

Externalism, Self-Knowledge and Explanation

A thesis submitted in partial fulfilment of the
Requirements for the Degree of

Master of Arts

In

The Philosophy Department

Of

Rhodes University

By

Richard Flockemann

February 2006

Abstract

In recent years, much attention has been given to the question of whether content externalism is compatible with an account of self-knowledge maintaining that we have an epistemically privileged access to the content of our propositional mental states. Philosophers who maintain the two are incompatible (incompatibilists) have put forward two major types of challenge, which I call – following Martin Davies – the Achievement and Consequence Problems, which aim to demonstrate that self-knowledge cannot be reconciled with externalism. These challenges have spawned a great deal of literature, and a diverse range of arguments and positions have emerged in response. In this dissertation, I intend to focus on examples of these different avenues of response, and show how none of them are adequate.

In the first chapter, I lay the groundwork for the debate, setting up how externalism and self-knowledge are to be understood, and outlining both the incompatibilist challenges as well as the available responses to them.

In the second chapter I examine these responses in more detail, concluding finally that the best available response is Tyler Burge's. Burge has two arguments that together establish his compatibilist position. First, he shows that even if externalism is true, our judgements about our occurrent thoughts are *immune from error*. This establishes that our judgements about our thoughts must be true. Second, he offers a transcendental argument for self-knowledge, arguing that our access to our mental states must be not only true, but non-accidentally true, in a way sufficient for genuine knowledge. This establishes that we possess the correct epistemic entitlement to our thoughts.

In the third chapter, I argue Burge's arguments do not, in fact, give us good reason to suppose externalism and self-knowledge to be compatible. This, I argue, is because Burge relies upon a transcendental argument, which, in this context, cannot establish that we have self-knowledge *if externalism is true*. All it establishes, I argue, is that we do possess self-knowledge. And this is insufficient to establish that externalism and self-knowledge are compatible.

Acknowledgments

First and foremost I would like to thank my supervisor, Dr. Ward Jones, for his patience and his confidence in me, but especially for his penetrating criticisms, and his astute guidance in focussing both my writing and my reading. Without him this dissertation would not have been possible. I would also like to thank my parents for all their support (both emotional and financial). Thanks must also go to various friends and family, as well as the Rhodes University Philosophy department, for helpful discussions, assistance, and – more commonly – just listening to my often indecipherable ramblings on the subject. In this respect, special mention must go to Dr Samantha Vice and Mr Francis Williamson for their time, and their very helpful discussions.

Pencil, ink marks and
highlighting can confuse
for other readers.

Contents

Introduction	1
Chapter 1	2
(I) Self-Knowledge and Externalism	2
(II) Incompatibilism	6
(III) Incompatibilism and Explanation	13
(IV) Compatibilism	18
Chapter 2	23
(I) McLaughlin & Tye and Two-Factor Theories	24
(II) Brewer's Principle of Acquaintance	33
(III) Davies' Limitation Principle	37
(IV) Warfield and the Relevant Alternative Approach	44
(V) Burge, Immunity from Error and Critical Reasoning	52
Chapter 3	59
(I) Burge and Explanation	60
(II) Transcendental Arguments: General Problems	63
(III) Burge's Transcendental Argument: Specific Problems	68
(IV) Implications: Incompatibilism as Paradox	76
(V) Implications: Self-Knowledge and Externalism Revisited	80
Bibliography	85

Introduction

Over the course of this dissertation I will present my argument for why the compatibility of content externalism with a privileged self-knowledge has not yet been established.

In Chapter 1, I explain the notions of “privileged self-knowledge” and “content externalism”, and demonstrate why there is a *prima facie* concern over whether they can be understood to be compatible. I articulate the two major problems that incompatibilists have used to present this worry, and sketch the compatibilist responses to these problems.

In Chapter 2, I examine the compatibilist responses in a little more detail. Of the available responses, I argue that Burge’s is the most plausible.

In Chapter 3 I argue that Burge’s solution does not provide any reason to think that our judgements about our thoughts count as knowledge, rather than belief, if externalism is true. The argument Burge provides, I argue, demonstrates that we do indeed possess self-knowledge, but in order to establish a robust compatibilism on this matter, Burge needs to give an answer to the above question as well. This he does not do, and I thus conclude that it is to the providing of this sort of reason, or to examining the prospects of a non-externalistic account of mental content, that the debate should proceed.

Chapter 1

In this first chapter, my aim is to set up the debate regarding the compatibility of content externalism with an understanding of self-knowledge entailing that we have any kind of privileged access, or special authority, when it comes to knowing what we ourselves are thinking. I will proceed as follows: In the first section, I will explain what is meant or entailed by this sort of privileged self-knowledge (henceforth, self-knowledge) and by content externalism (henceforth, externalism). I will show that there is *prima facie* reason to suppose externalism is incompatible with self-knowledge. In the second section, I will sketch what I understand to be the two major problems that philosophers have relied upon to argue that externalism and self-knowledge are incompatible. I call these, following Martin Davies (2000), the Achievement Problem and the Consequence Problem. In my analysis, I will also draw out what these problems imply about knowledge in general. In the third section, I will examine why this debate is significant, and also what precisely would be required in an adequate solution to these problems. In the fourth section, I will sketch the various avenues of response to these problems, also drawing out the implications. By the end of this chapter, I hope to have established more clearly what is at stake in this debate, and what options philosophers seeking to establish the compatibility of self-knowledge and externalism (compatibilists) have available.

(I) Self-Knowledge and Externalism

What is meant by *self-knowledge* must be understood generally; there are various theories of self-knowledge available that understand our access to our own thoughts to be epistemically privileged in some way. At this stage I will not commit myself to any particular view. In Chapter 2, in my criticism of some compatibilist arguments, I will argue for a more particular understanding of self-knowledge, but I will not get into this yet. Put broadly, self-knowledge is understood to be different from our knowledge of the external world, in that it is epistemically privileged, as well as direct and non-inferential.

