be different to the concept expressed by 'water' on the lips of the

denizen of Dry Earth 4. We could also say that the concepts will also be non-identical to the respective ‘water’ concepts of the denizens of Dry Earth 1 and of Dry Earth 2. We thus now have *four* different illusory substances constituting the sole application conditions of the four respective concepts.

4.212: The Third Problem: Boghossian’s Account of the Provenance of Content Externalism

Boghossian’s *reductio* requires that it be the case that it is not possible that a subject could be deploying a *non-atomic* concept, whilst another subject who is exhibiting a (qualitatively) identical brain state is deploying an *atomic* concept. This, as we’ve seen, is something that is contestable. Boghossian’s *reductio* also requires that it be the case that Thirsty Oscar’s ‘water’ concept cannot be atomic. This, as we’ve seen, is also something that is contestable. Boghossian’s *reductio* also faces a third problem. Boghossian’s *reductio* requires that it must be supposed that Oscar’s concept is atomic. Boghossian contends that this follows from what he takes to be the provenance of content externalism. The problem is that Boghossian mischaracterises the provenance of content externalism:

...let us remind ourselves how Putnam’s thought experiment is supposed to work. Whereas Oscar, an ordinary English Speaker, lives on Earth, his molecular and functional duplicate, Toscar, lives on Twin Earth, a planet just like Earth except that the liquid that fills its lakes and oceans, while indistinguishable from Earthly water in all ordinary circumstances, is not H₂O but some other substance with a different chemical composition – call it XYZ. Going by whatever criteria are relevant to such matters, water and twin water are distinct kinds of substance, even though a chemically ignorant person wouldn’t be able to tell them apart. Now widespread intuition appears to have it that, whereas Oscar’s tokens of ‘water’ apply exclusively to H₂O, Toscar’s tokens of ‘water’ apply exclusively to XYZ. Widespread intuition appears to have it, in other words, that Oscar’s and Toscar’s ‘water’ tokens have distinct extensions.
(Boghossian, 1997, p.164)

Boghossian claims that for this “widespread intuition” to be generated:

...the concepts expressed by the Earthly and Twearthly tokens of ‘water’ have to be thought of as atomic concepts, not compound concepts that are compositionally built up out of other concepts in well defined ways. For example, the experiment presupposes that water [i.e. the concept expressed by ‘water’] can’t be thought of as capable of being defined as: A tasteless, odourless liquid that flows in the rivers and faucets. For if it were

a compositional concept of that sort, its extension would be determined by the extension of its ingredient parts. Hence, a conclusion to the effect that water and twater have different extensions would have to proceed differently than it does in Putnam's original experiment, by showing that one of the *ingredients* of water—the concept expressed by 'liquid', for example—has a different extension from that expressed by its Twin counterpart.

(*ibid.*, p165)

Boghossian misattributes an argument to Putnam, which involves *allegedly alleged* intuitions, that is, intuitions that Boghossian alleges that Putnam and his defenders allege are generated via consideration of Twin Earth. This argument is in fact a very weak one. Essentially, the argument is that intuitively it seems that it must be the case that certain thoughts of Oscar that he'd express using the word 'water', will – if it assumed that that concept is atomic, otherwise *mutatis mutandis* for a re-orientated thought experiment – have different contents to the corresponding thoughts of Twin Oscar. Thus, there can be instances in which two atom-for-atom duplicates can be thinking thoughts with different contents. Thus, content externalism must be true. Boghossian is alleging that it is alleged that consideration of Twin Earth is meant to deliver the strong – or at least *fairly* strong – intuition that Oscar and Twin Oscar's thought contents cannot be identical, and thus the conclusion that content doesn't supervene on intrinsic properties. The problem is that this just simply isn't the case. Consideration of Twin Earth – in itself – does not generate any such intuition.

For Boghossian's *reductio* to work – for it to get off the ground – it does need to be supposed that Oscar's and Twin Oscar's respective concepts will be atomic. Boghossian's has presented us with an illegitimate reason for why we should suppose this would be the case. That does not necessarily mean that it won't be the case. Nevertheless, Boghossian has presented us with an illegitimate reason for supposing it would be the case.

This ends our discussion of Boghossian's *reductio*. We have, in this present chapter, examined Boghossian's *reductio* argument, in order to try to assess whether it leads us to incompatibilism. In doing so we have possibly given the compatibilist hope by identifying three potential – potentially solvable *perhaps* – problems for the defender of Boghossian's argument. The first is that it is not clear that it must be the case that it is not possible that a subject could be deploying a non-atomic concept, whilst another subject who is exhibiting an identical brain state is deploying an atomic concept (§4.22-4.24). The second problem is that it is not clear that Thirsty Oscar's 'water' concept cannot have application conditions if it is supposed that it must be atomic (§4.25-4.211). The third problem is that Boghossian's reason for why we should suppose that Oscar's concept is atomic is illegitimate.

In the next chapter we consider whether the compatibilist cause fares any worse in the face of another *reductio* argument that is meant to establish incompatibilism. We examine Brown's *reductio* argument.⁴⁰

⁴⁰ Boghossian's *reductio* may also face a fourth and even a fifth problem.

There is a certain sort of valid argument where it is possible to have a degree of epistemic entitlement to the premises without also having that same epistemic entitlement to the conclusion – even though the conclusion may follow directly from the premises. These arguments, which we can call *warrant transmission failure arguments*, have as their first premise what is usually some everyday fact that it can be taken for granted that we know, or that we have at least good very grounds for believing, i.e. we have *warrant* for believing. They have as their second premise a true conditional that has as its antecedent that everyday fact, and has as its consequent the negation of some outlandish scenario that could not be case if the everyday fact were true, and that crucially we don't have the same warrant for believing to be false as we have for believing the everyday fact to be true. They have as their conclusion that consequent.

If there are indeed such warrant transmission failure argument, then Moore's 'proof' (Moore, 1959) of the external world would be the most infamous example of one. I have hands. I surely have warrant for believing that I have hands. *Here's one, here's another!* If I have hands, then I am not a BIV. It should therefore follow that I am not a BIV. It seems somehow, however, that, contra Moore, I am not warranted to conclude therefore that I am not a BIV. The warrant I have for believing I have hands somehow fails to transmit to what would be an obvious consequent of my having hands.

Boghossian *reductio* involves a subject engaging in what is effectively a modus ponens argument. A subject is meant to start with the *ex hypothesi a priori* knowledge that she is thinking that water is wet. Via consideration of the relevant philosophical arguments she is meant to come to know *a priori* that if she is thinking that water is wet, then that entails that water exists. That in turn is meant to deliver to the subject the *a priori* knowledge that water exists.

It has been suggested (Davies, 2003; Wright, 2000; 2003; et al – also see Ch. 7 of Brown, 2004; Beebe, 2001; Brueckner, 2008; Schiffer, 2005; Wright, 2011; et al) that this modus ponens argument could be a warrant transmission failure argument, and this could allow us to say that although the premises can both be conferred the epistemic status of knowledge, that same epistemic status cannot be conferred on the conclusion. This potentially constitutes a further problem for Boghossian's *reductio*.

As I noted in the Introduction, this very cumbersome topic is something that I omit from this present thesis. This is largely due to considerations of space, and I may properly consider the topic in potential sequels.

Brown's own (earlier) *reductio*, which is the subject of the next chapter, and my own *reductio* argument, which I present in Chapter 6, involve modus ponens arguments that are similar to the modus ponens argument that forms the centre piece of Boghossian's *reductio*. They both involve a subject deducing – and thus coming to know *a priori* – something that she could not possibly know *a priori* based on her *ex hypothesi a priori* knowledge that she is thinking some thought with some content. Thus, they are both potentially vulnerable to similar warrant transmission failure considerations.

Boghossian's *reductio* also involves another modus ponens argument. Boghossian's *reductio* requires that based on a subject's *ex hypothesi a priori* knowledge that they are thinking that water is wet, she is entitled to move to the conclusion that therefore she is thinking a thought. (That in turns enables her to conclude that she is not in the same predicament that Thirsty Oscar is in, as Thirsty Oscar couldn't be thinking a thought, as the concept he would be deploying if he were thinking a thought could neither be atomic nor non-atomic.) Thus, we have a modus ponens argument that has as its first premise that the subject is thinking that water is wet. Its second premise is that if the subject is thinking that water is wet, then she is thinking a thought. Its conclusion is that the subject is thinking a thought.

Brown (2004, pp.287-9) has argued that it can be the case that you know you're thinking that water is wet without also knowing that you're thinking a thought. Brown's cites a "widely accepted" (ibid., p.288) necessary condition for knowledge:

...S knows that p only if, were p false, she wouldn't believe that p.
(ibid.)

Brown notes that in the possible worlds in which a subject wouldn't be thinking that water is wet would be worlds in which she doesn't believe she is thinking that water is wet. In a world in which she is thinking that water is wet she wouldn't believe that she is thinking that water is wet. She would believe instead that she is thinking that water is wet. This would not be so in possible worlds in which the subject is not thinking a thought at all. In such worlds the subject will still believe falsely that she is thinking a thought. She won't, however, believe falsely that she is thinking that water is wet. Thus, the subject can be said to know that she is thinking that water is wet, and not know that she is thinking a thought.

It does seem extraordinary to suppose, as Brown does, that it could be possible to know that you're thinking that water is wet without also automatically knowing that you're thinking a thought. It is like supposing that I could know that I am wearing a red T-shirt without knowing that I am wearing a T-shirt.

It could be argued that what Brown as in fact done is present an inconsistent quadruple – the only way out of which is to reject either ex hypothesi claim the we can know a priori what we are think or content externalism. The conclusion that is it possible for a subject to know that she is thinking that water is wet, and not know that she is thinking a thought would essentially rests on three assumptions, which are the a priori claim, content externalism, and the necessary condition that Brown cites – that a subject knows that p only if, were p false, the subject would not believe that p. So, we must reject the conclusion's converse – i.e. except the conclusion – or we must reject one of the three assumptions – we have an inconsistent quadruple that consists of the conclusion's converse and the three assumptions. What should we reject? We *should*, contra Brown, maintain that if someone knows that they are thinking that water is wet, then they should also know that they are thinking a thought – supposing otherwise is like supposing that I could know I am wearing a red T-shirt, without knowing that I am wearing a T-shirt. In other words, we should not accept the conclusion. In other words, we should not reject the conclusion's converse. That leaves the a priori claim, content externalism, or the necessary condition. If the necessary condition is as "widely accepted" has Brown intimates, then it is perhaps not the necessary condition that we should reject. Thus, we need to either reject the priori thesis or content externalism. In other words, we have established incompatibilism!

Chapter 5: Inferring from Mind to World II: Brown's Reductio

5.0: Introduction to Chapter 5

In the previous chapter we considered Boghossian's attempt to demonstrate that if content externalism is indeed true – or more specifically if it is true that Oscar and Twin Oscar, the co-stars of Putnam's 'Twin Earth', do indeed think thoughts with differing contents – and if it is also true that we have an a priori epistemic access to our thoughts, then it will absurdly be the case that we would be able to know a priori that water exists. In this present chapter I consider a similar attempt by Brown at demonstrating that if content externalism is true, specifically if *Burgean externalism* (see Chapter 1, §1.2) – or rather Brown's interpretation of what Burgean externalism is (see below) – is true, and it is also true we have an a priori access to our thought contents, then it would follow that we could know a priori something that we similarly ought not to be able to know a priori.⁴¹

⁴¹ The two attempts are both implementations of what McLaughlin and Tye call "McKinsey's recipe" (after McKinsey, 1991):

Michael McKinsey (1991) has described a recipe for making a case that externalist theses are incompatible with privileged access [actually *a priori* access – see my Chapter 7, §7.2]. Suppose a version of externalism implies that certain thought contents are individuated, at least in part, by environmental factors. Let *that p* be a thought content of the type in question. McKinsey's recipe for trying to show that the version of externalism is incompatible with privileged access is essentially the following: find some *E* such that (1) *E* cannot be known a priori...yet (2) the version of externalism implies that it is a conceptual truth that if one is thinking that *p*, then *E*.

(McLaughlin and Tye, 1998a, p.290)

5.1: Brown's Reductio: The Basic Argument

Brown (1995) explicates two conditionals that she takes Burge to be committed to:

Q: Necessarily, if x has a thought involving the concept of a natural kind k and x is agnostic about the application conditions of the concept k , then either x is in an environment which contains k , or x is part of a community with the concept k .

(Brown, 1995, p.152)⁴²

and:

R: Necessarily, if x has a thought involving a non-natural kind concept c , and x is agnostic about the application conditions of c , then x is part of a community which has the concept c .

(ibid, pp.154)

The first conditional, Q, results from a generalisation of:

P: Necessarily, if x has a thought involving the concept of water, and x is agnostic about the application conditions of the concept of water, then either x is in an environment that contains H₂O, or x is part of a community which has a term 'water' which applies to something if and only if it is H₂O.

(ibid, p.152)

Brown infers Burge's commitment to this conditional P from a passage from 'Other Bodies' (Burge, 1982) in which Burge considers a subject, Adam, who has no notion of 'H₂O' and who has neither interacted with water or with other people:

What seems incredible to suppose is that Adam, in his relative ignorance and indifference about the nature of water, holds beliefs whose contents involve the notion, even though neither water nor communal cohorts exist.

⁴² If we're to be pedantic, then we'd chastise Brown for her use of the definite article. If there is natural kind k , then (surely) there will potentially be more than one concept of that kind. The conditional should really be as follows:

Q: Necessarily, if x has a thought involving a concept of a natural kind k and x is agnostic about the application conditions of that concept, then either x is in an environment which contains k , or x is part of a community that has a concept of k .

(Burge, 1982, p.116⁴³, quoted in Brown, 1995, pp.150-1)

In explicating the conditionals P and Q, Brown introduces terminology that Burge does not use (see my Chapter 1, §1.2). Brown introduces the term ‘agnostic about the application conditions’. For the antecedent of P to be applicable to a subject, that subject must be ‘agnostic about the application conditions’ of the concept ‘water’. For the antecedent of Q to be applicable to a subject, the subject must be in possession of a natural kind concept of which she is ‘agnostic about the application conditions’. This term ‘agnostic about the application conditions’ proves to be of considerable significance, as we’ll be seeing as we progress through this chapter.

Brown infers Burge’s commitment to the second conditional, R, as following from the reasoning he employs in ‘Individualism and the Mental’:

Imagine that Oscar is agnostic about the application of the word, ‘sofa’. For example, he may apply it firmly and correctly to what we call ‘sofas’, but be unsure about whether it also applies to broad single seat armchairs. According to Burge, if Oscar is part of an English speaking community then, despite his agnosticism, he has thoughts involving the concept sofa. But if, counterfactually, Oscar had been part of a community in which ‘sofa’ applied both to what we call ‘sofas’ and to broad single seat armchairs, then Oscar would have *chofa* thoughts, where the concept of *chofa* applies both to what we call ‘sofas’ and to broad single seat armchairs, ([Burge, 1979], pp.77-83). Now imagine that there are no other speakers in Oscar’s environment. How could Oscar have propositional attitudes involving the concept sofa? Since sofa is not a natural kind concept, Oscar’s environment cannot help him to acquire the concept. There are no other speakers. Nothing seems to show that his attitudes involve the concept sofa as opposed to *chofa* (Brown, 1995, p.153-4)

Brown then combines these two conditionals Q and R to get the following conditional, which is a logical consequent of the conditionals Q and R:

S: Necessarily, if **x** has a thought involving a concept *c*, and **x** is agnostic about the application conditions of *c*, then either **x** is in an environment which contains instances of *c* and *c* is a natural kind concept, or **x** is part of a community which has the concept *c*, whether or not *c* is a natural kind concept.

⁴³ Pg.98 in Burge 2007.

(Brown, 1995, p.154-5)

If I know a priori that, when I am thinking water is wet, I am thinking that water is wet, then I should also know a priori I am deploying a concept 'water'. If I am like Burge's Adam and ignorant with respect to the chemical composition of water, then by Brown's lights I will be 'agnostic' with respect to the application conditions of my concept 'water'. Furthermore, I should be able to know that a priori. Thus, I will be able to know a priori that I meet the antecedent of the conditional S. I should be able to come to know a priori that the conditional S is true through reading Burge, and through reading Brown's paper, and through consideration of the a priori arguments found therein. So, since I know a priori that the antecedent is true with respect to me, I should know a priori that the consequent of S will be true with respect to me. Thus, I'd know a priori that either water exists or I am part of a socio-linguistic community that possesses the concept 'water'. That would be absurd. Therefore, either the ex hypothesi claim that I should know a priori what I am thinking must be false, or Burgean externalism must be false.

5.2: Responding to Brown's Reductio

5.2.1: Doubting Q

Brown's conditional Q is a generalisation of the conditional P. The conditional P is meant to follow directly from the Burge quotation. However, it is not entirely clear that it does. Burge's Adam is ignorant with respect to the nature of water. He doesn't know that it is H₂O. What might follow from the Burge quotation is the conditional:

P*: Necessarily, if x has a thought involving the concept of water, and x does not know that water is H₂O, then either x is in an environment that contains H₂O, or x is part of a community which has a term 'water' which applies to something if and only if it is H₂O.

Brown's conditional P and the conditional P* would amount to the same thing if not knowing that water is H₂O was the same as being 'agnostic' with respect to the application conditions of the concept water. It is not clear that that should be the case, however. Consider the case of *Confident Oscar*. Confident Oscar is ignorant with respect to chemistry. He doesn't know that water is H₂O. Confident Oscar nevertheless feels that he knows exactly what water is. He knows that water is anything that is the *same stuff* as the clear, odourless liquid that he finds all around him, anything that isn't the same stuff as the clear, odourless liquid he finds all around him isn't water. Although

Confident Oscar does not know that water is H_2O , it would (surely) be wrong to say that he is ‘agnostic’ with respect to the concept of water’s application conditions.

The case of Confident Oscar and his ‘water’ concept and the case of the “Oscar”, who we can call *Lonely Oscar*, and his ‘sofa’ concept that Brown discusses in the quotation are markedly different. Confident Oscar has a very firm view of what does and does not fall under his concept. Something falls under the concept if and only if it is the same stuff as what Confident Oscar finds all around him, which, coincidentally, is equivalent to something falling under the concept if and only if it is H_2O . Lonely Oscar does not likewise have a firm view of what does and does not fall under his concept. He is unsure whether “broad single seat armchairs” fall under his concept.

When presented with a clear, odourless liquid, Confident Oscar may in fact be unsure whether to call it ‘water’. This would be due to Confident Oscar’s lack of knowledge regarding the constitution of what he calls ‘water’ and the constitution of the clear, odourless liquid that is presently before him. If Confident Oscar was given a quick crash course on chemistry and told that what he calls ‘water’ is made of H_2O molecules, and then told that the clear, odourless liquid is also made of H_2O molecules, he’d without hesitation say that the clear, odourless liquid is ‘water’. If, on the other hand, Confident Oscar was given a quick crash course on chemistry and told that what he calls ‘water’ is made of H_2O molecules, and then told that the clear, odourless liquid before him is not also made of H_2O molecules, he’d without hesitation say that the clear, odourless liquid is not ‘water’. Lonely Oscar is unsure whether to call “broad single seat armchairs” ‘sofas’ for reasons that are entirely different. No amount of non-conceptual information, regarding sofas and broad singles seat armchair could help Lonely Oscar settle on whether broad single seat armchairs are sofas.

(It might be worth recalling here Burge’s characterisation of Twin Earth (see my Chapter 1, §1.23). It is not entirely clear that within that characterisation that Oscar is unsure about the application conditions of ‘water’ in the same way that Lonely Oscar is unsure about the application conditions of ‘sofa’. It is rather his “ignorance of expert knowledge” – that he does not know that water is H_2O – that, according to Burge, is pertinent. It is not immediately clear that there is anything, in Burge’s characterisation, that would rule out the original Oscar from having the same relationship with ‘water’ that Confident Oscar has.)

To come to the rescue of Brown’s reductio, the Burgean can in fact be shown to be committed to Brown’s conditional P (and hence Q), though it is in a roundabout way, and though it is not for the reasons that Brown gives.

Consider *Hesitant Oscar*. Hesitant Oscar inhabits a world in which as well as water (H_2O), there also exists another natural kind that we can call *pinkish water*. Pinkish water has a chemical composition that is distinct from water (H_2O) – hence it being a distinct natural kind. Superficially speaking however, pinkish water and water are almost identical. Pinkish water looks, tastes, feels etc. almost exactly like water, and, on Hesitant Oscar’s world, is found in all the same places that water

(H₂O) is found. The only superficial difference between water (H₂O) and pinkish water is that under the right lighting conditions pinkish water exhibits a pinkish tinge. Suppose that like Confident Oscar, Hesitant Oscar is ignorant with respect to chemistry. He doesn't know anything about the chemical composition of water (H₂O), or pinkish water. Suppose that Hesitant Oscar believes that, by the standards of his socio-linguistic community, it is definitely correct to refer to water as 'water'. Suppose, however, that Hesitant Oscar is unsure with respect to whether it is correct to refer to pinkish water as 'water'. He is *hesitant* to refer to pinkish water as 'water'. Finally, suppose that, by the standards of Hesitant Oscar's socio-linguistic community, it is incorrect to refer to pinkish water as 'water'. It is correct to refer to something as 'water' if and only if it is water (H₂O). By Burgean lights, this should mean Hesitant Oscar is in possession of a concept of water (H₂O), i.e. a concept such that something falls under it if and only if it is water (H₂O). The case of Hesitant Oscar is far closer to the case that features in the Brown quotation. Hesitant Oscar's reluctance to refer to pinkish water as 'water' mirrors exactly Lonely Oscar's reluctance to refer to "broad single seat armchairs" as 'sofas'.

Suppose now that counterfactually Hesitant Oscar was all alone. He was not part of any socio-linguistic community. By the reasoning invoked by Brown in the quotation (see above), this should mean that Hesitant Oscar would not be in possession of the concept water. This gives strong support for the conditional:

P***: Necessarily, if x has a thought involving the concept of water, and x is agnostic about the application conditions of the concept water, then x is part of a community which has a term 'water' which applies to something if and only if it is H₂O.*

Brown's conditional P follows trivially from the conditional P***. (If F→G, then F→GVH.) From P should follow the generalisation of P that is Brown's conditional Q.*

As well as the truth of Q, along with the truth of R (see next two sections), Brown's reductio also requires that there should be a case in which a subject meets the conditions stipulated in the antecedent of Brown's combined conditional S, and in which they can know a priori that they meet the conditions stipulated in the antecedent of S. That is, there needs to be a case in which a subject is thinking a thought involving a concept of which they are 'agnostic' about the application conditions, and in which they can know a priori that they are thinking a thought involving a concept of which they are 'agnostic' about the application conditions. The considerations of this present section have shown that my being chemically indifferent would not ensure that my thinking that water is wet constitutes such a case. The question then is what would. This is a matter we'll be returning to very shortly (in §5.24).

5.22: *Doubting R I: The Concept 'gilver'*

As well as the conditional:

Q: Necessarily, if x has a thought involving the concept of a natural kind k and x is agnostic about the application conditions of the concept k , then either x is in an environment which contains k , or x is part of a community with the concept k .

Brown's *reductio* also requires that the Burgean be committed to the conditional:

R: Necessarily, if x has a thought involving a non-natural kind concept c , and x is agnostic about the application conditions of c , then x is part of a community which has the concept c .

It has been questioned whether the Burgean needs to be committed to R.

Falvey (2000, p.138) presents the following alleged counterexample to R. Consider the case of *Isolated Iris*. Isolated Iris happens to be in a similar predicament to Lonely Oscar. Iris happens to be in a world in which there are no other speakers. Suppose that Iris has had interactions with the two elements gold and silver but is ignorant with respect to their chemical composition. According to Falvey, the Burgean would concede that by virtue of such interactions Iris will have the concepts *silver* and *gold* – i.e. she'll have a concept that applies to something if and only if it is gold (Au), and a concept that applies to something if and only if it is silver (Ag). Suppose that Iris stipulates to herself that 'gilver' be anything that is either gold or silver. Iris would thus be in possession of the concept *gilver*, which will be a concept such that anything that is either gold (Au) or silver (Ag) would fall under it. The concept *gilver* would be a non-natural kind concept. And according to Falvey, because Iris is indifferent with respect to the chemical composition of gold and silver, that means that she is agnostic with respect to the application conditions of the concepts *gold* and *silver*. That, in turn, means that Iris will be agnostic with respect to the application conditions of the concept *gilver*. Thus, Iris being in possession of the concept *gilver* would be in flat contradiction to Brown's conditional R.

There are a number of things that can be said with respect to this alleged counterexample.

Firstly, Falvey makes the same mistake that Brown makes. As we've seen in the previous section, being indifferent about the chemical composition of a natural kind does not necessarily make you 'agnostic' with respect to the relevant concept's application conditions.

Secondly, it could be argued that even if Iris were agnostic about the application conditions for the concepts *gold* and *silver*, it would not follow that she is agnostic about the application conditions of the concept *gilver*. Iris will have a very good idea of what it takes for something to fall

under the concept *gilver*. She will know that anything that falls under it will be something that either falls under the concept *gold* or falls under the concept *silver*.

Thirdly, the Burgean need not be committed to the idea that a subject can come into possession of the concept of a natural kind through interactions with it in the way that the counterexample requires. The Burgean might have the commitment in *addition* to their Burgean commitments, but they need not.

Fourthly and finally, there is what Brown herself says with respect to this alleged counterexample:

...we can treat the *gilver* case as an exception to (R) without undermining (R)'s original motivation. (R) should not be taken to apply where a subject has a concept and is agnostic about it, where this agnosticism stems wholly from his agnosticism about a component concept in terms of which the first concept is defined. This exclusion clause poses no threat to the *reductio* as a whole, for the *gilver* subject can know a priori that his agnosticism about *gilver* stems wholly from his agnosticism about *gold* and *silver*.
(Brown, 2001, p.218-9)

There is we can suppose a hidden exemption clause in the antecedent of the conditional R, of which a subject can know a priori whether it applies to her, and thereby whether her case is one which is or isn't ruled out. This hidden clause will carry over into the combined conditional S. A more explicit rendering of S would thus be:

S+: Necessarily, if *x* has a thought involving a concept *c*, and *x* is agnostic about the application conditions of *c*, then, unless *c* is a composite of two or more concepts and *x* is agnostic about the application conditions of at least one of those concepts, either *x* is in an environment which contains instances of *c* and *c* is a natural kind concept, or *x* is part of a community which has the concept *c*, whether or not *c* is a natural kind concept.

5.23: *Doubting R II: The Concept 'zofa'*

Consider again Brown's case for the conditional R:

Imagine that Oscar is agnostic about the application of the word, 'sofa'. For example, he may apply it firmly and correctly to what we call 'sofas', but be unsure about whether it also applies to broad single seat armchairs. According to Burge, if Oscar is part of an English speaking community then, despite his agnosticism, he has thoughts involving the

concept sofa. But if, counterfactually, Oscar had been part of a community in which ‘sofa’ applied both to what we call ‘sofas’ and to broad single seat armchairs, then Oscar would have *chofa* thoughts, where the concept of *chofa* applies both to what we call ‘sofas’ and to broad single seat armchairs, ([Burge], 1979, pp.77-83). Now imagine that there are no other speakers in Oscar’s environment. How could Oscar have propositional attitudes involving the concept sofa? Since sofa is not a natural kind concept, Oscar’s environment cannot help him to acquire the concept. There are no other speakers. Nothing seems to show that his attitudes involve the concept sofa as opposed to *chofa* (Brown, 1995, p.153-4)

Brown’s case for the conditional R rests on Burge’s externalism having the entailment that who we are calling *Lonely Oscar* – who Brown calls “Oscar” in the quotation – could not be in possession of the concept ‘sofa’ because he is not part of a socio-linguistic community.

It has been suggested (McLaughlin and Tye, 1998a; Falvey, 2000) that the Burgean need not be committed to this. It has been suggested that on the lips of *Lonely Oscar* ‘sofa’ expresses the concept *zofa*. This concept *zofa* will be distinct from the concept *sofa* – which has an extension that encompasses only sofas – and it will be distinct from the concept *chofa* – which has an extension that encompasses sofas and broad single seat armchairs. The concept *zofa* will apply to sofas – i.e. what we call ‘sofas’. However, there will be no fact of the matter with respect to the question of whether broad single seat armchairs fall under the concept *zofa*. It is indeterminate.⁴⁴

⁴⁴ Depending upon *Lonely Oscar*’s dispositions it may also be the case that there are other types of furniture, as well as armchairs, such that there is no fact of the matter with respect to whether the concept *zofa* applies. This may also be the case with respect to chaise-lounges. If when presented with a chaise-lounge, *Lonely Oscar* was not sure whether to call it a ‘sofa’, then there would similarly be no fact of the matter with respect whether chaise-lounges fall under *Lonely Oscar*’s concept.

Depending upon the dispositions of *Lonely Oscar*, the concept *zofa* may or may not apply to other types of furniture also. For instance, it may or may not apply to chaise-lounges. It doing so or it not doing is dependent upon *Lonely Oscar*’s dispositions. If *Lonely Oscar*, when presented with a chaise-lounge, was adamant that what was before him could not be a ‘sofa’, then chaise-lounges would not fall under the concept *zofa*. If, on the other hand, *Lonely Oscar* was adamant that what was before him was most definitely a ‘sofa’, then chaise-lounges would fall under the concept *zofa*.

The concept *zofa* would be similar to a vague concept, such as the concept *tall for an adult male*. A vague concept such as *tall for an adult male* does do *some* of the work of carving up reality that a concept is obliged to do, but falls short when it comes to a certain subset. Any man who is shorter than 5’7” does not fall under the concept *tall for an adult male*. Any man who is taller than 6’2” does. There will however be a range of heights somewhere in between 5’7” and 6’2” such that for a man whose height is within that range, there will be no fact of the matter with respect to the question of whether or not that man falls under the concept *tall for an adult*. Similarly, the concept *zofa* would do *some* of the work of carving up reality that a concept is obliged to do, but would fall short when it comes to a certain subset. That subset would comprise armchairs as well as possibly other things, such as chaise-lounges, depending upon *Lonely Oscar*’s dispositions.

Brown has in later works (Brown, 2001; 2004) conceded that Lonely Oscar *could* be in possession of a concept, that he *could* be in possession of the concept *zofa*. However, she still maintains that this does not mean the Burgean is permitted to jettison R. For the antecedent of R to be applicable to a subject, the subject must be ‘agnostic’ with respect to her concept’s application conditions. To be ‘agnostic’ with respect to a concept’s application conditions – as Brown in her later work clarifies – requires two things. Firstly, it must be the case there is at least one type of thing such that the subject is unsure whether her concept applies to that type of thing. Secondly, the subject’s concept must be such that there is a fact of the matter with respect to whether the type, or types, of thing that the subject is unsure about fall under the concept. So, for a subject to be agnostic with respect to the application conditions of the concept she expresses by ‘sofa’, not only must she be unsure whether the concept expressed by ‘sofa’ applies to broad single seat armchairs, it must also be the case that there is a fact of the matter with respect to the question of whether it does or not. It must either be the case that ‘sofa’ *does* apply to broad single seat armchairs, or it must be the case that the word ‘sofa’ *does not* apply to broad single seat armchairs.⁴⁵

5.24: *How can you Know A Priori that you’re ‘Agnostic’?*

It may be questionable whether we can have concepts such as the concept *zofa*, which have the sort of indefinite application conditions that we are supposing that the concept *zofa* has. Let us, nevertheless, suppose that such concepts do exist.

⁴⁵ There is another option open to the would-be defender of Brown’s reductio. Rather than contend that a subject being ‘agnostic’ about a concept’s application requires that that concept have definite application conditions, a would-be defender could suggest that a clause could be added to Brown’s conditionals. R would thus be replaced with:

R⁺: Necessarily, if **x** has a thought involving a non-natural kind concept *c*, and **x** is agnostic about the application conditions of *c*, and *c* has definite application conditions, then **x** is part of a community which has the concept *c*.

The combined conditional S would become:

S⁺: Necessarily, if **x** has a thought involving a concept *c*, and **x** is agnostic about the application conditions of *c*, and *c* has definite application conditions, then either **x** is in an environment which contains instances of *c* and *c* is a natural kind concept, or **x** is part of a community which has the concept *c*, whether or not *c* is a natural kind concept.

The difference between this alternative proposal and the proposal that we accept Brown’s amended/clarified definition of ‘agnostic’ are superficial. All of what I go on to discuss in the next section will apply *mutatis mutandis* to the alternative proposal. (If Clive could know a priori that bureaus are a particular sort of furniture, then he would be able to know a priori that the antecedent of the conditional S⁺ is applicable to him.)

Let us also accept Brown's amended/clarified definition of 'agnostic'. A subject is 'agnostic' with respect to a concept if (i) they are unsure about the concept's application conditions, and (ii) the concept does have definite application conditions.

Consider again the conditional that forms the heart of Brown's reductio:

S: Necessarily, if x has a thought involving a concept c , and x is agnostic about the application conditions of c , then either x is in an environment which contains instances of c and c is a natural kind concept, or x is part of a community which has the concept c , whether or not c is a natural kind concept.

For Brown's reductio to work there must be a case in which the antecedent of S applies to a subject, *and* in which that subject knows a priori that the Brown's conditional S applies to her. There needs to be a case in which a subject is thinking a thought involving a concept of which she is agnostic about the application conditions, and in which she can know a priori that she is agnostic about that concept's application conditions. Suppose I am thinking a thought involving a concept c . Ex hypothesi, I can know a priori that I am thinking such a thought. However, the reductio also requires that I should also know a priori that I am agnostic with respect to the concept c 's application conditions. This is not merely a matter of knowing a priori that I am unsure about the application conditions of c . It is also a matter of knowing a priori that c does in fact have determinate application conditions.

Brown suggests the following case:

Suppose that a subject has a partial understanding of bureau [the concept that is expressed by 'bureau'] and, in virtue of this, knows that it applies to a type of piece of furniture, but not what type. This subject can know a priori that he is unsure what type of furniture bureau applies to. It might also be argued that he can know a priori that it applies to some particular type of furniture. According to a traditional view, subjects can gain a priori knowledge of how a concept applies in virtue of understanding it. For example, a subject who understands the concept of a toothbrush can know a priori that it is the concept of an artefact, and someone who understands the concept of a dress can know a priori that it is the concept of a piece of clothing. Now, our subject is stipulated to know that bureau applies to a particular type of furniture in virtue of his partial understanding of it. Thus, someone might argue that his knowledge that bureau applies to a particular type of furniture constitutes a priori knowledge. But given that he also knows a priori that he is unsure what type of furniture bureau applies to, he can have a priori knowledge that he is agnostic about [the application conditions of] bureau.

(Brown, 2001, pp.219-220)

Such a subject, let's call him Clive, will know a priori that he meets the conditions of the antecedent of the conditional S, *as long as* he is able to form the belief that bureaus are a type of furniture – i.e. the concept 'bureau' has definite application conditions – via a priori means, and *as long as* that belief constitutes knowledge. Ex hypothesi, when Clive thinks a thought involving the concept, he should be able to know a priori that he is doing so. If the conditional S is true, then Clive could – if he were to attend philosophy classes – come to know a priori that the conditional S is true via consideration of Burge and Brown's arguments. Thus, Clive will be able to know a priori that either he belongs to a socio-linguistic community or his concept *bureau* corresponds to a natural kind that exists in his environment. That'd be absurd. Thus, we will have our reductio.

Whether we do in fact have our reductio will be dependent upon whether Clive would be able to come to believe that his concept applies to a particular type of furniture via a priori means, and upon whether that belief would be knowledge constituting. The success of Brown's reductio hinges upon whether Clive could know a priori that bureaus are a particular type of furniture.