We can come to know what we occurrently believe, without having to provide or have available any evidence that would support this claim. When we make judgements

about what other people believe, we will have to pay attention to, and cite, behavioural evidence, things that this person has said or done. We do not need to cite such evidence in support of our judgements about our own occurrent beliefs. This is something it does not normally make sense to ask of someone. Further, it is understood that our judgements about our own beliefs are more reliable than our judgements about other people's beliefs. Our access to our own mental states is in this sense privileged – we can better know what we occurrently believe than anyone else can. In explaining self-knowledge in this way, there are thus two aspects that must be noted: first, our access to our own thoughts is *generally* more reliable than the access other people have to our thoughts. There may be cases where some other person – perhaps a psychologist – may be better placed to tell us what we are thinking than we are. However, it must be acknowledged that such cases are rare or unusual – *usually* we are better placed than anyone else to judge what we are thinking, and while there may be many thoughts we are unaware of, there are also many thoughts that we are knowledgeably aware of as well. The presence of unconscious desires and beliefs, or self-deception, does not detract from my ability to judge accurately and knowledgeably about many other of my beliefs and desires. I believe right now that the sun is shining outside my window, and I *know* I believe this. The presence of the unconscious, or even self-deception, ought not detract from my ability to know that right now, I believe that the sun is shining. The notion of epistemically privileged access at work in this debate is thus to be understood as being compatible with self-deception, and with holding that we have many thoughts of which we are entirely unaware, and with holding that we might in some circumstances be wrong when we make judgements about our thoughts. To accept that we have some sort of epistemically privileged access, I take it, is merely to accept that there are occasions when we can uncontroversially attribute knowledge of our mental states to ourselves.

Second: our access to our own thoughts or mental states *usually* is not inferential. Our knowledge of other people's mental states is inferred from their behaviour, but our knowledge of our own mental states is in some sense direct – not based upon examinations of what we have said or done. Again, if a psychologist explains to us that we have an unconscious belief or desire, we may indeed accept that she is correct, and come to hold that we do indeed possess such a belief or desire. But again, such cases are

not the norm – despite this, I do know that right now I believe the sun is shining, and I do not infer this from anything I have said or done, or anything someone else has said or done. My access to my thoughts, usually, is direct. It is *a priori*, in the sense that normally when we do have knowledge of our own mental states, this knowledge does not depend upon anything we could have inferred, or learned through sensory observation for its justification.

A final point, which will feature importantly in Chapters 2 and 3, is that an adequate account of self-knowledge must also hold that we know the representational or semantic *content* of our thoughts. Not only do I know that I think that the sun is shining, but also I recognise this thought as representing something about the world – that is, I recognise it as being a thought *about the sun*. There may be many interesting features to our mental states – the relation they bear to our brains, for example, is one much discussed feature. Yet this is not what is at issue for self-knowledge. Rather, it is this representational content of our mental states that an account of self-knowledge is properly interested in. When we talk knowledgeably about our thoughts in the privileged and direct manner discussed above, we are talking about our thoughts *as* things with representational content.

Like self-knowledge, externalism must be understood generally as there are different externalist theories of content. I will restrict my understanding of externalism to the general view that the Putnamian Twin Earth thought experiments indicate that the actual content of our thoughts, our actual psychological states, are at least in some cases dependent upon our relations to our environment. If we were related differently to the environment around us, then our thoughts in some cases would be different, even though our brains and bodies might be in precisely the same state.

The Twin Earth thought experiments, pioneered in Putnam (1975), work as follows: imagine there is a Twin Earth exactly like ours in everyway, except that in the oceans, lakes, puddles and taps on Twin Earth, instead of having stuff chemically composed of H₂O, they have stuff that is phenomenologically indistinguishable from water, but which in fact has a chemical composition of XYZ. People on Twin Earth call this stuff “water”. Now imagine that Oscar, who lives on Earth, has a molecule for molecule duplicate on Twin Earth, Toscar. When Oscar, who knows nothing about the

chemical composition of water, thinks the thought “water is wet” this is a different thought from when Toscar thinks it – because the term “water” *means* something different on Earth and on Twin Earth.

Originally, Putnam was making a claim about meaning – a claim belonging more in the realm of philosophy of language, rather than in philosophy of mind. Yet these thought experiments have been taken by many philosophers¹ to indicate that the actual content of our thoughts are externally individuated – what we are actually thinking is dependent in certain ways upon our environment. Not only does the word “water” mean something different when Oscar utters it, from what it means when Toscar utters it, when Oscar thinks “water is wet” he is thinking something different from what Toscar is thinking, when he thinks “water is wet”. Oscar’s claim, say, that “there is water in that glass” has different truth conditions from Toscar’s claim “there is water in that glass”. Oscar’s claim will be true iff there is H₂O in the glass, while Toscar’s claim will be true iff there is XYZ in the glass. This will be the case even if both of them *believe* their term “water” to apply to any clear, tasteless, non-toxic liquid. Thus the content of their thoughts is different, and this difference can be attributed to a difference in their environment.

I will present the incompatibilist challenge as applying to any theory of mental content that draws upon the intuitions brought out by these thought experiments – on my rendition, the incompatibilist is denying that we can know in a privileged way what the contents of our thoughts are, if the contents of our thoughts are externalistically individuated in the manner described above. The question, then, is whether the relation between world and thought contents, as it is understood in these thought experiments, allows that epistemically privileged access to the contents of certain kinds of our own thoughts is possible.