The relevant passages of Brown's paper (Brown, 2001) are uncharacteristically unclear. But what (I believe) she may have had in mind is that there is a *mechanism* by which Clive could come to believe that his concept applies to a particular type of furniture that is either the same as or is similar to the mechanism that is employed in acts of *critical reasoning*. Suppose that Clive comes to belief that his friend Frank believes that bureaus are hopelessly antiquated. Clive would be able to reason that therefore Frank believes that a particular type of furniture is hopelessly antiquated. Clive would come, via his act of reasoning, to the *new* belief that Frank believes that a particular type of furniture is hopelessly antiquated. The way in which Clive comes to his new belief would (presumptively) be entirely a priori. Brown could be envisaging that something similar is going on when Clive comes to believe that his concept applies to a particular type of furniture.

So much for the belief, let's now consider whether the belief could count as knowledge.

Brown talks of "some sort of transcendental argument" (ibid., p.222) that would aim to demonstrate that we must admit the possibility of a priori knowledge of analytic truths, if we are to maintain that we can come to know things through the process of *critical reasoning*. The following is what I believe she may have had in mind. Suppose that Clive were to come to belief that there is a particular type of furniture that Frank thinks is hopelessly antiquated, in the way that I described, i.e. through his reasoning that because Frank thinks that bureaus are hopelessly antiquated, there is therefore a particular type of furniture that Frank thinks is hopelessly antiquated. That true belief would surely qualify as knowledge. Through his act of reasoning, Clive has come to *know* that there is a type of furniture that Frank thinks is hopelessly antiquated. Now, if Clive's belief that bureaus are a

particular type of furniture is gotten by the same – or a similar – mechanism, then that would suggest that that true belief must also constitute knowledge.

This “transcendental argument” requires a number of things.

Firstly, you should recall the distinction between *local reliability* and *global reliability* from Chapter 3 (§3.22). When a subject forms a true belief that <p> via a method of belief formation, that belief formation method will be *locally reliable* just in case there are no relevant alternatives in which <p> is not the case, but the method would still lead the subject to form a false belief. The belief formation method will be *globally reliable* just in case it is and would be locally reliable with respect to all propositions that are sufficiently similar to <p>. The argument requires that global reliability and not just local reliability is necessary for knowledge. If just local reliability were necessary for knowledge, then that could allow us to say that Clive’s belief that there is a particular type of furniture that Frank thinks is hopelessly antiquated counts as knowledge, without us also having to admit that his belief that bureaus are a particular type of furniture also counts as knowledge. The manner in which both beliefs are formed is the same. It just so happens – we could say – that that belief forming process is locally reliable in the case of the first belief, but not in the case of the second – i.e. it is not globally reliable.⁴⁶

Secondly, the argument requires that the process by which Clive acquires his belief that Frank thinks a particular type of furniture is hopelessly antiquated be, in an appropriate sense, *the same as* the process by which he acquires his belief that bureaus are a particular type of furniture. If the process by which the first belief is formed is *not* the same as the process by which the second belief is formed, then this would allow us to say that the first belief qualifies as knowledge, without us having to admit that the second belief qualifies as knowledge. And, we could do so whilst conceding that global reliability is required for knowledge. If the second belief doesn’t qualify as knowledge, then ipso facto the process by which the belief is formed is neither locally nor globally reliable. However, that won’t mean the process by which the first belief is formed is therefore not globally reliable, as it would, *ex hypothesi*, be a different process.

Thirdly and finally, you should also recall from Chapter 2 what Brown calls the *generality problem*. A method of belief formation will be globally reliable just in case it is locally reliable with respect to a range of propositions that are similar to <p>. The generality problem is the problem of what determines which propositions are similar to <p>. The argument requires a solution to the generality problem under which the proposition that Frank thinks a particular type of furniture is hopelessly antiquated and the proposition that bureaus are a particular type of furniture would be deemed sufficiently similar.

⁴⁶ Brown in a footnote (Brown, 2001, p.222, fn.7) does state that the soundness of her argument could depend on upon the issue of whether knowledge requires global or local reliability.

The argument can be written more formally:

Manifest Premise: Through an act of critical reasoning, Clive can come to know that Frank thinks a particular type of furniture is hopelessly antiquated.

Observation 1: Through a similar act of critical reasoning Clive can form the true belief that bureaus are a particular type of furniture.

Assumption 1: The act of reasoning that would lead Clive to form the true belief that bureaus are a particular type of furniture would be a priori.

Assumption 2: In order for a belief formation method to confer knowledge, it must be the case that that method is globally reliable.

Assumption 3: The act of critical reasoning that would lead to Clive's knowledge that Frank thinks a particular type of furniture is hopelessly antiquated, and the act of critical reasoning that would lead to Clive's true belief that bureaus are a particular type of furniture would be tokens of the same method of belief formation

Assumption 4: The correct solution to the generality problem would rule that the proposition that Franks thinks a particular type of furniture is hopelessly antiquated and the proposition that bureaus are a particular type of furniture are 'sufficiently' similar.

Transcendental Requisite: If Clive's true belief that bureaus are a particular type of furniture would not constitute knowledge, then Clive could not come to know that Franks thinks a particular type of furniture is hopelessly antiquated through an act of critical reasoning. (From Assumptions 2-4.)

Transcendental Conclusion: Clive's true belief that bureaus are a particular type of furniture would constitute knowledge. (From Manifest Premise and Transcendental Requisite.)

Further Conclusion: Clive would know a priori that bureaus are a particular type of furniture. (From Assumption 1 and Transcendental Conclusion.)

If the argument – whether it be what Brown had in mind or not – is sound, then we could, as the reductio requires, suppose that subjects could know a priori that they are 'agnostic' – per Brown's amended definition – with respect to a concept's application conditions. The argument being sound would mean that Clive would be able to know that he is agnostic with respect to the application conditions of his 'bureau' concept. Thus, it will be possible for Clive to know a priori that he meets the conditions stipulated in the antecedent of Brown's conditional:

S: Necessarily, if x has a thought involving a concept c , and x is agnostic about the application conditions of c , then either x is in an environment which contains instances of c and c is a natural

kind concept, or x is part of a community which has the concept c , whether or not c is a natural kind concept.

Thus, Clive could come to know a priori the consequent of S that either he inhabits a world that contains a natural kind that instances of which are called ‘bureaus’, or he belongs to a socio-linguistic community that has the concept ‘bureau’.

The later Brown, i.e. the Brown of *Knowledge and Anti-Individualism*, claims that it is “hard to see” how a subject could know a priori that they are agnostic about a concept’s application conditions (Brown, 2004, pp.298-301). Brown argues that if a subject is unsure about the application conditions of her concept, then it could be the case that the concept does not have definite application conditions, and thus would not be agnostic about the concept’s application conditions. So, since the subject cannot rule out that possibility a priori, it is surely the case that she cannot know a priori that she is agnostic about the concept’s application conditions.

If the argument I characterise here is sound, then the later Brown must be wrong. A subject *could* know a priori that she is agnostic with respect to her concept’s application conditions. Such a subject perhaps wouldn’t be able to rule out a priori the possibility that that concept could have indeterminate application conditions. But, if the argument is sound, then such possibilities won’t always constitute relevant alternatives.

Clive would not be able to rule out a counterfactual scenario in which the word ‘bureau’ as used by Clive’s socio-linguistic peers doesn’t have definite application conditions. However, if the earlier Brown’s argument – i.e. the argument I have inferred from Brown – is sound then *a fortiori* such a counterfactual scenario cannot constitute a relevant alternative. Because if it did then Clive’s true belief that bureaus are a type of furniture could not constitute knowledge – the a priori method by which Clive acquired his belief could not have involved him ruling out such a counterfactual scenario. Clive’s true belief will constitute knowledge if the argument is sound. *Ipsa facto*, such a counterfactual scenario will not constitute a relevant alternative.

For the earlier Brown’s argument to be sound the four assumptions I’ve explicated must *all* be true. I won’t here try to make a case for or against assumptions 2-4, which all essentially pertain to wider epistemological matters. I will, however, comment on **Assumption 1**, which is that the act of critical reasoning that leads Clive to form the belief that bureaus are a particular type of furniture would be a priori. I am not entirely convinced that the way in which Clive would come to the belief that bureaus are a particular type of furniture and to his belief that his friend Frank thinks that there is a particular type of furniture would be entirely a priori. In both cases Clive would essentially be

making judgements about how his socio-linguistic community – about how they use the word ‘bureau’ – which should surely be a posteriori.⁴⁷

Brown’s reductio requires a case in which someone is in possession of a concept of which they are ‘agnostic’ about the application conditions and know a priori that they are ‘agnostic’ about the application conditions. This, problematically, means that they must know a priori that their concept has definite application conditions. We’ve been considering the case that Brown suggests could meet these desiderata. I have tried to reverse engineer an argument that would demonstrate that the case that Brown suggests would indeed meet the desiderata. It is not entirely clear that the argument I have come up with – whether it is what Brown had in mind or not – is sound. Thus, it is questionable whether we have a case in which someone is in possession of a concept of which they are ‘agnostic’ about the application conditions and know a priori that they are ‘agnostic’ about the application conditions.

⁴⁷ Brown, incidentally, worries that her subject being in possession of a priori knowledge that bureaus are a particular type of furniture would be precluded by Quinean considerations with respect to the analytic/synthetic distinction (Quine, 1951). The subject knowing a priori that bureaus are a particular type of furniture would constitute an instance of a priori knowledge of an “analytic truth” (Brown, 2001, p.220). Quine’s famous deliberations, Brown worries, entail a denial that there can ever be an instance of a priori knowledge of an analytic truth. She also worries that the possibility would also be precluded by the arguments that Burge himself puts forward in his (1986) ‘Intellectual Norms and the Foundations of Mind’, as those arguments similarly entail a denial that there can ever be an instance of a priori knowledge of an analytic truth. It is not clear whether – or to what extent – my worry corresponds to such worries.

Conclusion to Chapters 3-5

In the preceding three chapters I have been assessing some of the most prominent strands in the debate that lasted from the late eighties to the mid noughties over whether content externalism is compatible with the claim that we should have an a priori access to the contents of our thoughts.

In Chapter 3, I considered what is known as the *discrimination argument*. Proponents of the *discrimination argument* hold that externalism allows for cases in which a subject forms a true – guaranteed-to-be-true (even) – belief about what she is thinking, but in which there will be *relevant alternatives* – i.e. pertinent counterfactual scenarios – in which things would seem to the subject exactly as they actually are, and in which the subject is not thinking what she is actually thinking. The most famous of those cases is the case of *Slow-Switching Sally*.

There are two ways of countering the discrimination argument. The first is to insist that no such case would or could actually arise. The problem with this is that, even if the case of Slow-Switching Sally is dismissed for being too contrived, there seem to be slow-switching cases that are more plausible such as the ones we encountered throughout the chapter – such as the cases of Biff and Griff.

The second way of countering the discrimination argument is to adopt the *reliabilist* approach. If content externalism were true, then there *would* be cases in which a subject forms a true belief about what she is thinking, and in which there will be relevant alternatives in which things would seem to the subject exactly as they actually are, and in which the subject is not thinking what she is actually thinking. In such cases, however, the existence of such relevant alternatives does not disqualify the subject's true belief from counting as knowledge, as within those relevant alternatives the subject would not form what would be the false belief that she is thinking what it is that she is actually thinking.

Brown's Wasp-Phobic Wendy case *might* pose a problem for this second approach – but could only do so if a number of conditions can be satisfied.

In chapter 4, I considered Boghossian's reductio argument. Boghossian argues that if content externalism is true, and it is true that we can know a priori what the contents of our thoughts are, then it follows absurdly that we could know a priori that water exists. If I know a priori that I am thinking that water is wet, I can deduce that I am not a denizen of Dry Earth, and thus I can come to know a priori that I am not a denizen of Dry Earth, and thus come to know a priori that water exists.

Boghossian's reductio rests on three claims.

Firstly, it must be the case that if we have a pair of (qualitatively) identical subjects it cannot be the case that one is deploying an atomic concept whilst the other is deploying a non-atomic concept.

Other than making cryptic allusions to the LOTH, Boghossian does not adequately explain why, if my 'water' concept is indeed atomic, it follows from that that my atom-for-atom Dry-Earthling duplicate's equivalent concept cannot therefore be non-atomic.

This could potentially be remedied if a convincing case for Boghossian's thesis could be provided. For instance, I have suggested that my Dry Earthling counterpart could not be said to *grasp*, and therefore possess, his putative 'water' concept if it were non-atomic, as, as far as my counterpart would be concerned, the concept would be atomic.

The second claim is that a Dry Earthling 'water' concept could not be atomic.

Boghossian does not exhaustively dispel the possible ways in which my Dry Earthling counterpart could still be said to be in possession of an atomic concept, even though on Dry Earth there would be no equivalent of water that could determine that concept's application conditions. For instance, Boghossian doesn't say anything about why the concept could not be *necessarily vacuous*.

The third claim is that Putnam's 'Twin Earth' thought experiment requires that Oscar's 'water' concept be atomic.

Boghossian does not adequately explain why this needs to be the case.

In Chapter 5 we saw how Brown derives the following conditional from the writings of Burge:

Q: Necessarily, if x has a thought involving the concept of a natural kind k and x is agnostic about the application conditions of the concept k , then either x is in an environment which contains k , or x is part of a community with the concept k .

and:

R: Necessarily, if x has a thought involving a non-natural kind concept c , and x is agnostic about the application conditions of c , then x is part of a community which has the concept c .

By combining these conditionals, Brown deduces that a Burgean must be committed to:

S: Necessarily, if x has a thought involving a concept c , and x is agnostic about the application conditions of c , then either x is in an environment which contains instances of c and c is a natural kind concept, or x is part of a community which has the concept c , whether or not c is a natural kind concept.

Brown argues that if Burgean externalism were true, a subject could know a priori that the antecedent of S applies to her. Thus, the subject could come, via consideration of the relevant philosophical arguments, to know a priori that the consequent of S also applies to her, which would be absurd.

It has been doubted whether the Burgean needs to be committed to the conditional R. It has been suggested that an isolated subject could be in possession of a concept even if she is unsure about that concept's application conditions, as long as it is supposed that it is possible for a concept to have indefinite application conditions. If an isolated subject is in possession of a concept and she is unsure whether that concept applies to a particular set of things, then it will be the case that there is no fact of the matter whether that concept applies to that set of things.

In response, Brown has conceded that there could be concepts with indefinite application conditions, but has contended that in order for a subject to be 'agnostic' with respect to a concept it must be the case that that subject is unsure about that concept's application conditions, *and* it must also be the case that the concept does have definite application conditions, there cannot be a set of things for which there is no fact of the matter whether the concept applies.

If there are concepts with indefinite application conditions, then in order for a subject to know a priori that she meets the conditions stipulated in the antecedent of Brown's conditional S she must be able to know a priori that her concept c has definite application conditions. Brown has presented a case in which a subject could meet this requirement, but only if the 'transcendental' argument that I have inferred from an obscure passage in one of Brown's papers is sound. It is not clear whether that argument is indeed sound.

In sum, the previous three chapters have all respectively considered prominent arguments for incompatibilism. None of these three arguments are entirely convincing. They contain too many potentially-pluggable-but-as-of-yet-unplugged dialectical holes.

Our enquiry could end here. Content externalism being incompatible with the claim we should know a priori what the contents of our thoughts are may pose a threat to content externalism. But we have no compelling reason to suppose that the two things are indeed incompatible. The arguments that we've considered are *potentially* salvageable. As-of-yet-unplugged dialectical holes could potentially be plugged. Nevertheless, we are possibly safe to conclude that we don't *as of yet*

have conclusive arguments for supposing that content externalism and the a priori claim are incompatible.

There is, however, the matter of why it should be supposed that content externalism and claims that we should be able to know a priori what we are thinking being incompatible would pose a threat to content externalism. If it is indeed the case that the two things can't both be true, then why should it be problematic to suppose that it is the latter and not the former that must be false? In the final chapter of the thesis – Chapter 7 – I essentially consider the feasibility of denying that we should be able to know a priori what we are thinking.

Also, Boghossian's reductio can (I believe) be given a new lease of life. In the next and penultimate chapter – Chapter 6 – I advance my own reductio argument that concerns Kripkean externalism that takes inspiration from Boghossian's reductio.

Chapter 6: Inferring from Mind to World III: An Additional Reductio

6.0: Introduction to Chapter 6

In the previous three chapters I assessed the arguments of others for the incompatibility of content externalism and the claim that we should be able to know a priori what we are presently thinking. In this present chapter, I present my own original argument for the incompatibility of Kripkean externalism and the claim that we should be able to know a priori what we are presently thinking. This argument is a reductio similar to those advanced by Boghossian and Brown. It takes inspiration from Boghossian's reductio in particular, but it trades the notions of conceptual atomicity and non-atomicity that fuel Boghossian's reductio for the Kripkean notions of *direct* and *descriptive* reference. Boghossian's reductio banks on the contention that in the event 'water' does not refer, 'water' could express neither an atomic nor a non-atomic concept, and therefore could not express a concept at all. Analogously, the reductio I propose banks on it being the case that in the event 'Aristotle' does not refer, 'Aristotle' could neither refer directly nor be shorthand for a definite description.

6.1: The Basic Argument

According to Kripkean theory, when I utter the name 'Aristotle' it should be the case that I am not using it as shorthand for some definite description that Aristotle uniquely satisfies. My reference is not mediated through the conceptual lens that is a definite description. Instead, I refer *directly* to Aristotle (see Kripke, 1982; Salmon, 1982; and Kaplan, 1989). Due to Kripke's commitment to Millianism (see my Chapter 1, §1.3), it should also be the case that if there were no 'Aristotle', then the thought I'd express 'Aristotle was an alcoholic' (say) could not have a content. The thought owes its very content to Aristotle. Thus, although even though I may believe that I am thinking a thought

that I'd express 'Aristotle was an alcoholic', if there were no 'Aristotle', I could not be thinking any such thought. And thus, if it is to be supposed that I know a priori that I am thinking the thought 'Aristotle was an alcoholic', then it should follow that I would be able to know a priori that there was an 'Aristotle', that Aristotle existed. This would be absurd.

6.2: Could 'Aristotle' be Shorthand for a Definite Description in the Event that he did not Exist?

My reductio requires it to be the case 'Aristotle was an alcoholic' *could not* express a thought if Aristotle had never existed. Thus, if it can be shown that 'Aristotle was an alcoholic' *could* express a thought, in the event that Aristotle never existed, my reductio will have failed.

One way in which this might be done is to argue that 'Aristotle' on my lips would, in the event that Aristotle never existed, be shorthand for some appropriate definite description. When I utter 'Aristotle' what I really mean, if it is indeed that case that Aristotle never existed, is 'the teacher of Alexander' (say), or 'the last great philosopher of antiquity' (perhaps). The thought 'Aristotle was an alcoholic' thus becomes 'The teacher of Alexander was an alcoholic', or 'The last great philosopher of antiquity was an alcoholic', and thus acquires a new Aristotle-independent content.

This is precisely what has been suggested by Ludlow (Ludlow, 2003). Ludlow suggests that we can mix Russellian and Kripkean semantic theory for a sort of best of both worlds approach. Kripke has shown that Russellian semantic theory, which holds that names are shorthand for definite descriptions, is in irreconcilable conflict with intuitions, made bare by Kripke, regarding the workings of proper names (Russell, 1905; Kripke, 1982). Russellian theory, nevertheless, apparently succeeds where Kripkean theory does not, in that it gives us a solution to the problem of *empty names* – names that fail to refer. Ludlow's suggestion is that we concede that Kripke is basically right in the case of names that do not fail to refer, but for names that do fail to refer we return to Russell. Even though *Naming and Necessity* rips up the old Russellian rule book in spectacular fashion, we still need to defer to its torn fragments in the problematic case of empty names, or as Häggqvist and Wikforss, in their critique of Ludlow (Häggqvist and Wikforss, 2007, pp.379-382), deliciously put it:

Russell, as it were, gets to clean up after Kripke's party.
(*ibid.*, p.379)

In the counterfactual scenario in which Aristotle did not exist, the name 'Aristotle' on my lips would be an empty name – it would be a name that fails to refer. Ludlow's suggestion is that because of this the name 'Aristotle', on my lips, should be ascribed the semantics that Russellian semantic theory would ascribe it. That is, the name 'Aristotle', on my lips, should be considered shorthand for some

definite description that Aristotle is believed to uniquely satisfy. Even though, in the actual case it is not. In the actual case, in which Aristotle exists, the name ‘Aristotle’, on my lips, will refer directly.

The problem with this suggestion is that the role that *speaker intentions* play in communication would be compromised. If I’m going to use a name as shorthand for some definite description, then I should do so with the intention to use that name as shorthand for a definite description. Likewise, if I’m going to use a name to refer directly, then it should be the case that I would do so with the intention to do so.

Under Ludlow’s proposal it seems that whether I use a name as shorthand for a definite description, or use it to refer directly to someone or something, would not depend upon my intentions, but would depend upon factors pertaining to my external environment – it will depend upon whether or not something in my external environment exists.

Ludlow contends that in both the actual and counterfactual scenario, I *do* exercise my semantic intentions to their full extent when I utter the sentence ‘Aristotle was an alcoholic’ in that in the both the actual case and in the counterfactual scenario, my semantic intentions are to express the thought that I am currently thinking. It just so happens that that thought is different in the counterfactual scenario to what it actually is. Not only is it different in terms of its content. It is also different in terms of its logical form. In the actual situation the thought is singular, whereas, in the counterfactual scenario the thought is descriptive.

Häggqvist and Wikforss complain that this is no better than the position that Ludlow himself rejects that he calls “bald externalism” (Ludlow, 2003, p.405), and that is the view that:

...the logical form of an utterance such as ‘Socrates was a philosopher’ depends entirely on whether Socrates exists.

(Häggqvist and Wikforss, 2007, p.380)

Häggqvist and Wikforss complain that what Ludlow proposes:

...can hardly be characterised as a situation in which the logical form of the utterance is *determined* by the individual’s intentions. Rather, the logical form of the thought *as well as* of the utterance is determined by the speaker’s environment – this is why there is a connection between the two, and not because there is a prior intention determining the logical form of the utterance. Hence this position seems to be no better off than “bald externalism” when it comes to giving intentions a semantic role.

(*ibid.*, Häggqvist and Wikforss’ emphasis)

Häggqvist and Wikforss (ibid., p.381) consider another suggestion for what role my semantic intentions could play. My intention to use ‘Aristotle’ to refer directly constitutes a *necessary condition* for ‘Aristotle’ referring directly. It is just that my having that intention is not also a *sufficient condition* for ‘Aristotle’ referring directly. If it is my intention to use ‘Aristotle’ as shorthand for some definite description, then I would be doing so both in the actual situation, and in the counterfactual scenario in which ‘Aristotle’ does not refer. If it is not my intention to use ‘Aristotle’ as shorthand for some definite description, if instead it is my intention to refer directly to Aristotle, then it will be the case that I will be referring directly to Aristotle *only* in the actual situation in which Aristotle existed. I will *not* be referring directly to Aristotle in the counterfactual scenario in which he did not exist. In the counterfactual scenario, I would be using ‘Aristotle’ as shorthand for some definite description, even though that would be contrary to my intentions.

Häggqvist and Wikforss complain that this again is no advancement on ‘bald externalism’, that it:

...seems only marginally less mysterious, if at all, than the position Ludlow labels bald externalism.

(ibid.)

There is also a third suggestion for how it could be that my intentions could still feature in my utterance that is mooted by Häggqvist and Wikforss. The suggestion is that my intentions are “disjunctive”. I use ‘Aristotle’ with the intention that it refer directly to Aristotle as long as he did exist but be shorthand for some definite description if it so happens that he did not.

Häggqvist and Wikforss’s complain that this third proposal:

...clearly requires ordinary speakers to have intentions concerning *semantics*, involving semantic theories, and attributing such meta-semantic intentions (i.e. intentions concerning foundational semantics) to ordinary speakers is utterly implausible – even more so when we consider that we may be asked to attribute them to speakers of, say, 11th century Norse.

(ibid., p.382)

We could here take issue with Häggqvist and Wikforss’s dismissal of the linguistic abilities of “ordinary speakers” – and of speakers of “11th century Norse” in particular. Ordinary speakers – including speakers of 11th century Norse – should have at least a *tacit* grasp of the distinction between referring to something directly and using a term as shorthand for some definite description, which those of us who have read *Naming and Necessity* have been made more aware of. Thus, ordinary

speakers should be able to *effectively* have the disjunctive intention that a name have the semantic properties that Kripkean semantic theory states that it would have in case that the named existed, and have the semantic properties that Russellian semantic theory states that it would have in case that the named did not exist.

In any case, there are independent reasons for supposing that this third proposal won't work.

Firstly, if I, or indeed anybody, can possess the disjunctive intention that 'Aristotle' refer directly to Aristotle in the case that he existed but be shorthand for a definite description in the case that he did not, then I should be able to have the *non*-disjunctive and *simpler* intention that 'Aristotle' should refer directly to Aristotle.

Secondly, there is no good reason why I or anyone should use 'Aristotle' with the more complex disjunctive intention.⁴⁸ The *raison d'être* of directly referring expressions and the *raison d'être* of definite descriptions coincide. Directly referring expressions and definite descriptions are both used by the speaker to identify a unique individual – or object, or natural kind (even – see Chapter 1, §1.32) – that the speaker believes exists, and that the speaker wishes to say something about. If Aristotle did not exist, then that is just as much a problem for the speaker who wishes to refer to Aristotle via the means of a definite description as it is for the speaker who wishes to refer to Aristotle directly. If Aristotle did not exist, both speakers would have failed to do what they've set out to do in their respective speech acts. It is unlikely that a speaker, who uses 'Aristotle' to refer directly, would want – as some sort of consolation prize – to be interpreted as using 'Aristotle' as shorthand for a definite description that would pick out no existing individual, if it so happens 'Aristotle' fails to refer.

Relatedly but more generally, an objection to my *reductio* might be that for most speakers, their intention in using a referring expression is simply to refer. They may not care whether that reference is achieved through their referring expression being shorthand for some definite description or through it being directly referential. Indeed, they may not even be aware that there is a difference. For our purposes, however, we need only demonstrate that there is at least one person, namely me, who would be able to acquire a priori knowledge that Aristotle exists to get us to the absurdity that my *reductio* requires. This means that, for our present purposes, we can in fact restrict ourselves to the case of someone who has read *Naming and Necessity*, and who is aware of the difference. We need only contend that *I* have a semantic intention that *my* use of 'Aristotle' refer directly to Aristotle, and that because of that 'Aristotle' on *my* lips cannot possibly be shorthand for some definite description.

⁴⁸ In words more couched in Russellian theory (Russell, 1905), it is unclear why someone would want to be understood as stating that Aristotle was an alcoholic, in the event that Aristotle existed, but stating that there exists someone who is the one and only teacher of Alexander and who was an alcoholic, in the event that Aristotle did not exist.

Alternatively, it may be argued that everyone is capable of having something that can be correctly described as an ‘intention to refer directly’ and is also capable of having something that can be correctly described as an ‘intention to refer descriptively’. It is just that some of us have been made more aware of such tacit intentions through reading things like *Naming and Necessity*.

Our reductio is maintained regardless, as it is I – someone who has read *Naming and Necessity* – who is its intended victim.

In sum, we cannot suppose that ‘Aristotle’ could be shorthand for a definite description in the event that Aristotle had never existed because if I’m going to use a name to refer directly, then I do so with the intention to refer directly, and if I were going to use a name as a shorthand for some definite description then I would do so with the intention to use that name as a shorthand for a definite description. In the event Aristotle never existed, I could not be using ‘Aristotle’ as a shorthand for some definite description, as that would run contrary to my semantic intentions.

6.3: Could ‘Aristotle’ Still Refer Directly even if Aristotle did not Exist (if You’re a Millian)?

The previous section should have established that the name ‘Aristotle’ could not be shorthand for some definite description, in the event that Aristotle did not exist. A question that could still be asked is: is there *any* other way in which ‘Aristotle was an alcoholic’ could still express a thought if it were the case that Aristotle did not exist? I venture that from the perspective of a Kripkean externalist the answer to that question must be no. This would, mean that as the reductio requires, I would be able to deduce that Aristotle exists based on my ex hypothesi a priori knowledge that I am thinking the thought ‘Aristotle was an alcoholic’. To be a Kripkean externalist is by definition – i.e. by my definition (see Chapter 1, §1.31) – to be a Millian. A direct consequence of that is that the name ‘Aristotle’ must refer to something, otherwise there can be no thought expressed by ‘Aristotle was an alcoholic’.

What I am venturing here requires some qualification. It might be possible, even from the Millian perspective, that ‘Aristotle’ could still refer in the event that there was no *flesh-and-blood* individual who bore the name.

Let’s consider *four* ways in which Aristotle might not have existed, but in which the name ‘Aristotle’ could still be in use.

Firstly, imagine a scenario in which ‘Aristotle’ was some sort of fictional character created by medieval Arab scholars, and the relevant Kripkean socio-linguistic chain that connects our present day use of the name ‘Aristotle’ connects us to these Arab scholars, and to their fictional creation.

Secondly, imagine a scenario in which the residents of fourth century BC Athens all suffered from a weirdly coincidental mass psychosis that caused them all to see a man who talked a lot about final causes, and who they all called ‘Aristotle’, but who was in fact an illusion. Suppose that the

relevant socio-linguistic chain connects our present-day use of the name ‘Aristotle’ to these psychotic fourth-century Athenians.

Thirdly, imagine a scenario in which the universe is only a few seconds old – all our memories, the fossil record, etc. are fabrications.

Fourthly and finally, consider the following scenario. When the Great Library was set alight two Alexandrian brothers named ‘Aris’ and ‘Stotle’ braved the flames, rescuing a fair number of scrolls. The brothers believed that the scrolls were all by the same author. They were unsure who that author was, or what he – they assumed that he was a ‘he’ – was called. So, they decided to name the mystery author after themselves. They decided to name him ‘Aristotle’. They thus essentially engaged in a Kripkean baptismal event in which the reference of ‘Aristotle’ was *apparently* established via a reference-fixing definite description. I say ‘apparently’ because, unbeknownst to the brothers, it so happens that the scrolls contained works by a number of different authors. Thus, the brothers failed to fix the reference of ‘Aristotle’. Suppose that our present-day use of ‘Aristotle’ is linked via the requisite socio-linguistic chain to this failed baptismal event. (And – if you like – suppose that the scrolls that the brothers rescued became what we now believe are the works of a single author called ‘Aristotle’.)

In the third and fourth scenarios the Millian would (surely) have to concede that because in those scenarios ‘Aristotle’ would fail to refer there cannot be a thought that is being expressed by the sentence ‘Aristotle was an alcoholic’. Things, however, aren’t so clear cut with respect to the first two scenarios. It could be argued that in the first scenario ‘Aristotle’ refers to the fictional creation of the Arab scholars in the same way that ‘Sherlock Holmes’ refers to the fictional creation of Sir Arthur Conan Doyle. The same thing *mutatis mutandis* can be argued with respect to the second case. The bearer of the name ‘Aristotle’ could arguably be the perceptual illusion jointly suffered by the fourth century Athenians.

If either, or both, of these two sorts of possibilities are allowed on behalf of the Millian, then my *reductio* may need modification. The charge may cease to be that if I could know a priori that I am thinking the thought ‘Aristotle was an alcoholic’, then I could know a priori that Aristotle existed. Instead the charge could be that I would be able to know a priori that either Aristotle existed (i.e. was flesh and blood) or there existed someone or some people who gave birth to a fictional character called ‘Aristotle’. Or, it could be that I would be able to know a priori that either Aristotle existed or there existed someone or some people who suffered an illusion of seeing a man they called ‘Aristotle’. Or, finally, the charge could be that I would be able to know a priori that either Aristotle existed or there existed someone or some people who gave birth to a fictional character called ‘Aristotle’ or there existed someone or some people who suffered an illusion of seeing a man they called ‘Aristotle’. Such a priori knowledge would maybe be less substantive than the a priori knowledge that Aristotle existed (was flesh and blood). However, it will still be the case that I should not have such a priori

knowledge. They are still things that we should not be able to know a priori. Thus, the reductio is maintained.⁴⁹

Alternatively, we could go the other way and contend that the ex hypothesi a priori knowledge that I am thinking ‘Aristotle was an alcoholic’ would allow me, absurdly, to rule out via a priori means scenarios like the third and fourth scenarios. The charge would thus become that, if I could know a priori that I am thinking the thought ‘Aristotle was an alcoholic’, I could know a priori that the universe is older than a few seconds. Or the charge would thus become that I could know a priori that if there was a Kripkean baptismal event in which the name ‘Aristotle’ was first introduced, then that baptism must have succeeded in establishing a reference. Again, such a priori knowledge may be less substantive than the a priori knowledge that Aristotle existed (was flesh and blood), but they are still things that I should not be able to know a priori. Thus, the reductio would still be maintained.

⁴⁹ Brown’s reductio presumes that it would be absurd for a subject to know a priori that either there is a certain natural kind or she belongs to a socio-linguistic community – see previous chapter. The Brown supposition seems very similar to a supposition that a subject can know a priori that either Aristotle existed or a community of people with whom the name originated existed. Thus, if the former is absurd, then (surely) the latter should also be considered absurd.

Chapter 7: Three *Broad* Positions on the Provenance of Our Self-Ascriptions of Thought and Other Mental States

7.0: Introduction to Chapter 7

The last five chapters of this thesis have all concerned the question of whether content externalism is incompatible with the claim that we should have an a priori access to our thoughts. In this final chapter I consider whether it would be a problem if it was. I argue that we may be able to deny a priori access by challenging a notion that is seemingly widespread throughout the debate, which is that privileged access and a priori access are equivalent.

Privileged access and a priori access would, I argue, only need to be held to be equivalent under a certain conception of what privileged access should amount to. Under this conception our having privileged access should involve each of us being the only one who is privy to the goings on in a personal private mental realm. Just as we via our senses are privy to the goings on in an external world that is observable by all, there should also be another realm of goings on that is our own private mental realm that we are privy to via some faculty that is analogous to the senses. It is only we ourselves that can discern the goings on in this private realm. No one else can. Furthermore, our judgements with regards to the goings on within this private realm will be a priori, as although they'll involve us using something that is akin to our senses, they won't require the use of our actual senses. Such a picture of privileged access would be incompatible with a denial of a priori access. I suggest, however, that, in the case of thought content at least, we can dispense with *such* a notion of privileged access whilst still maintaining that we have a privileged access of another sort to our thoughts.

7.1: Outline of Chapter 7

I start the chapter (§7.2) briefly considering two prominent instances in which it is apparent that privileged access and a priori access are conflated.

I then (§7.3) characterise three broad positions that are possible to have regarding the method by which we self-ascribe a given class of mental states. The first position grants privileged access to the given class of mental states, but also requires a priori access. The second position holds that the way in which we self-ascribe mental states of a given class doesn't necessarily require a priori access. This second position, however, denies privileged access. The third position grants privileged access to mental states of a given class, without necessarily requiring a priori access. I argue that in the case of thought there are grounds for rejecting the first and second positions and opting for the third that are independent of the compatibility debate.

I then (§7.4) sketch an account of the method by which we self-ascribe thought content, which adheres to the third position. This account should grant us privileged access to our thought contents, without necessarily requiring us to have a priori access. Thus, it should demonstrate how it could be that we can have both externalism and privileged access, without needing a priori access as well.

In the final section (§7.5) I consider how my account intersects with what has been discussed in previous chapters.

7.2: McKinsey and Brown and the Supposed Equivalence of A Priori and Privileged Access

The implicit assumption that privileged access and a priori access are equivalent is widespread throughout the compatibility debate. For one prominent instance consider the opening paragraph of McKinsey's 'Anti-Individualism and Privileged Access':

It has been a philosophical commonplace, at least since Descartes, to hold that each of us can know the existence and content of his own mental states in a privileged way that is available to no one else. This has at least seemed true with respect to those 'neutral' cognitive attitudes such as thought, belief, intention, and desire, whose propositional contents may be false. The crucial idea is not that one's knowledge of these states in oneself is incorrigible, for surely we can make mistakes about what one believes, intends, or desires. Rather the idea is that we can in principle find out about these states 'just by thinking', without launching an empirical investigation or making any assumptions about the external world. I will call knowledge obtained independently of empirical investigation *a priori* knowledge. And I will call the principle that it is possible to have *a priori* knowledge of one's own neutral cognitive attitude states, the Principle of Privileged Access, or just 'privileged access' for short.