There is some *prima facie* reason to suppose externalism and self-knowledge are incompatible. The difficulty stems from the consideration that externalism entails that thoughts need to be individuated differently if certain extrinsic, environmental factors were to become different. Oscar and Toscar, remember, may both *believe* that when they are thinking about the stuff they both call “water” they are thinking about any clear,

¹ For a prominent and influential example, see Burge (1979).

tasteless, non-toxic liquid. Yet this is not the case. If Oscar were to somehow travel to Twin Earth, and strike up a conversation with a Toscar, Oscar might well believe that he and Toscar were talking about the same stuff when they were talking about “water”. Yet Oscar would be wrong about this. They are talking about different stuff. Oscar’s claim “there is water in that glass” will be false if the liquid in that glass is chemically composed of anything other than H₂O. Likewise, Toscar’s corresponding claim “there is water in that glass” is false if the liquid in that glass is composed of anything other than XYZ. Oscar and Toscar are having different thoughts – Oscar is thinking and talking about *water* (the stuff chemically composed of H₂O) while Toscar is thinking and talking about *twater* (the stuff chemically composed of XYZ). Whether Oscar is thinking a thought involving the concept “water” or “twater” is partly dependent upon what is in the world around him. If environmental conditions were to have been different, so too would have been the concept water that Oscar possesses. Whether Oscar possesses the concept “water” or “twater” thus seems to be knowable only in virtue of first knowing certain facts about his environment. This, then, is a *prima facie* worry regarding the compatibility of externalism and self-knowledge – what must be explained, therefore, is *how* Oscar can know he is thinking that there is *water* in that glass – without first discovering whether he is actually thinking of water or twater. In the next section, I will examine the two major ways in which incompatibilists have articulated worries of this kind.

(II) Incompatibilism

One way of articulating the incompatibilist worry is to raise what I will call the Achievement Problem – this maintains that, given the implications of externalism, the sufficient conditions for self-knowledge cannot be achieved. The following thought experiment is used to support this intuition: imagine that, unbeknown to him, Oscar is slowly switched between Earth and Twin Earth, so that he spends enough time in each place to acquire both concepts, water and twater. Now, it seems that when he says “water is wet”, meaning to express a belief he has about the world, the belief he will express at a certain time will be one belief he has (water is wet) and at another time will be another

belief he has (twater is wet). However, since he is unaware of the fact that he is switching between Earth and Twin Earth, he will not know this. In this case, the argument goes, Oscar could not be properly said to know whether he occurrently believes that water is wet, or twater is wet. This challenge was put forward first in Boghossian (1988).

In formalising this concern as an argument against compatibilism, I will follow Anthony Brueckner². On this formulation, the Achievement Problem poses a sceptical challenge to our knowledge of our thoughts, which may be presented as follows:

- A1) If I know that I believe that water is wet, then I must know that I do not believe that twater is wet.
- A2) I do not know that I do not believe that twater is wet.
- A3) Therefore, I do not know that I believe that water is wet.

In the above argument, the slow switch possibility is used to underwrite the argument's second premise – if Oscar was being slowly switched, then whenever he had a thought “water is wet” at some points he would be thinking that H₂O is wet, while at other points in his life, he would be thinking that XYZ was wet.

Note that the above sceptical argument against self-knowledge closely resembles the Cartesian sceptical argument against our knowledge of the external world. While there are a variety of ways of understanding how a sceptical argument is supposed to work, one understanding of a standard Cartesian sceptical argument may present it as follows:

- S1) If I know that I have hands, then I must know that I am not a brain in a vat
- S2) I do not know that I am not a brain in a vat
- S3) Therefore, I do not know that I have hands.

Like the Achievement Problem, this sceptical argument uses a thought experiment to underwrite the second premise – the sceptic invites us to imagine a world where we are indeed brains in a vat, having our sense organs simulated by electro-chemicals to produce

² See Brueckner (1990).

the sensations that we rely upon when we judge that, say, the sun is shining. Unless we can rule out the possibility that this is in fact the case, the sceptic claims, we cannot know that we have hands. Likewise, it seems, the incompatibilist is suggesting that unless we can rule out the possibility of a slow switch, we cannot know that we believe that water is wet – because we may in fact believe that twater is wet. The incompatibilist, in presenting this sceptical challenge, it seems, is endorsing standards for knowledge, which – if we generalised them – would render knowledge of the external world impossible. In order to be able to properly know that I am thinking that water is wet, the incompatibilist claims, I must be able to know that I am not thinking that twater is wet.

I will not be dealing explicitly or implicitly with the problem of scepticism here. I mention scepticism in this context only because it seems that given the sort of challenge being posed here, the problems the compatibilist faces may be analogous to the problem faced by anyone attempting to refute Cartesian scepticism. In this sense, the references I make to scepticism may be suggestive, but the strength of this analogy is not something I will explore. My comments on scepticism are used only to illustrate what I see as useful points for the debate surrounding externalism and self-knowledge, and should be treated as parenthetical.

At first glance it may seem that the sceptic sets an implausibly high standard for knowledge – given that if generalised, these standards imply that we do not know a great many things that we previously assumed we did know. What exactly the sceptic's criteria for knowledge are, is debatable; yet one understanding – which Barry Stroud thinks originates from Descartes – claims these standards are based on our everyday ascriptions of knowledge³. To borrow Stroud's example, imagine I see a yellow bird out my window. I have a limited experience of birds, but I come to believe that the bird outside my window is a canary. Now imagine that I can't tell the difference between canaries and goldfinches (in fact quite true). If the bird outside my window had been a goldfinch, and not a canary, I still would have formed the belief that there is a canary outside my window. Do I know that there is a canary outside my window? It is obvious that I don't know this. To suggest that this is something I might know seems incorrect – I am

³ See Stroud (1984A).

certainly not a reliable canary-detector of any kind, and there is no reason to suggest that my belief that there is a canary outside my window is anything other than mere belief.

How are we to explain why I don't know that there is a canary outside my window in this case? Stroud explains:

Reflecting... on the uncontroversial everyday examples alone can easily lead us to suppose that it is something like this: if somebody knows that something, p , he must know the falsity of all those things incompatible with his knowing that p (or perhaps all those things he knows to be incompatible with his knowing that p).⁴

Given that this is based on the way in which we ascribe knowledge to ourselves and others in everyday life, there is great intuitive support for this account. And if this is correct, then we can use the following argument to explain why I do not know that there is a canary outside my window:

- Ca1) If I know that the bird outside my window is a canary, then I must know that it is not a goldfinch.
- Ca2) I do not know that it is not a goldfinch.
- Ca3) Therefore, I do not know there is a canary outside my window.

Premise Ca1 is supported by Stroud's principle. If the bird was a goldfinch, then I could not be properly said to know that it is a canary. Thus I must know it is not a goldfinch, if I am to know the falsity of all the things incompatible with my knowing it is a canary. Premise Ca2 is supported by the fact that I can't tell canaries from goldfinches – that if it really was a goldfinch, I would still believe that it was a canary. This appears to be a good explanation of why I do not know that there is a canary outside my window.

The problem is that we can generalise cases like this so that they apply to all our knowledge claims regarding the external world. Since this argument works in the same way as the above sceptical argument, it seems that if it is correct, then so is scepticism. I understand this to be why scepticism is such a disturbing problem. To defeat scepticism

⁴ Stroud (1984A pp. 29 -30).

would entail showing either that these standards are wrong, or that they are not universal. That is, we can deny that this is why I do not know that there is a canary outside my window, *or* we can claim that Stroud's explanation is correct but only in that particular kind of case. I am not always required to meet these requirements in order to genuinely possess knowledge. But this second option means that to say that the incompatibilist is endorsing these requirements for self-knowledge is not to commit her to a general scepticism. The incompatibilist need not think these standards are universal – all she is committed to is the claim that these requirements must be met if we are to possess *self-knowledge*.

Nonetheless, there is reason to suppose these requirements may be considered too strict – since, if generalised, they lead to scepticism. Thus at this point we might already suspect that a key difference between incompatibilist and compatibilist on this matter might be that compatibilist supports lower requirements for self-knowledge than does incompatibilist. I will return to this thought briefly in Section IV, and again in Chapter 2.

After the Achievement Problem, the second problem incompatibilists use to generate a worry for compatibilism I call the Consequence Problem⁵. In this argument, the incompatibilist attempts to demonstrate that such a compatibilism would have absurd consequences. It offers the following argument:

- C1) If I am thinking that water is wet, then I must be in an environment that contains or did contain water
- C2) I am thinking that water is wet.
- C3) Therefore, I am in an environment that contains or did contain water.⁶

If compatibilism is true, then it seems that C1 and C2 are both knowable, if at all, *a priori*. Self-knowledge implies that I need not justify my judgements about my own thoughts by any *a posteriori* examination. I know I believe that the sun is shining,

⁵ To my knowledge, the first person to describe it as such was Martin Davies in Davies (2000).

⁶ In formulating the problem as simply as this, I am setting aside the considerations, originally used against Michael McKinsey who pioneered this challenge, that an externalist is not necessarily committed to C2. However, Jessica Brown's sophisticated rendering of the challenge, I think, can successfully and fairly obviously overcome that particular problem, so I will leave the matter aside for the sake of simplicity. See Brown (1995).

without having to investigate my environment. Externalism, also, is a thesis that could be known simply by armchair reflection – one need not necessarily have read Burge in order to arrive at those conclusions about mental content, one could conceivably arrive at them by *a priori* reflection. But then it follows that we could reason from C1 and C2, and thus come to have *a priori* knowledge of our environment. That is absurd – we can surely only have *a posteriori* knowledge of those kinds of specific facts about the external world.

Note that the above argument also relies on a principle philosophers have associated with scepticism – the idea that knowledge be *closed under known entailment*; that is, if I know that *p*, and I know that *p* entails that *q*, then I must be in a position to know that *q*. Philosophers like Fred Dretske have argued that denying such a principle applies to our knowledge of the external world is in fact the only way to refute scepticism⁷ – which suggests, for Dretske, given that scepticism is such an unacceptable position, that knowledge being closed under known entailment simply must be false in those sorts of cases. Again, this might perhaps be taken to suggest that the requirements for knowledge endorsed by the incompatibilist are higher than those supported by the compatibilist – the incompatibilist, it might be said, sees a failure of closure implying a failure of knowledge, whereas the compatibilist disagrees. I will return to this thought later.

There are other background assumptions at work in the Consequence Problem, which must be made clearer. Note that in order for this strategy to work, it must assume that to know that we are thinking water is wet, means that we must know that we are thinking the proposition water is wet, and that we are not thinking the proposition twater is wet. If Oscar is to know he is thinking water is wet, it is assumed that he must know the thought he is thinking is a water-thought and not a twater-thought, or any other thought. If it turned out that he did not need to know whether the thought he is thinking is a water- or twater-thought in order to know that he *what* thought he is thinking, then the Consequence Problem would not get far off the ground. For if he did not need to know whether he was thinking a water-thought or a twater-thought, or any other thought had in any other imagined environmental conditions, then what he would be able to infer about his environment, would be that his environment is one of an infinite number of different

⁷ Dretske (1999).

environments, consistent with the term “water” expressing a natural kind concept, or merely with Oscar *thinking* it expresses a natural kind concept. It would be consistent with Oscar being on any number of Twin Earths, or a Dry Earth, where there is in fact no water, but where the community of language-speakers collectively hallucinate that there is⁸, or any other possibilities of this kind. If this were to be all that Oscar is capable of inferring, then the conclusion here would be too vague to have any force. And if, as this would suggest, the conclusion is merely that he must be in an environment upon which having his thought “water is wet” depends, then the conclusion becomes trivial. Of course I can know by reflection that I am in the environment that I am, in fact, in. The incompatibilist, however, wishes the conclusion to be more specific than that. In this way, in offering this challenge, the compatibilist imagines that in order to genuinely know I am thinking water is wet, I must be able to know enough *about* that thought, to be able to infer something particular about the world in this way. What precisely the incompatibilist thinks I am required to know about water is not entirely clear, yet it seems it must be specific enough to infer specific facts about my environment. What the incompatibilist has in mind here, or how plausible this requirement is, is at this stage not clear, but I will leave the issue for now. In Chapters 2 and 3, I will revisit the question of what we need to know *about* our concepts in order to possess them, or in order to know we possess them.