(McKinsey, 1991, p.9, McKinsey's Emphasis)

McKinsey here starts talking about the notion that "each of us can know the existence and content of his own mental states in a privileged way that is available to no one else". He then switches to talk of how each of us can know the existence and content of their own mental states "without launching an empirical investigation or making any assumptions about the external world". It is clear that McKinsey regards the two things to be equivalent.

For a second equally prominent instance, consider the following passage from Brown's *Anti-individualism and Knowledge*:

A subject's first-person access to her own propositional attitudes is strikingly different from her third-person access to the propositional attitudes of other subjects. To know what someone else thinks, a subject must use empirical evidence about behavior, whether linguistic or non-linguistic. For instance, I might attribute to you the belief that today is Tuesday, on the grounds that you say it is, or that your actions fit the activities in your diary for Tuesday. In other cases, the link might be less direct. In your absence I might attribute to you the belief that the government's new education policy is wrong. Here, I haven't used any information about your current behaviour. Instead, I use my general knowledge of your political views gained in the past using behavioral evidence. By contrast, a subject can know what she herself thinks without basing this on evidence about her own behavior, whether past or present. For example, I can know that I believe that today is Tuesday without first observing my behavior, say, hearing myself say 'Today is Tuesday', or noticing myself go about the activities in my diary for Tuesday. Indeed, I can know that I have this belief without evidencing it in my behaviour at all (perhaps, I have only just woken up). Similarly, I can know that I believe that the government's new policy is wrong without inferring this from a general knowledge of my own political views, gained through observation of my own past behavior. This distinctive way of gaining knowledge seems available for a wide range of propositional attitudes. Any subject has a host of beliefs about mundane matters of fact that she can know herself to have without using behavioural evidence, such as beliefs about geographical facts (e.g., that Paris is the capital of France), beliefs about the properties of object and kind (e.g., that water is wet), beliefs about numbers (e.g., that $1+1=2$), and so on. There are numerous examples of other attitudes a subject can know herself to have without reliance on behaviour, including certain desires (e.g., that this meeting not go on much longer), hopes (e.g., that I don't buy my grandmother the same present for Christmas as my brother) and intentions (e.g., that I will go shopping later on).

(Brown, 2004, pp.33-34)

Brown here argues for what McKinsey claims has been “a philosophical commonplace, since Descartes” to hold, which is that “each of us can know the existence and content of his own mental states in a privileged way that is available to no one else”. Brown then draws a conclusion about what McKinsey labels “the Principle of Privileged Access”, which is the principle that each of us can know the existence and content of their own mental states “without launching an empirical investigation or making any assumptions about the external world”:

It seems, then, that whereas knowledge of others’ propositional attitudes is based on empirical evidence concerning behavior, knowledge of one’s own propositional attitudes is typically independent of empirical evidence, whether about behavior or the external world in general. In short, a subject can typically have a priori knowledge of her propositional attitudes, that is, she can have such knowledge without basing it in a justificatory way on perceptual experience.

(Brown, 2004, p.35)

It is clear that Brown also views privileged access and a priori access as amounting to the same thing.⁵⁰

7.3: Three Broad Positions Concerning the Method by Which We Self-Ascribe Mental States

7.31: *The First Position*

Viewing privileged access and a priori access as being equivalent would be valid *if* a certain broad and generalised position on the way in which we self-ascribe mental states were true. This position is, however, only the *first* of what are *three* broad and generalised positions that I believe it is possible to have regarding the ways in which we self-ascribe mental states. This first position holds that a subject self-ascribes her mental states, or a given class of her mental states – e.g. her thoughts – via her discerning the goings on in a personal and private realm that only she is privy to.

An adherent to this first position will hold that the way or ways in which a subject discerns her mental states are akin to the ways in which she discerns the goings on in her external environment,

⁵⁰ Although in the quotations both McKinsey and Brown talk more broadly of the epistemic access we have to our propositional attitudes, when they get into the main meat of their respective lines of argument, they switch exclusively to talk of the epistemic access we have to our occurrent thoughts.

but are entirely separate, and therefore, for that reason, are necessarily both privileged and a priori. The way or ways in which a subject discerns her mental states will be privileged because they will only afford the subject herself epistemic access to her own private mental realm. They will be a priori because it is only the goings on of the external realm, not the internal realm, that should require *a posteriori* discernment, and the private and the external realms are separate.

Suppose that, as part of some sort of party game, you have both of your hands plunged into two boxes. You are to guess what is in each of those boxes through touch alone. The judgements that you make via the use of your left hand with regards to what is in the left-hand box are separate from the judgements that you will make via the use of your right hand with regards to what is in the right-hand box. An advocate of the first thesis holds that a subject *metaphorically* has two different hands plunged into two different boxes. One of those metaphorical boxes is the external realm, and the metaphorical hand that she has feeling around in that metaphorical box is her senses. The other of those two metaphorical boxes is the subject's own private mental realm, and the metaphorical hand that the subject has feeling around in that metaphorical box is something akin to but other from her senses. There are potentially other hands in the former box, but it is only the subject's own hand that is able to feel around in the latter. It is only through our senses that we make a posteriori judgements. Everything else is a priori. The judgements that the subject makes with regards to the second metaphorical box will be a priori. They will also be privileged, because it is only the subject herself that has a hand in the second box.

The sixth chapter of Ryle's *The Concept of Mind* is essentially an attack on the first position, which Ryle saw there being widespread assent for. Consider the opening paragraph:

A natural counterpart to the theory that minds constitute a world other than 'the physical world' is the theory that there exist ways of discovering the contents of this other world which are counterparts to our ways of discovering the contents of the physical world. In sense perception we ascertain what exists and happens in space; so what exists or happens in the mind must also be ascertained in perception, but perception of a different and refined sort, one not requiring the functioning of gross bodily organs.

(Ryle, 1949, p.148)

The first position is essentially what Ryle here sees as the "the theory" that is a "natural counterpart to the theory that minds constitute a world other than 'the physical world'".

Now, the great aim of *The Concept of Mind* was a comprehensive attack on dualism. Ryle was a physicalist. He – as is evident in the quotation – saw the first position as a "natural counterpart" to dualism, and therefore, for that reason, something in need of refutation. It is important to note, however, that that connection need not necessarily be made. It is possible to be anti-dualist and still

adhere to the first position. Indeed, as we are about to see, Armstrong et al are adherents to the first positions and yet they, like Ryle, are physicalists. The first position holds that there are private mental realms, and it is only the owners of those private realms that are privy to the goings on therein. Such private mental realms need not ultimately be non-physical as Ryle's choice of language suggests.

There might also be further options. A holder of the first position need not necessarily be a dualist or a physicalist. There are no obvious reasons why a holder of the first position could not be (say) an *idealist* – and hold that everything is mental – or (say) a *neutral monist* – and hold that everything is ultimately made of something that is neither physical nor mental.

Ryle goes on to characterise *two* ways in which he claims that it is supposed that the “contents” of the “other world” may be discerned. Firstly, Ryle characterises a supposed “constant awareness” (ibid.) of the things that are meant to occur in the other world. There are processes that occur in the other world that a subject cannot help but be aware of:

The states and operations of the mind are states and operations of which it is necessarily aware, in some sense, and this sense of ‘aware’, and this awareness is incapable of being delusive. The things that a mind does or experiences are self-intimating, and this is supposed to be a feature which characterises these feelings not just sometimes but always. (ibid., p.152)

For explanatory aid Ryle gives a couple of similes:

...it is supposed that mental processes are phosphorescent, like tropical sea-water, which makes itself visible by the light which itself admits. Or, to use another simile, mental processes, are ‘overheard’ by the mind whose processes they are, somewhat as a speaker overhears the words he is himself uttering. (ibid.)

Secondly, Ryle characterises a process which, in contrast, a subject deliberately engages in known as ‘introspection’. In addition to being made aware of the goings on within our private mental realms by their very “phosphorescence”, we can also consciously attend to such goings on:

It is supposed that much as a person may at a particular moment be listening to a flute, savouring wine, or regarding a waterfall, so he may be ‘regarding’ in a non-optical sense, some current mental state or process of his own. The state or process is being deliberately and attentively scrutinized and so can be listed among the objects of observation. (ibid., p.157)

An adherent to the first position need not *necessarily* hold that the ways in which a subject discerns are exactly as the two that Ryle characterises here. But, it is hard to see how the adherent to the first position could characterise the epistemic relationship that a subject has with respect to the goings on within her private mental realm, without that characterisation implicitly involving either the subject being made aware of those goings on, or the subject consciously making an effort to discern those goings on, or both.

The first position can be characterised as having a stronger and a weaker version. The stronger version concerns *both* the provenance of our self-ascriptions of our mental states, *and* the epistemic worth of our self-ascriptions of mental states. The weaker version concerns the provenance of our self-ascriptions of our mental states *only*. The weaker version holds that we can *form true beliefs* regarding our mental states through being privy to a private mental realm. The stronger version of the first position holds that we can *know* our mental states through being privy to a private mental realm. We can form true beliefs regarding our mental states through being privy to a private mental realm, *and* those true beliefs constitute knowledge.

Gertler, in her surveys of the various positions on self-knowledge (Gertler, 2011; 2020), characterises two *major* theories that essentially purport to demonstrate how it is that the first position is true, and furthermore that it is not just the weaker version that is true, the *stronger* version is also true. These two theories both purport to give an account of a way in which a subject could form true beliefs regarding her mental states that involves the subject discerning the goings on in a private mental realm, and explain why such true beliefs would constitute knowledge. The two theories purport to explain *how* a subject could form true beliefs regarding her mental states, and furthermore purport to explain *why* such beliefs would constitute knowledge. In both cases, those explanations picture the subject as being privy to the goings on in a private mental realm.

The first of the two theories Gertler labels *The Acquaintance Theory*. The acquaintance theory states simply that a subject will be *acquainted*, in the Russellian sense, with at least some of her mental states, and it is through that acquaintance that the subject is able to acquire knowledge of those mental states. Who Gertler fashions as *contemporary acquaintance theorists* (Bonjour, 2003, Chalmers, 2003, Conee, 1994; Fales, 1996; Feldman, 2004; Fumerton, 1995; Gertler, 2001; Levine, 2006; Pitt, 2004) maintain that the epistemic relation – that of ‘acquaintance’ – that Russell (Russell, 2010-11; 2012) famously claimed held between us and sense impressions, and between us and universals, holds between us and certain of our mental states. Furthermore, they maintain that it is

through such acquaintance that we come to self-ascribe those mental states. *Furthermore*, they maintain that that is the reason why such self-ascriptions constitute knowledge.⁵¹

Gertler (Gertler, 2011, p.95) suggests that the class of mental states that we could most plausibly know through being acquainted with are what she terms ‘sensations’, by which she means phenomenal mental states of experiencing a certain sensation, e.g. the phenomenal state of feeling a certain pain. Gertler (*ibid.*, p.123) also suggests that there could be other mental states that we could, for similar reasons, plausibly know through being acquainted with, if those states are similarly phenomenal states.

The second major theory that adheres to the first position that Gertler characterises is *The Inner Sense Theory*. The inner sense theory states that just as a subject is able to discern the goings on via the use of her senses and thus come to know things about the world, she is also able to discern the goings on in her own private mental realm and thus come to know her mental states via the use of a special ‘inner sense’.

Gertler outlines a *historical* version of the inner sense theory that she attributes to Locke:

This Source of Ideas, every Man has wholly in himself...And though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense.

(Locke, 1687, II.i.iv; as quoted by Gertler, 2011, p.39, p.129)

According to Locke’s famous *tabula rasa* doctrine there is no such thing as an ‘innate idea’. All ideas have an experiential source. They can come from the external world via our senses. Alternatively, ideas can come from the source that is being referred to in the quotation, which necessitates an internal mental realm that “every Man has wholly in himself”, and a means of discerning the goings on therein that “might properly enough be call'd internal Sense”.

Gertler also outlines a more *contemporary* version of the inner sense theory that she attributes to Armstrong (and to Lycan⁵² and to others⁵³). Armstrong characterises “a self-scanning process in the brain” (Armstrong, 1968; Gertler, 2011, p.132) that allows a subject to know her mental states. This self-scanning process is like a sense in that it allows a subject to know her mental states without knowing their true nature. A table is an arrangement of atoms. By seeing and touching a table I may

⁵¹ Through the course of her chapter on the acquaintance theory (Gertler, 2011, Ch. 4, pp.87-128), Gertler elaborates a more detailed account of how it is that we come to form knowledge constituting beliefs about our mental states through being acquainted with them. The details of which we need not here concern ourselves with.

⁵² Lycan, 1996.

⁵³ E.g. Goldman, 2006.

not necessarily be aware that it is an arrangement of atoms, but I can still know that it is there, and that it has certain properties. For Armstrong mental states are, as are tables and everything else, essentially arrangements of atoms. Armstrong is a physicalist. It is through the use of “a self-scanning process in the brain” that we can come to know our mental states, that we are in possession of them, and that they have certain properties, without necessarily knowing their true nature.

In terms of their broad outlines, the acquaintance theory and the inner sense theory are similar. Acquaintance is akin to a sense, but it is not quite like a sense. An ‘inner sense’ would similarly be akin to a sense, but not quite like a sense. Both the acquaintance theory and the inner sense theory propose that there is a means that we have of discerning the goings on in our own cognition that is akin to, but not quite like a sense.

There are ways in which the two theories differ that we need not overly concern ourselves with here. For instance, one way in which the acquaintance theory and the inner sense theory differ is that Russellian acquaintance comes pre-packaged with a guarantee that something described as being like a ‘sense’ may not have. If we are acquainted with some thing – in the Russellian sense – then there is absolutely no way that that thing cannot exist. If we are acquainted – in the Russellian sense – with our mental states, then it will necessarily be the case that we are subject to those mental states.⁵⁴ If, on the other hand, we in some sense *sense* our mental states then we won’t necessarily have that same guarantee. The phenomenon of perceptual illusion allows that we can sense in the ordinary sense of ‘sense’ – see, hear, etc. – things that are not in fact there. If the way in which we discern our mental states is characterised as being an ‘inner sense’ that isn’t Russellian acquaintance exactly, but will nevertheless be akin to a sense, then that allows for the possibility that such an ‘inner sense’ could be something that, although reliable most of the time, could, in exceptional circumstances, lead us astray in a way that acquaintance could not.

Another way in which, according to Gertler, *contemporary* versions of the inner sense theory, such as those of Armstrong (et al), and the acquaintance theory are different is that contemporary versions of the inner sense theory are epistemologically *externalist*, whereas the acquaintance theory is epistemologically *internalist*.⁵⁵

⁵⁴ This would not necessarily mean that the self-ascriptions that we’d make through such acquaintance would be infallible – see Gertler 2011, Ch.4.

⁵⁵ I hear footsteps. I thus form the true belief that there is someone walking nearby. Supposing that that true belief constitutes knowledge, there are broadly two reasons that could be given as to why it does.

The first is that when I heard the footsteps, I experienced the phenomenology of hearing footsteps. That is, I experienced a phenomenological episode that has a certain character that constitutes my hearing footsteps. My experiencing that phenomenological episode with that character is what *justifies* my belief that there is someone walking nearby, and that is why my true belief constitutes knowledge.

7.32: *The Second Position*

The position characterised in the previous subsection one of three positions that I believe that it is possible to have regarding the ways in which we self-ascribe mental states. I characterise the third position in the next subsection. I characterise the second in this present subsection.

The second position consists of two commitments. The first commitment is that, for a given class of mental states, the self-ascription of those mental states does *not* and *cannot* involve the consultation of a private mental realm that only the self-ascriber is privy to. The second commitment is that, for that given class of mental states, a third party could, in principle, be at least as equally able, and as epistemically qualified, to discern those mental states as the self-ascriber is.

In chapter six of *The Concept of Mind*, after Ryle has dispensed with – what I am calling – the first position, he goes on to argue that the second position will be true with respect to a potentially

A second reason is that the belief forming process – i.e. my sense of hearing – that leads to my belief that there is someone walking nearby is *reliable*. I may have experienced a phenomenological episode with a specific character when I heard the footsteps. But that is irrelevant. What matters, for the question of knowledge, is that my belief “will be the outcome of a reliable process, causally linked to its truthmaker” (Gertler, 2011, p.131), which in this case is the person who is walking nearby.

If we are to suppose that a subject can discern her mental states because of a special epistemic relationship that she has with respect to the goings on within her own cognition, *and* if we also suppose that the beliefs about her mental states that the subject would thereby form would constitute knowledge, then there will similarly be two broad reasons as to why they would.

Firstly, it could be maintained that the phenomenology that a subject experiences when she self-ascribes a mental state justifies her self-ascription, and thereby confers it as knowledge. Just as I experience a phenomenological episode that has a specific character when I judge that there is someone walking nearby, I experience a similar phenomenological episode that also has its own specific character when I judge that I believe that Paris is the capital of France. It is my experiencing that phenomenological episode with that specific character that is, firstly, responsible for my self-ascribing the belief that Paris is the capital of France, and that, secondly, is the reason why that self-ascription constitutes knowledge.

Secondly, it could be maintained that what matters is the reliability of the process that leads to a subject’s belief that she has a certain mental state. The subject may well experience a phenomenological episode when she self-ascribes her mental state that is particular to the self-ascribing of that mental state. Such phenomenology, however, is irrelevant, when it comes to the question of whether the subject knows that she is subject to the mental state. In forming my belief that I believe that Paris is the capital of France I may experience a phenomenological episode that has a certain character, which will be specific to the mental process that is my forming the belief that I believe that Paris is the capital of France. But that as no bearing on whether my belief that I believe that Paris is the capital of France constitutes knowledge. The reason why my belief that I believe that Paris is the capital of France constitutes knowledge is that it “will be the outcome of a reliable process, causally linked to its truthmaker”, which in this case will be my belief that Paris is the capital of France.

Who Gertler characterises as acquaintance theorists emphasise the first reason, whereas contemporary versions of the inner sense theory, such as those of Armstrong and Lycan and others, emphasise the second.

very large range of mental states. Ryle (1949, pp.166-177) argues that, for a large range of mental states, a subject knowing or being aware of those mental states – what we might describe as ‘her knowing or being aware of those mental states’ – is essentially a matter of her being aware, or her “being alive to” what it is she will/would do or say when/if certain contingencies arise. That being aware of or “being alive to” what it is she will or would do essentially consists of the subject making accurate predictions regarding what she will or would do. Such predictions can only be based on things that are publicly observable, such as the subject’s past behaviour. There is nothing that is privately observable to base them on. There is no private mental realm. Thus, the predictions that the subject makes are predictions that can potentially be made by a third party – who knows the subject exceptionally well, who (say) has been watching her since birth. Furthermore, the predictions made by such a third party can potentially be of the same epistemic worth, as they can potentially be based on the same data set, namely the subject’s past behaviour and other things that are publicly observable.

A subject who is walking alone along a seafront promenade is unlikely to avow aloud ‘I want ice cream’ to no-one in particular. She may, however, in a counterfactual scenario, in which (say) she was walking with a companion, avow ‘I want ice cream’ to that counterfactual companion. Such a counterfactual avowal would – or at least could⁵⁶ – be indicative of what might properly be described as ‘a desire for ice cream on the part of the subject’. Essentially, Ryle’s contention is that it is the subject’s consideration of such counterfactual scenarios, and her judging on how she’d act in such counterfactual scenarios, her judging that she’d avow ‘I want ice cream’, that leads to, and even warrants, the subject self-ascribing to herself a desire for ice cream. Furthermore, the subject’s judgement that she’d behave in a certain way can only be based on things, such as her past behaviour, that are publicly observable. It could not possibly be based on anything else. The subject’s self-ascription of a desire for ice cream is essentially a prediction. It is a prediction that the subject makes regarding how she would behave. It is a prediction that can *only* be based on things that a third party could in principle be privy to. There is nothing that could contribute to the subject’s calculation that only she can be privy to. Thus, it is a prediction that could, in theory, be made with the same epistemic confidence by a third party.⁵⁷

⁵⁶ Whether it is indicative of ‘a desire for ice cream on the part of the subject’ may be dependent upon the precise details of what is actually the case, and upon the precise details of the counterfactual scenario in which the subject would make the avowal, and also upon the precise details of other past, present, future, actual, and counterfactual behaviours of the subject.

⁵⁷ A more contemporary adherent to the second position is Carruthers (Carruthers, 2011; Gertler, 2020). Carruthers proposes that the way in which we discern many of our own mental states is via a special faculty that originally evolved in order to allow us to discern the mental states of others, but was later in our evolutionary history co-opted for the purpose of the discernment of our own mental states. We self-ascribe our own mental states based on the same cues, such as externally observable behaviour, that lead us to ascribe mental states to others. The provenance of our self-ascriptions of mental states,

The first position, considered in the previous section, grants that we have privileged access to our mental states, but requires that we have a priori access. In contrast, the second position, which we are considering in this present section, does not require that we have a priori access, but flat out denies that we have privileged access. The third position, which I consider in the next section, allows that we can have privileged access in such a way that may not require a priori access.

7.33: The Third Position

The two positions I characterised in the previous two subsections are not (I believe) exhaustive of the positions that it is possible to have regarding the ways in which we self-ascribe mental states. There is also a third position. This third position consists of the second position's first commitment and the *denial* of the second position's second commitment. An adherent to the third position *agrees* with the adherent to the second position that the self-ascription of the given class of mental states does not and cannot involve the consultation of a private mental realm. An adherent to the third thesis, however, *denies* that a third party could in principle be at least as equally able and just as epistemically qualified to ascribe mental states, of the given class, to a subject as the subject herself is.

A famous and often quoted Evans passage can be interpreted – see below (§7.4) – as demonstrating how it could be that the third thesis is true in the case of *belief*:

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

(Evans, 1982, p.225)

I take stock of current world events and come to belief that there is going to be a third world war. By taking stock of those same current world events I can also come to believe that I believe that there is going to be a third world war. I don't self-ascribe the belief based on the goings on in some private mental realm that only I am privy to. I self-ascribe my belief based on facts about current world events which in principle are available to everyone. But it is only I that can ascribe to myself a belief based on those facts with any great certainty. A third party may *surmise*, given that I am privy to the facts about current world events, and given that I am a rational and reasonably intelligent person, that

therefore, is such that it is the case that a third party could, in theory, ascribe mental states to us with the same degree of epistemic confidence as we ourselves can.

I therefore should surely believe that there is going to be a third world war. But they can *only* surmise that I believe that there is going to be a third world war. A third party could not be as certain that I believe that there is going to be a third world war as I myself am, even though both I and the third party may be basing our respective judgements on a convergent data set. The epistemic access I have to my belief is privileged. A third party could not be as equally able or as epistemically qualified to ascribe to me the belief as I myself am. The privileged epistemic access I have to my belief, however, is not privileged because of my being privy to the goings on within some private mental realm.

7.34: Why be a Third Positionist?

I have now characterised three broad positions on how it is that we come to self-ascribe mental states of a given sort. The first position maintains that we self-ascribe mental states, of the given sort, via being privy to a private mental realm. We have privileged access because we have a priori access to that private mental realm. The second position denies that there are such things as private mental realms. It maintains that an external observer is in principle able to ascribe mental states to a subject with the same authority as the subject herself is, and thus we *do not* have privileged access. The third position also denies that there are such things as private mental realms, but maintains, nevertheless, that an external observer *could not* in principle be able to ascribe mental states to a subject with the same authority as the subject herself is, and thus we *do* have privileged access. Moreover, the third position may allow that we could have privileged access that is not predicated on a priori access.

In the next section I present a new account of what could be the provenance of our self-ascriptions of thought content. This new account should demonstrate how it could be that the third position could be true with respect to thought content. It should demonstrate how it could be that we self-ascribe our thoughts based on publicly discernible facts, but, nevertheless, do so in such a way that a third party could not.

My reason for doing so in this present thesis is to demonstrate how we can have privileged access to our thoughts without it needing to be a priori, and thus how we can have privileged access even if content externalism is true. There is however an independent and possibly stronger reason for such an approach, which is that, in the case of thought, there are grounds for the rejection of the other two positions.

If we are to suppose that we have privileged access to our thoughts then we must reject the second position.

The first position would be undesirable because of something that was observed in the opening paragraph of this thesis. If we do self-ascribe thought via our being privy to some private mental realm, as a first-positionist approach would suppose that we do, then (surely) it should be the case that thoughts produce their own phenomenology. Otherwise, there would be no basis on which

we self-ascribe thought. Our thoughts should *announce* themselves to us, and they could only do that via the phenomenology that they would need to give off – they should be “phosphorescent, like tropical sea-water, which makes itself visible by the light which itself admits”. As was observed in the opening paragraph of this thesis, thoughts produce no phenomenology. It, therefore, (surely) cannot be the case that we self-ascribe thought via being privy to some private mental realm.⁵⁸

There is a serious drawback to taking a third-positionist approach. If we self-ascribe thought, or any given class of mental states, via the obtuse means that a third-positionist approach would suppose, then the epistemic relationship we have with those mental states will be suspect. Consider again the Evans example. If I base my belief that there is going to be a third world war on it being the case that there is going to be a third world war, then I will have ‘inferred’ – if what I will have done can be called that – a fact about my internal mentality from something that has nothing to do with my internal mentality. The same will be true of *thought* – as we’ll be seeing in the next section.

This may seem to be quite a bullet that must be bitten. However, it does seem reasonable to suppose that the three positions I’ve characterised may be exhaustive of the relevant dialectical space. It is hard to see how a theory regarding the provenance of our self-ascriptions of mental states could be neither first-positionist, second-positionist, nor third-positionist. (It is hard to imagine what *fourth positionism*, or even *fifth positionism*, could look like.) If we do assume exhaustion, then we can contend that our bullet biting is necessitated not so much because of the possible incompatibility of content externalism and claims that we should have an a priori access to our thoughts, but rather because of the undesirability of first positionism and second positionism in the case of occurrent thought.⁵⁹

7.4: A Third-Positionist Account of the Method by Which We Self-Ascribe Thought

Consider again the famous Evans quote:

[I]n making a self-ascription of belief, one’s eyes are, so to speak or occasionally literally, directed outward, upon the world. If someone asks me “Do you think there is going to be

⁵⁸ This is not intended as a *water-tight* refutation of first positionism. For starters, it requires that one agrees with me that thoughts produce no phenomenology. Also, it is – strictly speaking – still possible to be a first positionist *whilst* agreeing with me that thoughts produce no phenomenology. One can be an inner sense theorist à la Armstrong et al but contend that not only is it the case that any phenomenology produced through mental state self-ascription is not relevant to assessing whether such ascriptions count as knowledge (see fn. 55), it is actually the case that no such phenomenology is produced. In other words, one can be an inner sense theorist (à la Armstrong et al) who denies epiphenomenalism.

⁵⁹ One could bite an alternative bullet by being a hard-nosed-materialist epiphenomenalism-denying inner sense theorist – see previous footnote.

a third world war?” I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question “Will there be a third world war?”

(Evans, 1982, p.225)

Evans can be interpreted as proposing that we self-ascribe belief via a sort of mental act that is almost like an act of reasoning but isn't exactly. A subject judges something to be the case, and that leads her to judge that she believes that that something is the case. I judge that there is going to be a third world war and based on that I judge further that I believe that there is going to be a third world war. I make a judgement concerning geo-politics and based on that I make a judgement concerning what it is that I believe. In an act of reasoning we start by judging that something is the case and based on that we judge that something else is the case. We may (say) judge that both $\langle p \rangle$ and $\langle p \rightarrow q \rangle$ are the case and based on that judge (correctly – if $\langle p \rangle$ and $\langle p \rightarrow q \rangle$ are both indeed the case) that $\langle q \rangle$ is the case. This would constitute an act of deductive reasoning. We may (say) judge that all swans so far encountered have been white and based on that judge (plausibly – but falsely) that all swans are white. This would be an example of induction. Finally, there is abduction. I may judge (say) that I am experiencing phenomenology that is consistent with the external world being real and from that judge that the external world being real is the more plausible explanation – as opposed to the usual sceptical alternatives – for why that is the case and thus that the external world is indeed real. Evans's act of quasi-reasoning is almost like such acts of reasoning in that it involves a subject judging that something is the case and then based on that judging that something else is the case. But it isn't exactly an act of reasoning.

This interpretation of Evans diverges from Moran's (Moran, 2001). Moran presents the Evans quote alongside the following quote from Edgley, in which the notion of 'transparency' is introduced:

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

(Edgley, 1969, p.90; quoted in Moran, 2001, p.61)

Moran views Evans and Edgley as getting at the same point, that they are both trying to bring to light the notion of 'transparency', even though it is only Edgley that actually uses the term “transparent”:

To claim that one question is “transparent” to another is not to claim that one question reduces to the other. The fact that answers the question about the war is different from the fact about a particular person’s belief. As Edgley points out, the correct answers to the two questions need not be the same. But nor is it right to say, as he does, that the two questions are indistinguishable from within the first-person point of view. After all, it isn’t as if, although the answers to the two questions are in fact distinct, I must remain somehow in the dark about this, or that I cannot see them pointing in different directions. It will be common knowledge, among anyone with the concept of belief, that although one believes something as true, the fact believed and the fact of one’s belief are two different matters. From within the first-person perspective I acknowledge the two questions as distinct in virtue of acknowledging that what my beliefs are directed upon is an independent world, and they may therefore fail to conform to it. So, rather than reducibility or indistinguishability, the relation of transparency these writers are pointing toward concerns a claim about how a set of questions is to be answered, what sorts of reasons are to be taken as relevant. The claim, then, is that a first-person present-tense question about one’s belief is answered by reference to (or consideration of) the same reasons that would justify an answer to the corresponding question about the world. (Moran, 2001, pp.61-62)

Moran arrives at the following conclusion:

...the vehicle of transparency in each case [i.e. in the cases of both Edgley and Evans] lies in the requirement that I address myself to the question of my state of mind in a *deliberative* spirit, deciding and declaring myself on the matter, and not confront the question as a purely psychological one about the beliefs of someone who happens also to be me.

(*ibid.*, p.63; Moran’s emphasis)

In our interpretation of Evans we are agreeing with Moran that, even from a first-person perspective, the question of whether one believes that there is going to be a third world war is distinct from the question of whether there is going to be a third world war. But we are rejecting Moran’s contention that the former question is being answered through the entering into of a “deliberative spirit”. According to *our* interpretation of Evans, the question of whether I believe that there is going to be a third world war *is* a “purely psychological one”. I ask myself a question about geo-political matters, which I answer through (presumably) a posteriori means. I then – through a rather epistemologically unsavoury intellectual manoeuvre (see below) – use that answer to answer a further question about my

internal psychology. The question of whether I believe that there is going to be a third world war does not become – is not ‘transparent to’ (if you like) – the question of whether I *ought* to believe that there is going to be a third world war. I am not necessarily assessing my reasons for believing that there is going to be a third world war. Rather, I am drawing a conclusion about my internal psychology based on – what is for me – the *fact* that there is going to be a third world war.⁶⁰

Adapting this Moran-divergent take on Evans, in which the normative aspect is absent, leads to a new account of occurrent thought self-ascription.

Tautologically, if I judge that there is going to be a third world war, then it will be the case I am at that moment judging that there is going to be a third world war. Something will be presently happening in my cognition, which is that I am judging that there is going to be a third world war. Suppose now that my judgement that there is going to be a third world war sparked me to make a further judgement. Suppose I make a further judgement that I am *at present*⁶¹ judging that there is going to be a third world war. That latter judgement will be correct, as it will be the case that I am presently judging that there is going to be a third world war. I will have thus engaged in a similar act of quasi-reasoning that we, inferring from Evans, are supposing leads to my judgment that I *believe* that there is going to be a third world war. A judgement about geo-politics has led me to make a further judgement about my internal cognition. The difference is that it has resulted in a judgement that is about my occurrent mentality as opposed to being about something that perdures with respect to my mentality. Beliefs are things that persist over time, whereas judgements are things that happen at a given moment in time. At the moment in time in which I judge that there is going to be a third world war, it should (presumptively – see next paragraph) be the case that at that moment I believe there is going to be a third world war. It should also remain true for some time after I make the judgement – that there is going to be a third world war – that I will continue to believe that there is going to be a third world war. (It may also have been true before I made the judgement – that there is going to be a third world war – that I believed that there is going to be a third world war.) Whereas, it

⁶⁰ It is thus wrong to refer to my account as ‘a transparency account of self-knowledge’ if that leads to one seeing it as invoking the ideas of Moran, as that can only lead to misunderstandings. There may be meaningful comparisons to be made, but that is something must wait until potential sequels. In the meantime, it is far better if we all put such considerations to one side. My account is in fact *far* closer to Burge, and his account of thought self-ascription (see Chapter 2, §2.2; and §7.54 of this chapter), than it is to Moran. (We are talking here exclusively of the Burge of Burge 1988 and not the slightly later Burge of Burge 1996 – in the later paper Burge begins to tread a similar path to the path that Moran treads.)

⁶¹ This temporal sensitivity means that the account of thought self-ascription I am outlining here, like the account of Burge (see Chapter 2, §2.2, and §7.54 of this chapter), cannot – on its own – explain how a subject can discern what she was thinking a few seconds ago (say), or what she was thinking when she was stood taking in the *Mona Lisa* during the weekend break in Paris that she enjoyed last summer (say).

is only true at the moment that I make the judgement – that there is going to be a third world war – that it is the case that I am presently judging that there is going to be a third world war.

This method of judgement self-ascription has a guarantee of accuracy that the method of belief self-ascription that we are inferring from Evans arguably does not. I may, in a more despairing moment, judge that there is going to be a third world war and based on that judge further that I believe that there is going to be a third world war. I then may, at a slightly later point when my mood has been lifted, start earnestly making plans about an eventual retirement, which I am not envisaging will take place during a nuclear winter. If I am entertaining the possibility of such a retirement, then arguably it cannot be said that I truly ‘believe’ that there is going to be a third world war, and that I didn’t really ‘believe’ that there is going to be a third world war when I made the judgement that there is going to be a third world war.⁶² Thus, we have an instance in which the Evans method of belief self-ascription – i.e. the method we are inferring from Evans – will have led to falsehood. It will still, however, have been the case that when I judged that there is going to be a third world war that I was indeed at that moment judging that there was going to be a third world war. Thus if I judged at that moment that I was judging that there is going to be a third world war, it will have indeed been the case that I was judging that there was going to be a third world war.⁶³

Suppose now that my judgement that there is going to be a third world war sparks me to judge (instead of or alongside my judgement I am presently judging that there is going to be a third world war) that I am thinking – i.e. entertaining – a thought with the content <there is going to be a third world war>. That judgement will also concern my occurrent mentality; it will be about a slightly different aspect of my occurrent mentality. That judgement will also be true; it will have the same guarantee of being true. It is made true by my judgement that there is going to be a third world war. If I am judging that there is going to be a third world war, then ipso facto I am thinking – i.e. entertaining – a thought with the content <there is going to be a third world war>. What we have here is a method for occurrent thought self-ascription. A subject can judge that <p>, and that can lead her to judge further that she is thinking a thought with content <p>. Call this the *conviction method*.

Suppose that instead of judging that there is going to be a third world war I judge that it is *true or false* that there is going to be a third world war.⁶⁴ At the moment that I make this judgement –

⁶² C.f. the Gallois case considered in fn. 23, and the numerous other similar examples that are to be found throughout published philosophy.

⁶³ There will of course be other ways of interpreting the example – i.e. possibly ones that would be more favourable with respect to the Evans method – hence my insertions of the qualifier ‘arguably’.

⁶⁴ This sort of judgement – in many circumstances – might better be understood as a ‘might’ judgement. I consider whether the common clear and odourless substance in my environment, which I and my socio-linguistic brethren call ‘water’, could be present, in its liquid form, on Venus. I judge that water (in its liquid form) might be on Venus, i.e. it might or might not be the case that there is water on Venus, i.e. it’s true or false that there is water on Venus.

that it is true or false that there is going to be a third world war – it will be the case that at that very moment in time I am entertaining a thought with the content <there is going to be a third world war>. Thus, if this judgement sparks me to judge further that I am thinking – i.e. entertaining – a thought with the content <there is going to be a third world war>, then that second judgement will be correct. What we have here is a *second* method for thought self-ascription. A subject can judge that it is true or false that <p>, and that can lead her to judge further that she is thinking a thought with content <p>. Call this the *contemplation method*.