In general, however, I think we can extract the following assumptions from the incompatibilist position: from the Achievement Problem, that the sceptic’s requirements for knowledge apply to self-knowledge, and from the Consequence Problem, that the relevant sort of knowledge is (philosophical reflection and self-knowledge) closed under known entailment, and that to know that I am thinking that “water is wet” requires me to know enough about the concept “water” to be able to infer something specific about the environment.

⁸ For a discussion of the Dry Earth possibility, see Boghossian (1998) and McLaughlin and Tye (1998).

(III) Incompatibilism and Explanation

In this section, I wish to consider the sort of question being asked by the incompatibilist in raising these two problems. It is important to note the ends to which the incompatibilist intends to use the Achievement and Consequence Problems, if we are to understand the kind of question she is asking, and subsequently the sort of answer that needs to be delivered. My discussion will also touch on the relevance of the debate.

First note that if it turns out that externalism and self-knowledge are indeed incompatible, the consequences would be significant. Brian McLaughlin and Michael Tye, for example, say that:

we find [privileged self-knowledge] sufficiently compelling that if Twin Earth externalism is incompatible with this privileged access thesis, that is a powerfully compelling reason for rejecting Twin Earth externalism. Indeed, were we ourselves to come to think that the incompatibilist view is right, we would reject Twin Earth externalism.⁹

That we have self-knowledge of this sort is not a universally held view. Certainly, there are deflationary accounts available, denying that our knowledge of our own mental states is genuinely special.¹⁰ Yet self-knowledge is certainly intuitively compelling – in everyday life, we operate under the assumption that we tend to know what we are thinking; our everyday ascriptions of thoughts to ourselves presuppose not only that we can know what we are thinking, but also that we can know it without having to provide any kind of evidence. If it turned out that we do in fact not have self-knowledge of this kind, this would constitute a profoundly deep sceptical challenge. Both Cartesian and Humean scepticism seeks to undercut our knowledge of the external world, yet these projects, if successful, would not cut as deeply as the sceptical challenge that would follow from the falsity of self-knowledge. Indeed, the Cartesian doubt relies upon us being able to know that we are thinking certain things – what it disputes is that we can

⁹ McLaughlin & Tye (1998, p. 289).

¹⁰ See, for example, Wright (1998).

know that the world around us corresponds to how we think it is. Without self-knowledge, however, it is not obvious how that question would even be possible. Without knowing what we are thinking of, it is difficult to imagine how we could begin to compare our thoughts to the world at all. Many philosophers reject Cartesian or Humean scepticism on the grounds that it simply *must be* wrong – its conclusion is one that we as agents are unable to take seriously¹¹. The idea that we know about the external world, or at least are justified in believing certain things about it, seems largely unshakeable. No matter how strong the sceptic's argument, we do not suspend our belief, for example, that the sun will rise tomorrow, or that it is shining now. The claim that we lack self-knowledge would be as difficult – if not more so – to take seriously in this sense. Just as the sceptical argument does not make me suspend my belief that the sun is shining, likewise, a sceptical argument against self-knowledge would not make me suspend my belief about what I am thinking; we make judgements about our thoughts all the time, and with unshakeable confidence. An argument claiming that I may well be wrong whenever I attribute thoughts or beliefs to myself would not have any affect on my beliefs about my thoughts. In this sense then, the claim that we lack self-knowledge is not easy to take seriously.

The fact that self-knowledge occupies such a central and undeniable role in our conception of ourselves, would, as McLaughlin and Tye suggest, spell trouble for externalism, if externalism and self-knowledge are indeed incompatible. Consider the following remark Paul Boghossian makes:

philosophers who embrace externalism don't do so because they regard it as a self-evident truth. They embrace it, rather, because their intuitive responses to a certain kind of thought experiment... appear to leave them little choice.¹²

If a consequence of externalism *were* that it rendered self-knowledge impossible, then externalists would find that they have conflicting intuitive responses, and, as such, a little more choice about whether to accept externalism or not. On intuition alone, certainly, it

¹¹ For a discussion of this, see Jones (2000).

¹² Boghossian (1998, p. 273).

appears easier to reject externalism than it does to reject that we possess self-knowledge. Disregarding the intuitive support for externalism, clearly, would be a lot easier if it did turn out to be incompatible with self-knowledge.

Yet abandoning externalism would be no small measure. As Burge has observed, in his view, externalism: “has carried the day among the vast majority of philosophers who have understood and reflected upon it”¹³. A great deal of work in philosophy of mind is accommodating to the externalist intuition – to name a few, Burge, Donald Davidson, Fred Dretske, Ruth Millikan, David Papineau, John McDowell and even Jerry Fodor, amongst many others, put forward theories of content that are in some sense externalist-friendly¹⁴. Externalism enjoys widespread support – abandoning externalism would fly in the face of a great deal of influential work in philosophy of mind, and should not be done lightly.

Now the fact that the conclusion that self-knowledge and externalism are incompatible would be so difficult to accept, may motivate some philosophers to dismiss or play down the importance of these challenges – given their conclusion, they might reason, these arguments must be wrong at some point. In reply to this manner of response to external world scepticism, philosophers like Stroud have stressed that we should not understand a philosopher endorsing the sceptical argument as attempting to demonstrate that we in fact lack knowledge of the external world. Rather, Stroud thinks, we should understand the sceptic to be someone already convinced that we do in fact possess knowledge of the external world, but who is uncertain as to how we can explain this, given the plausibility of scepticism.¹⁵ The sceptic does not need to be shown *that* we have knowledge, for Stroud; what must be shown is *how* we have it.

In a similar way, the plausibility of both externalism and self-knowledge should not motivate us to dismiss incompatibilism out of hand. Rather, we can understand incompatibilists not as attempting to show that we *do* lack self-knowledge, but, in the same way as Stroud understands the sceptic, to be asking a question about *how* we have self-knowledge, assuming externalism to be true. In support of this, Boghossian says:

¹³ Burge (2003B, p. 397).

¹⁴ For example, see Burge (1979), Davidson (1983), Dretske (1989), Fodor (1995), McDowell (1986) Millikan (1989), Papineau (1984).

¹⁵ Stroud (1989).