If we do self-ascribe thought via either of these two methods, then we will have a *privileged access* to our thoughts. The methods only work from the first-person perspective. *You* may judge that there is going to be a third world war, or judge that it is true or false that there is going to be a third world war, and, for some bizarre reason, based on that, judge further that *I* am presently thinking the thought <there is going to be a third world war>. I may, by sheer coincidence, indeed be presently thinking that very thought. But it is extremely unlikely. If *I*, on the other hand, judge that there is going to be a third world war, or judge that it is true or false that there is going to be a third world war, and based on that judge that *I* am presently thinking that <there is going to be a third world war>, then it will most definitely be the case that that is what *I* am presently thinking.

Depending upon the nature of the thought that is being self-ascribed, self-ascription via the conviction and contemplation methods may or may not necessarily involve a posteriori judgement. My judgement that there is going to be a third world war will be a posteriori. If that a posteriori judgement leads me to self-ascribe the thought ‘there is going to be a third world war’, then that self-ascription will require a posteriori judgment. Suppose I judge instead that it is true or false that there is going to be a third world war. That judgement is possibly a priori. If it is, and it leads to my self-ascribing the thought ‘there is going to be a third world war’ then I will have self-ascribed a thought without that self-ascription requiring a posteriori judgement.

Interestingly, the self-ascription of a *singular thought* will involve an a posteriori judgement regardless of whether it is via the conviction method or the contemplation method.

A judgement that Aristotle was an alcoholic will obviously be a posteriori. Thus, if I were to self-ascribe the thought ‘Aristotle was an alcoholic’ based on it being the case that Aristotle was an alcoholic – i.e. via the conviction method – then that self-ascription would be based on an a posteriori judgment.

Interestingly, the judgment that it is true or false that Aristotle was an alcoholic will also be a posteriori. Thus, if I were to self-ascribe the thought ‘Aristotle was an alcoholic’ based on it being the case that it is true or false Aristotle was an alcoholic – i.e. via the contemplation method – then that self-ascription would be based on a posteriori judgment. The unsubstantiveness of the judgement that it is either true or false that Aristotle was an alcoholic may, *prima facie*, make it seem purely a priori. It is, however, that very unsubstantiveness that makes the judgement a posteriori. Because of its

unsubstantiveness, the judgement boils down to – essentially is – the judgement that Aristotle exists, which is a posteriori.

We might, due to a commitment to Millianism – and hence Kripkean externalism (see Chapter 1, §1.3) – or otherwise, suppose that the content of the thought ‘Aristotle was an alcoholic’ is dependent upon Aristotle. That is, we might suppose that changing the object of the thought changes its content. If we are to suppose this, then we must also suppose that changing the object also changes the content of the second-order judgements about the thought that would be formed via either the conviction or contemplation method. Imagine a counterfactual in which the Kripkean reference-determining socio-linguistic chain, which in actuality connects us to the actual Aristotle, connects us to the numerically non-identical *Arisnotle* (see Chapter 1, §1.3). If we are to suppose that in such a counterfactual scenario, the content of ‘Aristotle was an alcoholic’ would be different, then we must also suppose that the corresponding judgement ‘I am thinking that Aristotle was an alcoholic’ would have a different content. In such a counterfactual scenario, instead of judging that it is true or false that Aristotle was an alcoholic and correspondingly judging that I am thinking that *Aristotle was an alcoholic*, I would be judging that it is true or false that *Arisnotle* was an alcoholic and correspondingly judging that I am thinking that *Arisnotle was an alcoholic*. The judgement would still be true. Its content would be reflective of the content of the thought of which it concerns.

In such a counterfactual scenario I would use the same words to express my internal ratiocinations. I would say that ‘I am judging that it is true or false that Aristotle was an alcoholic’, and I would based on that say that ‘I judge that I am thinking that Aristotle was an alcoholic’. I would, however, be using ‘Aristotle’ to refer to *Arisnotle* as opposed to (the actual) Aristotle.

Depending upon how its details are cashed out, I might not be able to distinguish my being in such a counterfactual scenario from what is actual. The judgements that I make about what I am thinking via the conviction or contemplation methods would still be true nonetheless.

Alternatively, we might suppose – as Burge does (see Chapter 1, §1.4) – that the contents of singular thoughts are independent of their object. That is, we might suppose that changing the object does not change the content. If we are to suppose this, then we must also suppose that changing the object does not change the content of the second-order judgements about the thought that would be formed via either the conviction or contemplation methods. Let us suppose that Alfred is able to think a number of singular thoughts with respect to apple 1, including (for instance) the one he’d express ‘*that* apple [pointing to apple 1] looks delicious’. Let us also suppose that in a counterfactual scenario in which apple 1 is replaced by the numerically non-identical apple 2, such singular thoughts would have the same content as they actually have. If we are to suppose this, then the second-order judgements that Alfred would make with respect to what he is thinking would have the same content to what they actually have.

Alfred would still self-ascribe his thoughts through a posteriori judgement regardless of whether he does so via the conviction or contemplation method. If Alfred were to self-ascribe the thought ‘*that* apple look delicious’ via the conviction method, then he would do so via making the clearly a posteriori judgement that *that* apple looks delicious. If he were to ascribe via the contemplation method, he would self-ascribe the thought through judging that it is true or false that *that* apple looks delicious. This judgement will also be a posteriori for reasons already stated. The judgement will essentially be that the apple exists.

Along with singular thought there is also another sort, or rather another three sorts, of thought that the self-ascription of which would necessarily involve a posteriori judgement regardless of whether the conviction or contemplation method is used. The self-ascriptions of the thoughts that Burge claims that we are able to think – and that he derives his content externalism from (see Chapter 1, §1.2) – would necessarily involve a posteriori judgement regardless of whether they are arrived at via the conviction or contemplation method.

Recall the case of Arthritic Alf. Alf is able to think thoughts that are expressible using a term, ‘arthritis’, that Alf incompletely understands. He is unaware that the term ‘arthritis’ is only applicable to ailments of the joints. If Alf were to self-ascribe the thought ‘arthritis is not as bad as cancer’ via the conviction method, then he would do so through judging that arthritis is not as bad as cancer. Such a judgement should be a posteriori. If he were to self-ascribe the thought via the contemplation method, then he’d do so through judging that it is true or false that arthritis is not as bad as cancer. This would also constitute an a posteriori judgement. It would be a judgement that Alf makes about his socio-linguistic community. It is a judgement that there is this ailment, of which he suffers from, that is called ‘arthritis’ that certain things can be true or false with respect to.

In Burge’s reading of Twin Earth, if Oscar were to self-ascribe the thought ‘water is wet’ via the conviction method, then he would do so by judging that water is wet, which is obviously a posteriori. If Oscar were to self-ascribe his thought via the contemplation method, then he’d do so through judging that it is true or false that water is wet. In making such a judgement Oscar will be making a judgement about his socio-linguistic community, and about the interactions that more scientifically literate members of his socio-linguistic community have had with a naturally occurring substance. Oscar will be judging that there is this naturally occurring substance called, by the members of the socio-linguistic community to which Oscar belongs, ‘water’, of which things can be true or false of, and of which there are members of his socio-linguistic who know a lot more than Oscar himself knows about its internal nature, and have a much better sense what should and what shouldn’t be considered ‘water’. The judgement will be simultaneously about what substances exist in Oscar’s external environment *and* about how there are more knowledgeable members of his socio-linguistic community who are better able to classify, and have classified, such substances.

Finally, recall Burge's 'sofa' thought experiment, in which there is a "Person A" who believes that sofas are a sort of religious artefact that would normally collapse if sat upon – i.e. has a "nonstandard theory" with respect to what 'sofas' are. If this person A were to self-ascribe the thought 'sofas normally collapse if sat upon' via the conviction method, then he'd do so through judging that sofas normally collapse if sat upon. That judgement would obviously be a posteriori. If Person A were to self-ascribe the thought via the contemplation method, then he'd do so through judging that it is true or false that sofas normally collapse if sat upon. This judgement would also be a posteriori. Person A will be judging that there are things he encounters daily that his wider socio-linguistic community call 'sofas', and it will be true or false of these 'sofas' that they will normally collapse if sat upon. (Person A believes that his socio-linguistic peers are wrong about the nature of these artefacts and are thus wrong about what should be considered their defining characteristics. But, in the case of thought self-ascription, that is irrelevant.)

As is the case *mutatis mutandis* with respect to the Millian/Kripkean view of singular thought, if Arthritic Alf were to use the conviction or contemplation methods to self-ascribe a relevant thought in the counterfactual scenario in which the term 'arthritis' encompasses a larger range of ailments – i.e. not just those that effect the joints – then that would still yield true self-ascriptions. The relevant second-order beliefs will have contents that are different but are reflective of the contents of the relevant first-order thoughts in such a way that that ensures their truth. The same will be true *mutatis mutandis* with respect to the other two Burgean cases.

Now, there are an infinite number of primes. This is something that would be true even if the universe was such that it was incapable of producing complex life. More particularly, it would still be true if there were no beings in the universe capable of comprehending that there are an infinite number of primes. It would be preposterous to suppose that because there are an infinite number of primes, then that means that there is a very specific sentient being – presumptively the endpoint of a complex evolutionary history – who at a very specific moment in time thought the thought <there are an infinite number of primes>. It would be even more preposterous to suppose that because it is *true or false* that there are an infinite number of primes, then that means that there is such a being who at a very specific moment in time thought the thought <there are an infinite number of primes>. Yet making such preposterous intellectual jumps is precisely what my proposed account of thought self-ascription supposes that we do when we self-ascribe thought. To self-ascribe the thought <there are an infinite number of primes> a subject either judges that there are an infinite number of primes or that it is true or false that there are an infinite number of primes, and it is that what leads her to judge that she is presently thinking the thought <there are an infinite number of primes>.

If a subject judges that she is thinking that there are an infinite number of primes, based on it being the case that there are an infinite number of primes (or on it being true or false that there are an infinite number of primes), then it will indeed be the case that she is thinking that there are an infinite

number of primes. It is *guaranteed* to be the case that she is thinking that there are an infinite number of primes. The proposed process is preposterous yet infallible.

Despite that infallibility, it might be argued that, because subjects' beliefs about what they are presently thinking are formed via such an epistemically dubious process such beliefs, therefore, cannot possibly count as knowledge. It might be argued further that regardless of what else it may hope to achieve – i.e. regardless of whether it demonstrates how we can have privileged access without a priori access – an account of thought self-ascription must allow that we can *know* our thoughts and thus if under my account we would not know our thoughts, then my account is ipso facto a failure. The response to this objection is to query whether the notion that we do have knowledge of our thoughts is so irrefutable. It may be the case that it *is* irrefutable that we are able to *discern* our own thoughts in a way that others cannot. But, must it also be the case that such discernment should jump through all the necessary epistemological hoops required for it to qualify as knowledge? There is an overarching point here. Even if we are to maintain that my account would confer knowledge, then we must still concede that there is something odd about that knowledge. Such knowledge would be acquired via our making a preposterous inference. The overarching point is that if we are to accept my account then we must choose between either conceding that there is something odd about the knowledge that we have of our thoughts, or conceding that it is odd that the provenance of our thought self-ascriptions is such that it can't be said that we have knowledge of our thoughts. Either way, if we are to accept my account of thought self-ascription, then we must concede that thoughts, and the epistemic relationship we have with them, are – as already alluded to in the opening sentence of this thesis – indeed strange.

My account of thought self-ascription if accepted may affect how we countenance ourselves as thinking beings in other ways that I won't here try to articulate or explore further. It should suffice to say there is something very uncomfortable about my account of thought self-ascription. This does not mean that my account must therefore be untrue. An uncomfortable truth is still a truth. Or rather, just because an idea may be uncomfortable does not necessarily mean that it is therefore untrue.

In any case, regardless of this discomfort, we do have our desired account of thought self-ascription that allows that we can have a privileged access to our thoughts without that privileged access necessarily requiring us to have an a priori access to our thoughts. This in turn should allow for an acceptance of content externalism that is free of the worry that such an acceptance of content externalism might be incompatible with the claim that we should be able to discern what it is on our minds better than a third party ever could.

7.5: Loose Ends

7.51: *The Three Reductios*

In chapters four, five, and six we considered reductio arguments that purport to show that if it is true that we have an a priori access to our thoughts, and if content externalism were true, then our having an a priori access to our thought would entail our having an a priori access to the external world. It is (hopefully) clear that this worry vanishes if my account of thought self-ascription is accepted, as my account denies that we need to have an a priori access to our thoughts. (We have a privileged access to our thoughts without needing an a priori access to our thoughts.)

Furthermore, the consequents of the conditionals that power the three reductios are essentially the sorts of things that according to my account we judge to be the case in order to ascertain what we are thinking. In each reductio the protagonist is ex hypothesi granted the a priori knowledge that they are thinking a certain thought. The protagonist is then shown to have a priori knowledge of a certain conditional that has as its consequent some fact about the external world of which it would be absurd suppose that they could have a priori knowledge of. Thus, it is shown that if the protagonist does indeed have the ex hypothesi a priori access to their thought, then absurdly they will know a priori the consequent of that conditional. According to my account the protagonists of each of the three reductios would not have an a priori access to the respective thoughts. Furthermore, the way in which the protagonists would ascertain that they are thinking what they are thinking is through their judging to be the case what it so happens are the consequents of the conditionals that power each of the three reductios. My account transforms what are apparently troublesome reductios into harmless and trivial cycles of ratiocination!

Let's run through the reductios in reverse order. The reductio in Chapter 6 involves a person deducing that Aristotle exists through ex hypothesi knowing a priori that they are thinking the thought 'Aristotle was an alcoholic'. According to my account the person would not be able to ascertain that they are thinking 'Aristotle was an alcoholic' through a priori means. Instead, they would do so, if they were to do so via the contemplation method, through judging that Aristotle existed, and thus it is true or false of him that he was an alcoholic. Brown's reductio centres around the conditional:

S: Necessarily, if x has a thought involving a concept c , and x is agnostic about the application conditions of c , then either x is in an environment which contains instances of c and c is a natural kind concept, or x is part of a community which has the concept c , whether or not c is a natural kind concept.

The consequent of S is roughly the sort of thing that from x 's point of view would need to be the case in order for a proposition involving c to be true or false, and thus what x would need to judge to be the case in order to self-ascribe a thought involving the concept c . For a proposition to be true or false the relevant concepts must exist in the right sense, which is to say that the concepts must have

determinate application conditions. Because of *x*'s 'agnosticism' with respect to the application conditions of *c*, such a judgement will necessarily be about how the concept's application conditions are determined by her external physical and socio-linguistic environment. Finally, Boghossian's charge is that water would need to exist in order for there to be a concept 'water', in the way that a content externalist would understand it. Thus, if a subject is granted the a priori knowledge that she is thinking 'water is wet', she would be able to deduce via a priori means that water exists, which would be absurd. According to my account, such a subject would not ascertain that she is thinking 'water is wet' through a priori means. She would do so, if she went via the contemplation route, through judging that it is true or false that water is wet, which is to judge that there is a concept 'water'. If such a concept is inextricably tied to the natural kind water in the way that Boghossian contends that content externalists must suppose that it is, then the judgement will effectively be that water exists.⁶⁵

7.52: *The Problem of the 'Empty' Case*

Although my account seems to circumvent the three reductios, there is nevertheless a related problem that an adoption of my account may need to navigate, which is the problem of the 'empty' case. Suppose that I judge that it is true or false that Aristotle was an alcoholic and based on that self-ascribe the thought 'Aristotle was an alcoholic'. Imagine now a counterfactual in which there is no 'Aristotle', there isn't even a fictional creation or perceptual illusion going by the name of 'Aristotle' (see Chapter 6, §6.3), but in which everything seems to me as it actually is.⁶⁶ It would seem that the Kripkean externalist would need to accept an *Evansian/McDowellian* (Evans, 1982; McDowell, 1986) conclusion that within such a counterfactual scenario my self-ascription process would be illusory from start to finish. I could not be judging that it is true or false that Aristotle was an alcoholic, and I could not be thinking that Aristotle was an alcoholic. I would nevertheless believe that I am self-ascribing a bona fide thought. In the actual situation, in which Aristotle exists, I would also believe that I am self-ascribing a bona fide thought. In the actual situation that belief will be true, and crucially its

⁶⁵ It is conceivable that there may be – though I at present am unable to see it – a non-first-positionist account/conception of privileged access that like the non-first-positionist account outlined in the previous section allows for privileged access without a priori access – i.e. would envisage some sort of privileged access that does not necessarily have to be a priori. This rival account may be susceptible to the reductios in a way my account is not. Call the imagined method of thought self-ascription *Method M*. Suppose that it is through Method M that I self-ascribe the thought *water is wet*. The a priori reasoning that the protagonist in Boghossian's reductio (say) is meant to follow may then lead me to the conclusion that because I am thinking that *water is wet* water must therefore exist. Depending on the precise nature of Method M this may be deemed a problematic route to the conclusion that water exists, even though it won't necessarily be entirely a priori. If that is so, then that is all the more reason to accept my account of thought self-ascription over this imagined rival account, as, for the reasons stated, this won't be the case in the case of my account.

⁶⁶ One way of doing this would be to imagine that the universe is only a few seconds old – see Chapter 6, §6.3.

provenance is accounted for. It is a result of a chain of reasoning – or rather quasi-reasoning – that begins with my judgement that it is true or false that Aristotle was an alcoholic, that in turn leads me to judge that I am thinking the thought ‘Aristotle was an alcoholic’, and *that* in turn leads me to judge that I am thinking a bona fide thought. In the counterfactual situation the corresponding belief will be false, but that is not the problem. The problem is how do we account for the belief’s provenance. It cannot be due to the same chain of reasoning because in the counterfactual scenario in which there is no ‘Aristotle’ no such chain of reasoning can take place. Imagine three persons hanging from a cliff edge. The third person holds on to the second person’s ankles, who in turn holds on to the ankles of the first person, who holds on to the edge. If the first person is unable to keep their grip, all three persons fall to their deaths.

7.53: *Slow Switching*

Recall the case of *Slow-Switching Sally*. Slow-Switching Sally is routinely switched between Earth and Twin Earth, and consequently switches between thinking that water is wet to thinking that twater is wet. If Slow-Switching Sally were to self-ascribe the thought that water/twater is wet via either the conviction or contemplation methods, she would arrive at a guaranteed-to-be-true belief about what she is thinking, regardless of which planet she is on. Slow-Switching Sally, however, would not be able to discriminate between a situation in which she is on Earth and self-ascribes the thought: water is wet, and a situation in which she is on Twin Earth and self-ascribes the thought: twater is wet. In Chapter 3 of this thesis we considered whether cases such as those of Slow-Switching Sally challenge the claim that we should be able to know a priori what we are currently thinking. We assumed that we are able to form guaranteed-to-be-true beliefs about what we are thinking through a priori means, and we considered whether those guaranteed-to-be-true beliefs could count as knowledge in the light of slow-switch cases. Those same considerations should apply mutatis mutandis to the question of whether the guaranteed-to-be-true-but-in-the-case-of-the-sort-externally-individuated-thoughts-in-question-not-a-priori beliefs that the conviction and contemplation methods would yield could fully qualify for the esteemed designation ‘knowledge’ (if content externalism is/were true). That is, if the considerations alluded to towards the end of the last section don’t already disqualify such beliefs from qualifying for that esteemed designation.

7.54: *Comparison with Burge’s ‘Cogito-Like Judgements’ Account of Thought Self-Ascription: A Cliff Hanger Resolved*

My account of thought self-ascription is similar to Burge’s account of thought self-ascription (Burge, 1988 – see my Chapter 2, §2.2) in that both my account and Burge’s feature similar ‘ipso facto’

relations between a thought and a judgement. I judge that I am thinking that writing requires concentration therefore I am thinking that writing requires concentration, and therefore my judgement is correct, or so Burge's account goes. If I judge that writing requires concentration or that it is true or false that writing requires concentration, it will be the case that I am thinking that writing requires concentration. Thus, if my judgement that writing requires concentration, or that it is true or false that writing requires concentration, leads to me judging further that I am thinking that writing requires concentration, then that further judgement will be true, or so *my* account goes.

The key difference between my account and Burge's account is that my account explains the provenance of our thought self-ascriptions. It may well be true that if I judge that I am thinking that writing requires concentration then I will be thinking that writing requires concentration, but Burge's account does not explain why I'd judge that I am thinking that writing requires concentration in the first place. My account does. According to my account it is my judgement that writing requires concentration, or that it is true or false that writing requires concentration, that leads me to judge that I am thinking that writing requires concentration.

7.55: The Redeployment Thesis

What is known as the *Redeployment Thesis* (Peacocke, 1983; Sawyer, 2001; – see my Chapter 2) states that we should be able to form second-order beliefs about our first-order thoughts such that the content of such second-order beliefs should be *reflective* of the content of the first-order thoughts of which they are about in such a way that guarantees truth. The redeployment thesis requires there to be an appropriate belief-forming process or appropriate belief-forming processes by which a subject comes to form a belief about what she is presently thinking linking that belief to the thought that she is presently thinking. The conviction and contemplation methods are candidates for what those processes could be.

Conclusion to Thesis

If our having a *privileged* epistemic access to our occurrent thoughts required that we have an *a priori* access to our thoughts, then content externalism being incompatible with the claim that we do have an *a priori* access to our thoughts would pose a very serious problem for content externalism. It is possible though questionable that content externalism is incompatible with the claim that we should have an *a priori* access to our thoughts. However, it is not clear that our having a privileged access to our occurrent thoughts does require that we have an *a priori* access to our thoughts.

References

- Armstrong, A (1968) *A Materialist Theory of Mind*, Routledge.
- Bach, K (1987) *Thought and Reference*, OUP.
- Beaney, M (ed.) (1997) *The Frege Reader*, Blackwell.
- Beebe, H (2001) 'Transfer of Warrant, Begging the Question and Semantic Externalism', *Philosophical Quarterly*, 51 (204) pp.356-374.
- Bernecker, S. (1996): 'Externalism and the Attitudinal Component of Self-Knowledge', *Nous*, 30, pp.262–275.
- Boghossian, P (1989) 'Content and Self-Knowledge', *Philosophical Topics*, 17, pp.5-26 (also in Ludlow and Martin, 1998)
- _____ (1997) 'What the Externalist Can Know a Priori', *Proceedings of the Aristotlean Society*, 97, pp.161-175 (also in MacDonald et al, 1998).
- Bonjour, L (2003) 'Back to Foundationalism' in Bonjour, L and Sosa, E (eds.) (2003) *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*, Blackwell.
- Bridges, J (2006) 'Davidson's Transcendental Externalism', *Philosophy and Phenomenological Research*, 73 (2), pp.290-315.
- Brown, J (1995) 'The Incompatibility of Anti-Individualism and Privileged Access', *Analysis*, 55, pp.149-156 (also in Ludlow and Martin, 1998).
- _____ (1998) 'Natural Kind Terms and Recognitional Capacities', *Mind*, 107, pp.275-303.
- _____ (2001) 'Anti-Individualism and Agnosticism', *Analysis*, 61, pp.213-224.
- _____ (2004) *Anti-Individualism and Knowledge*, MIT Press.
- Brueckner, A (1992) 'What an Anti-Individualist Knows A Priori', *Analysis*, 52, pp.111-118 (also in also in Ludlow and Martin, 1998).
- _____ (2002) 'Anti-Individualism and Analyticity', *Analysis*, 62, pp.87-91.
- _____ (2004) 'Brewer on the McKinsey Problem', *Analysis*, 64, pp.41-3.

- _____ (2008) 'Wright on the McKinsey Problem', *Philosophy and Phenomenological Research*, 76 (2), pp.385-391.
- _____ (2011) 'A Defense of Burge's "Self-Verifying Judgements"', *International Journal for the Study of Skepticism*, 1:1, pp.27-32.
- Burge, T (1973) 'Reference and Proper Names', *The Journal of Philosophy*, 70, pp.425-439.
- _____ (1977) 'Belief *De Re*', *The Journal of Philosophy*, 74, pp.338-362 (also in Burge, 2007).
- _____ (1979) 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4, pp. 73-121 (also in Burge, 2007).
- _____ (1982) 'Other Bodies', in Woodfield, A (ed.) *Thought and Object*, OUP (also in Burge, 2007).
- _____ (1986) 'Intellectual Norms and Foundations of Mind', *Journal of Philosophy*, 83, pp.697-720 (also in Burge, 2007).
- _____ (1988) 'Individualism and Self-Knowledge', *Journal of Philosophy*, 85, pp.649-663, (also in Ludlow and Martin, 1998).
- _____ (1996) 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society*, 96, pp.91-116, (also in Ludlow and Martin, 1998).
- _____ (2007) *Foundations of Mind*, OUP.
- Carruthers, P (2011) *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, OUP.
- Chalmers, D (2003) 'The Content and Epistemology of Phenomenal Belief' in Smith, Q and Jokvic, A (eds.) *Consciousness: New Philosophical Perspectives*, OUP.
- Conee, E (1994) 'Phenomenal Knowledge', *Australasian Journal of Philosophy*, 72, pp.136-150.
- Davidson, D (1987) 'Knowing One's Own Mind', *Proceedings of the American philosophical Association*, (also in Ludlow and Martin, 1998).
- _____ (2001a) *Essays on Action and Events (2nd Edition)*, Clarendon Press.
- _____ (2001b) *Inquiries into Truth and Interpretation (2nd Edition)*, Clarendon Press.
- Davies, M (2003) 'The Problem of Armchair Knowledge' in Nuccetelli, 2003.
- Edgley, R (1969) *Reason in Theory and Practice*, Hutchinson.
- Evans, G (1982) *The Varieties of Reference*, OUP
- Fales, E (1996) *A Defense of the Given*, Rowman and Littlefield.
- Falvey, K (2000) 'The Compatibility of Anti-Individualism and Privileged Access', *Analysis*, 60, pp.137-142.
- Falvey, K and Owens, J (1994) 'Externalism, Self-Knowledge, and Skepticism', *Philosophical Review*, 103 pp.107-137.
- Feldman, R (2004) 'Foundational Justification' in Greco, J (ed.) (2004) *Ernest Sosa and His Critics*, Blackwell.
- Frege, G (1892) 'On Sinn and Bedeutung' in Beaney, M (1997) *The Frege Reader*, Blackwell.
- _____ (1918) 'Thought' also in Beaney, 1997.

- Fitch, G and Nelson, M (2018) 'Singular Propositions', *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/propositions-singular/>.
- Fumerton, R (1995) *Metaepistemology and Skepticism*, Rowman and Littlefield.
- Gallois, A (1996) *The World Without, The Mind Within: An Essay on First-Person Authority*, Cambridge University Press.
- Gerken, M (2008) 'Is Internalism About Knowledge Consistent with Content Externalism?', *Philosophia*, 36, pp.87-96.
- Gertler, B (2001) 'Introspecting Phenomenal States', *Philosophy and Phenomenological Research*, 41, pp.91-112.
- _____ (2011) *Self-Knowledge*, Routledge.
- _____ (2020) 'Self-Knowledge', *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2020/entries/self-knowledge/>.
- Gibbons, J (1996) 'Externalism and Knowledge of Content', *The Philosophical Review*, 105, pp.287-310.
- Goldberg, SC (2000) 'Is "I Am Presently Thinking That p" Self-Verifying?' Presented at the Central States Philosophical Association Meeting in Lincoln, Nebraska.
- _____ (2003) 'On Our Alleged A Priori Knowledge that Water Exists' *Analysis*, 63, pp.38-41.
- _____ (2005) 'An Anti-Individualistic Semantics For 'Empty' Natural Kind Terms', *Grazer Philosophische Studien*, 70, pp.147-168.
- _____ (ed.) (2015) *Externalism, Self-Knowledge, and Skepticism*, Cambridge University Press.
- Goldman, A (1977) 'Discrimination and Perceptual Knowledge', *Journal of Philosophy*, 73, pp.771-791.
- _____ (2006) *Simulating Minds*, OUP.
- Häggqvist, S and Wikforss, Å (2007) 'Externalism and A Posteriori Semantics', *Erkenntnis*, 67, pp.373-386.
- Horgan, H and Tienson, J (2002) 'The Intentionality of Phenomenology and the Phenomenology of Intentionality' in Chalmers, D (ed.) (2002) *Philosophy of Mind: Classical and Contemporary Readings*, OUP USA.
- Jeshion, R (ed.) (2010) *New Essays on Singular Thought*, OUP.
- Kallerstrup, J (2011) 'Recent Work on McKinsey's Paradox', *Analysis*, 71, pp.157-171.
- Kaplan, D (1989) 'Demonstratives' in Almog, J, Perry, J, and Wettstein, H (eds.) (1989) *Themes From Kaplan*, OUP.
- Korman, DZ (2006) 'What Externalists Should Say about Dry Earth', *Journal of Philosophy*, 103, pp.503-520.

- Kripke, S (1982) *Naming and Necessity*, Blackwell.
- _____ (2011) 'A Puzzle about Belief' in Kripke, S (2011) *Philosophical Troubles: Collected Papers, Volume I*, Oxford Scholarship Online.
- _____ (2013) *Reference and Existence: The John Locke Lectures*, OUP.
- Lasonen-Aarnio, M (2006) 'Externalism and A Priori Knowledge of the World: Why Privileged Access is Not the Issue', *Dialectica*, pp.433-445.
- Levine, J (2006) 'Phenomenal Concepts and the Materialist Constraint' in Alter, T and Walter, S (eds.) (2006) *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*, OUP.
- Loar, B (1976) 'Semantics of Singular Terms', *Philosophical Studies*, 30, pp.353-377.
- Locke, J (1689) *An Essay Concerning Human Understanding* (Historical text, multiple editions are available, no page numbers referenced).
- Lormand, E (1996) 'Nonphenomenal Consciousness', *Noûs*, 30 (2), pp. 242-261.
- Ludlow, P (1997) 'On the Relevance of Slow-Switching', *Analysis*, 57, pp.285-6 (also in Ludlow and Martin, 1998).
- _____ (2003) 'Externalism, Logical Form and Linguistic Intentions' in Barber, A (ed.) *Epistemology of Language*, OUP.
- Ludlow, P and Martin, N (eds.) (1998) *Externalism and Self-Knowledge*, CSLI Publications.
- Lycan, WG (1996) *Consciousness and Experience*, MIT Press.
- MacDonald, C (1998) 'Externalism and Authoritative Self-Knowledge' in MacDonald et al, 1998.
- MacDonald, C, Smith, BC, and Wright, C (eds.) (1998) *Knowing Our Own Minds*, OUP.
- McDowell, J (1986) 'Singular Thought and the Extent of Inner Space' in McDowell, J and Pettit, P (eds.) (1986) *Subject, Thought, and Context*, OUP.
- McKinsey (1991) 'Anti-Individualism and Privileged Access', *Analysis*, 51, pp.9-16.
- _____ (2007) 'Externalism and Privileged Access are Inconsistent' in McLaughlin, P and Cohen, J (eds.) *Contemporary Debates in the Philosophy of Mind*, Blackwell.
- McLaughlin, BP and Tye, M (1998a) 'Externalism, Twin Earth, and Self-Knowledge' in MacDonald et al.
- _____ (1998b) 'Is Content-Externalism Compatible with Privileged Access?' *Philosophical Review*, 107, pp.349-380.
- Moore, GE (1959) 'Proof of the External World' in Moore, GE (1959) *Philosophical Papers*, Allen and Unwin
- Moran, R (2001) *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press.
- Noordhof, P (2005) 'The Transmogrification of A Posteriori Knowledge: Reply to Brueckner', *Analysis*, 65, pp.88-9.

- Nuccetelli, S (ed.) (2003) *New Essays on Semantic Externalism and Self-Knowledge*, MIT Press.
- Peacocke, C (1983) *Sense and Content*, OUP.
- _____ (1996) 'Entitlement, Self-Knowledge and Conceptual Redeployment', *Proceedings of the Aristotelian Society*, 96, pp.117-158.
- Perry, J (1979) 'The Problem of the Essential Indexical', *Noûs*, 13, pp.3-21 (Also in Perry (1993) *The Problem of the Essential Indexical and Other Essays*, OUP).
- Pitt, D (2004) 'The Phenomenology of Cognition, Or, What Is It Like to Think That P?', *Philosophy and Phenomenological Research*, 69, pp.1-36.
- Putnam, H (1975) 'The Meaning of "Meaning"' in Putnam, H, *Philosophical Papers Vol. 2: Mind, Language and Reality*, Cambridge University Press.
- Quine, WVO (1951) 'Two Dogmas of Empiricism', *The Philosophical Review*, 60, pp.20-43 (also in Quine (1953) *From a Logical Point of View*, Harvard.)
- Recanati, F (1993) *Direct Reference: From Language to Thought*, Blackwell.
- Russell, B (1905) 'On Denoting', *Mind*, 14, pp.479-493.
- _____ (1910–11) 'Knowledge by Acquaintance and Knowledge by Description', *Proceedings of the Aristotelian Society*, 11, pp.108–128 (Also in Russell, B (1917) *Logic and Mysticism and Other Essays*, Allen and Unwin.)
- _____ (1912) *The Problems of Philosophy*, OUP.
- _____ (1921) *The Analysis of Mind*, Spokesman.
- Ryle, G (1949) *The Concept of Mind*, Penguin.
- Sawyer, S (1998) 'Privileged Access to the World', *Australasian Journal of Philosophy*, 76, pp.523-533.
- _____ (2001) 'In Defence of Burge's Thesis', *Philosophical Studies*, 107, pp.108-128.
- Salmon, N (1982) *Reference and Essence*, Blackwell.
- _____ (1986) *Frege's Puzzle*, MIT Press.
- Schiffer, S (2005) 'Paradox and the A Priori' in Gendler, TS and Hawthorne, J (eds.) (2005) *Oxford Studies in Epistemology Vol. I*, OUP.
- Segal, G (2000) *A Slim Book about Narrow Content*, MIT Press.
- Segal, G (2001) 'Two Theories of Names', *Mind and Language*, 16, pp.547–563.
- Soames, S (1987) 'Direct Reference, Propositional Attitudes, and Semantic Content', *Philosophical Topics*, 15, pp.47-87.
- Spicer, F (2009) 'On Always Being Right (About What One is Thinking)', *Canadian Journal of Philosophy*, 39, pp.137-160.
- _____ (2011) 'Two Ways to be Right about What One is Thinking', *International Journal for the Study of Skepticism*, 1:1, pp.33-44.
- Tye, M (1998) 'Externalism and Memory', *Proceedings of the Aristotelian Society*, 72, pp.77-94.

- Tye, M and Wright, B (2011) 'Is There a Phenomenology of Thought?' in Bayne, T and Montague, M (eds.) (2011) *Cognitive Phenomenology*, OUP.
- Warfield, TA (1992) 'Privileged Self-Knowledge and Externalism are Compatible', *Analysis*, 52, pp.232-7 (also in Ludlow and Martin, 1998).
- _____ (1997) 'Externalism, Self-Knowledge and the Irrelevance of Slow-Switching', *Analysis*, 57, pp.282-4 (also in Ludlow and Martin, 1998).
- Wright, C (2000) 'Cogency and Question-Begging: Some Reflections on McKinsey's Paradox and Putnam's Proof', *Philosophical Issues*, 10, pp.140-163.
- _____ (2003) 'Some Reflections on the Acquisition of Warrant by Inference' in Nuccetelli, 2003.
- _____ (2011) 'McKinsey One More Time' in Hatzimoysis, A (ed.) (2011) *Self-Knowledge*, Oxford Scholarship Online.

Novel Referred To

Stanislaw Lem, *The Futurological Congress*, trans. Michael Kandel, Harcourt.

End of Thesis

Word Count (including footnotes, excluding bibliography): 68,056

I, the author (Simon Stubbs), confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.