I consider the sceptical claim about self-knowledge to have the status of a paradox: apparently acceptable premises lead to an unacceptable conclusion. For I do not seriously envisage that we do not know our own minds.¹⁶

The argument does not intend to press the claim that we lack self-knowledge – that we have self-knowledge is something that an incompatibilist like Boghossian accepts already. His argument is thus not to be taken as an argument against self-knowledge, but rather as presenting a challenge to the compatibilist to explain *how* we have it. This construal seems to agree with Stroud, and closely parallels the way in which Crispin Wright understands the standard sceptical challenge; Wright says that the problem with scepticism (about the external world) is that it leads us to an absurd conclusion (like I cannot know I have hands) by means of a “seemingly well-motivated route”. According to Wright, an adequate answer to scepticism cannot merely offer reasons to suppose that knowledge in the domain in question is possible. For whether or not that knowledge is possible is not, or not all, the sceptic is asking, according to Wright. What is further required, he says, is that a satisfactory answer to a sceptical challenge must inform us of where we have gone wrong in our reasoning. A solution to a sceptical challenge must offer “a properly detailed diagnosis and exposé of its power to seduce”¹⁷.

The question being asked, then, is: how do we have knowledge of the content of our own occurrent thoughts, if externalism is correct, given the sceptical challenge raised by the Achievement Problem, and the *reductio* argument raised by the Consequence Problem? In this way, in responding to these problems, the compatibilist may not merely insist that self-knowledge must be possible – for it is not that question that is, or is only, at issue. If the challenges being posed are not expected to undermine self-knowledge, the question becomes one of explaining *how* exactly self-knowledge is possible, given the considerations that the incompatibilist offers.

What is required in an explanation of how self-knowledge is possible? Plausibly understood, to explain something, *p*, is to answer the question: “why *p* rather than *q*”? As

¹⁶ Boghossian (1988, p. 150).

¹⁷ Both from Wright (1991, p. 89).

Bas van Fraassen says, one needs a “rather” clause in an explanation to distinguish what is it about a proposition one wants explained: if we asking for an explanation of why Peter ate the apple, it is ambiguous what exactly we are looking for. Do we want to know why *Peter*, rather than Paul, ate the apple, or why Peter ate the *apple* rather than the orange? For clarification, then, one must add the “rather” clause¹⁸. This seems uncontroversial.

Now in asking for an explanation of self-knowledge and externalism, I take it that we must add in a further “if” clause. Thus the question being asked is: “why are our judgements about our own thoughts knowledge, rather than mere belief, *if externalism is true*?” If we can understand why we should take such judgements to count as knowledge rather than mere belief, then this, it seems, will explain how it is we possess self-knowledge if externalism is true. Thus the explanation can be provided by supplying reasons to suppose these judgements count as knowledge rather than mere belief, if externalism is true.

These reasons must also be stronger than the reasons to think our judgements to be mere belief rather than knowledge. If we have *no better* reason to think of our judgements as knowledge, than we do to think of them as mere belief, then I take it that we do not have a good reason to think of these judgements as knowledge. This might be seen as somewhat one-sided; since the incompatibilist needs only to draw, whilst the compatibilist must win. All the same, without the compatibilist’s reasons being stronger, I cannot see how we can take ourselves to have good reason to consider our judgements about our thoughts to be knowledge, rather than mere belief. If this is not obtained, it is difficult to see how an adequate explanation to have been given.

Note that if this is the case, and this is indeed what the compatibilist needs to provide, the incompatibilist is not necessarily *committed* to Stroud’s standards for knowledge, articulated earlier with regard to the Achievement Problem. The incompatibilist need not accept outright that these standards are correct. All the incompatibilist is asking is that *given the plausibility of those standards*, why are our judgements about our thoughts knowledge rather than belief, if externalism is true? Those standards provide some reason to think that our judgements, if externalism is true, are

¹⁸ For example, see van Fraassen, (1977).

mere belief. What is required in response is some reason to think that our judgements count as knowledge, if externalism is true. I will return to these matters in Chapter 3; in responding to Burge's solution to the Achievement and Consequence Problems, I will argue that Burge does not provide this sort of explanation of how self-knowledge is possible, and that his solution is for this reason inadequate.

In the next section, I will offer a cursory examination of the compatibilist responses to these challenges. I will not examine the plausibility of these responses here – that task I will leave until Chapter 2 – instead, I will merely provide an overview of the landscape; I mean only to sketch the available options, and their *prima facie* implications. I have chosen these particular responses because the *positions* they argue for seem to me to encompass all available positions one can take in response to the Achievement or Consequence Problems. There may indeed be other available *arguments* for these positions, but I can see no other available positions. As such, when I respond to these arguments in Chapter 2, I will focus my criticisms on the positions they adopt, rather than on their arguments for these positions. In this way, I hope to make my criticisms of compatibilism as general as possible.

(III) Compatibilism

There are two ways of responding to the Achievement Problem, and three ways of responding to the Consequence Problem. The first way of responding to the Achievement Problem is to accept that Oscar, if he were indeed to undergo a slow switch, would not know that he is thinking that water is wet. However, the response maintains, the possibility of a slow switch is too implausible to undermine self-knowledge – given that we are in fact *not* undergoing a slow switch, we still can be said to know that we are thinking that “water is wet”. Ted A. Warfield explains that indeed for Oscar, if he is being slowly switched between Earth and Twin Earth, then the possibility of error introduced in the Achievement Problem is a relevant alternative – and that the argument does establish that Oscar, in this case, does lack self-knowledge. What it does not establish, he argues, is that we, who are not undergoing a slow-switch, lack-self-knowledge in this way. What this argument establishes, for Warfield, is that it is not

necessary that any agent be able to know their thoughts non-inferentially – but that is not the same as the point that we *do not* know our thoughts non-inferentially¹⁹.

This argument, then, denies that Stroud’s condition must be met in order to be properly said to know something. We can, on this response, know that we are thinking that water is wet, even though we cannot rule out the possibility of a state of affairs obtaining (a slow switch) which would undermine our knowledge here. If we accept Warfield’s point here, then we deny that Stroud’s condition for knowledge is correct.

Burge has another response to the Achievement Problem. He points out that the second-order thought: “I am thinking that water is wet” is immune from error. That is, whenever you make a claim about what you believe in that sort of way, it is entirely impossible for that claim to ever be false. In this way Burge concludes that the negative force raised by the possibility of a slow-switch is dissipated. Thus the Achievement Problem does not generate a genuine threat to self-knowledge.

His argument runs as follows: *cogito*-like (judgements about what I am currently thinking) are never wrong, because thinking the second-order thought (“I am thinking that water is wet”) is possible to do only whilst thinking the corresponding first-order thought (“water is wet”) so that the concept (“water”) expressed in the first-order is conceptually redeployed to the second-order. In this way, whatever the term “water” means in the first-order thought, it will mean the same thing in the second-order thought as well. So whilst we might not be able to tell whether we are expressing one particular concept by our term “water” or another distinct concept, we can still be said to know that we are thinking that water is wet. So, the possibility of a switch does not impact on our ability to know what we are thinking²⁰.

So, on Burge’s account, even though we cannot tell whether we are having water-thoughts or twater-thoughts, we can still be said to know what we are thinking, when we say: “water is wet”. I will leave any further discussion of this response until the next chapter. My general conclusions from this preliminary sketch of the Compatibilist options will exclude Burge’s response; but I will discuss his argument at length in the following two chapters.

¹⁹ Warfield (1992, pp. 215 – 221).

²⁰ Burge (1988, pp. 111 – 128).

The Consequence Problem, recall offers the following argument as a *reductio* of compatibilism:

- C1) If I am thinking that water is wet, then I must be in an environment that contains or did contain water.
- C2) I am thinking that water is wet.
- C3) Therefore, I am in an environment that contains or did contain water.

Now one available response to this problem is to deny that C1 is knowable without empirical observation. McLaughlin and Tye respond in this manner – they claim that knowing I am thinking that “water is wet” is not sufficient to know that there is water in my environment. For all I know, they claim, I could be on Twin Earth, or Dry Earth, or any one of an infinite number of possible worlds. As a result, they argue, I cannot know, on the basis of my thought that “water is wet” that actual H₂O exists in my environment. They argue that Oscar may still know in a privileged way, that he is thinking that “water is wet” even though he may not know whether his concept is the concept “water” “twater” or any other Twin concept. This response, then, denies the incompatibilist view that in order to properly know that I am thinking “water is wet” I need know enough about a concept in order to be in a position from which I could use externalism to infer specific facts about the environment. McLaughlin and Tye thus do disagree with the incompatibilist that I am required to know enough about the application conditions of my concepts to be able to infer specific things about the environment, in order to genuinely possess that concept, or know I possess it²¹.

Another available response is to deny that premise C2, that I am thinking that water is wet, is knowable by anyone who is not already in a position to know C3, the conclusion. In this way, this strategy denies that the Consequence Problem offers us a means of acquiring new knowledge of the world, and thus denies that it can be successfully used as a *reductio*. This is Bill Brewer’s strategy²². However, if faced with the Achievement Problem, this strategy would have to adopt a response similar to that of

²¹ McLaughlin and Tye (1998, pp. 285 – 319).

²² Brewer (2000, pp. 415 – 433).

Warfield, which was mentioned above. If I am being slowly-switched between Earth and Twin Earth, I will not, I take it, be in a position to know that there is water in my environment. My environment will genuinely be changing every so often, so whatever empirical observations I have made in the past will not establish that there is H₂O around me *now*. However, Brewer maintains that if I am not in a position to know that there is water in my environment, then I cannot genuinely know that water is wet. Therefore, Oscar, if slowly switched, would not genuinely know that he is thinking that water is wet. In this way, Brewer, like Warfield, must deny Stroud's requirements for knowledge – which again indicates that the requirements for knowledge assumed here are less stringent than those assumed by the incompatibilist.

A third strategy is to accept both premise C1 and C2, but deny that having warrant to believe those premises gives us warrant to believe the conclusion. This is Martin Davies' strategy²³. Davies argues for a limitation principle for the transfer of epistemic warrant, so that, in arguments like the *reductio* argument used in the consequence problem, having warrant to believe each of the premises does not give us any reason to believe the conclusion. In this way, according to Davies, whilst I may know that I am thinking that water is wet, and I may know that if I am thinking that water is wet, then I am in an environment that contains or did contain water, but I may not use these known premises to deduce that I must be in an environment that contains or did contain water. If I am to genuinely know that, or be warranted in believing that, the warrant for that belief must come from somewhere else. In this way, Davies rejects the conclusion of the *reductio* argument – one cannot, if his limitation principle is correct, come to gain knowledge of the external world in the manner the Consequence Problem describes. It appears that Davies might be committed to the denial of deductive closure – seen superficially, his position seems consistent with the idea that someone could genuinely know that *p*, and genuinely know that *p* entails that *q*, but yet not know that *q*. If that is the case, then once more it seems that the requirements for knowledge adopted by the compatibilist are less stringent than the requirements adopted by the incompatibilist.

In this way, with the exception of Burge's response, a superficial examination of the various compatibilist avenues available all appear to presuppose lower requirements

²³ Davies (2000, pp. 384 – 414).

for self-knowledge than those presupposed by the incompatibilist. Incompatibilists, I argued, presuppose or in some way endorse the following closely-related requirements for knowledge: first, that I need to be able to rule out the possibility that any state of affairs obtains, under which I would fail to have self-knowledge, in order to genuinely have knowledge. Second, that if I genuinely know that p , and genuinely know that p entails that q , then I must in a position to genuinely know that q . Third, that in order to genuinely know I am thinking that water is wet, my understanding of what the stuff “water” I refer to, must be detailed enough in order to allow me to infer certain *specific* things about the world. Setting Burge aside, compatibilists, it seems, deny one or more of the above requirements.

Warfield and Brewer, it seems, must deny that in order to know I am thinking water is wet, I must be able to rule out the possibility of something like a slow-switch, the presence of which – they think – would undermine my knowledge that I am thinking that water is wet.

McLaughlin and Tye deny that I need to know specific details about the concept “water” in order to genuinely know that I am thinking “water is wet”. According to McLaughlin and Tye, we have seen, this is precisely the area where the incompatibilist goes wrong.

Davies, we have seen, appears to deny closure. If this is the case, then Davies denies that knowing that p , and that if p then q , entails that I be in a position to know that q . So, to revert to a standard sceptical example, to know that I have hands, does not require me to also know that I am not a brain in a vat.

In this way, a cursory examination of the compatibilist options appears to indicate that for the most part compatibilists are committed to lower requirements for knowledge than are incompatibilists. In the next chapter, I offer a more substantive examination of these strategies, focussing both on whether or not this is indeed an accurate assessment of their views, and upon whether their responses to the incompatibilist challenges are adequate.

Chapter 2

In this chapter I will proceed as follows: In Sections I- IV I will discuss the responses to the two incompatibilist challenges that I deem to be implausible responses.

Unfortunately, due to space constraints, I will not be able to give these solutions the attention they deserve – my responses to them will be regrettably quick. Nonetheless, I think I will be able to convey the problems with these positions. I will begin by discussing the three responses to the Consequence Problem, and then discuss the two responses to the Achievement Problem.

In Section I, I will argue that the response of McLaughlin and Tye appears to commit itself to a two-factor theory of mental content, or a two-factor theory of thought types – the difference between the two apparently being merely “verbal”. There are two problems with this: first, whether or not self-knowledge is compatible with a two-factor theory of content is distinct from the question of whether self-knowledge is compatible with *externalism*, so this does not address the incompatibilist concern. Second, there are independent reasons to think that two-factor theories, insofar as they maintain a commitment to externalism, cannot do the work McLaughlin and Tye require of them.

In Section II, I will argue that Brewer’s account is committed to a Warfield-like solution to the Achievement Problem.

In Section III, I will argue that Davies is forced either to agree with Warfield and Brewer on this matter, or is forced to deny that knowledge of this sort is closed under known entailment, and that this second option is not a plausible one.

In Section IV I will argue that Warfield is committed to an implausible account of self-knowledge.

In Section V, I will discuss Burge’s solution; which appears much stronger. Burge, I will argue, offers the most plausible response to incompatibilism. Of all these options, Burge’s strategy is the strongest and most promising. However, in Chapter 3, I will argue that Burge’s explanation of how we can have knowledge of the content of our thoughts does not give us reason to think that externalism is compatible with self knowledge.

(I) McLaughlin & Tye and Two-Factor Theories

I will begin by examining the compatibilist responses to the Consequence Problem, starting with the response of McLaughlin and Tye. McLaughlin and Tye's project is focussed on responding to the specific formulations of the Consequence Problem endorsed by Michael McKinsey, Jessica Brown and Paul Boghossian²⁴. I will not go into any detail regarding this exchange, rather I will move straight into their positive account.

McLaughlin and Tye deny that C1 is knowable from the armchair. They maintain that we cannot have privileged access to whether our thoughts are *singular propositions* (propositions that are dependent on features of the world or community in the sort of way that "water" is). If a singular proposition enters into the content of our thoughts, then this aspect of our thought content is not knowable to us. The idea of self-knowledge:

does not imply or presuppose that we have privileged access to the *singular contents* of such kind-dependent thoughts. Indeed, there can be illusory kind-dependent thoughts. The thought that phlogiston caused the fire would be an example: a thinker of the thought might be under the cognitive illusion that it is a kind-dependent thought.²⁵

In this way then, on their view we do not have any sort of privileged access to the singular contents of our kind-dependent thoughts, like our water thoughts, and this is because, it seems, of the possibility of illusory kind-dependent thoughts. Because it is always possible that our thoughts do in fact not have singular contents, or have different singular contents, we cannot have privileged access to this sort of content. The possibility of alternative singular contents, or illusory singular contents, which we cannot rule out, undermines our ability to know our singular contents. But what, then, is it that we have privileged access to, if not the singular content of our thoughts?

McLaughlin and Tye acknowledge that mental states can be individuated in any number of ways. We could, presumably, type mental states according to their actual

²⁴ See McKinsey, (1991), Brown (1995), Boghossian (1998), and McLaughlin and Tye (1998).

²⁵ McLaughlin and Tye (1998, p. 292). Emphasis mine.

content, or according to their physical realisers, or according to the functional role they play in the thinker's cognitive processes. Yet when we are to talk knowledgably about our own thoughts, in the manner described in the previous chapter, what we can know about our thoughts will be limited. I could not know without empirical research, for example, that my thought that the sun is shining outside correlates with a particular brain-state I am in (assuming that it does). When using our privileged self-knowledge, we are only warranted to speak of our thoughts under certain descriptions of those thoughts – not all descriptions. Our thoughts do possess properties about which we cannot speak of knowledgably without empirical research. So under what descriptions of our thoughts are we entitled to speak of them? According to McLaughlin and Tye, self-knowledge is concerned only with mental states individuated from one another:

in as *fine-grained* a way as is necessary for the purposes of *any* rationalising explanation. Whether typing occurrent thoughts by their contents will suffice for typing them in a way fine-grained enough for such purposes will depend on what sorts of contents are in question²⁶.

They also hold that individuating thoughts in terms of their actual content, when something like a natural kind concept is evoked, does not type them in as fine-grained a way as is necessary for rationalising explanation. Thinking “water is wet” might not always be the same as thinking that “H₂O is wet” – Oscar may believe that the first of these claims is true, but perhaps not the first. As McLaughlin and Tye say:

the mental state type *thinking that water is wet* can play a different role in rationalizing explanation from the role played by the mental state type *thinking that H₂O is wet*. They are distinct mental state types. One can be in a state of the first type without being in a state of the second type²⁷.

²⁶ *Ibid*, p. 293.

²⁷ *Ibid*, p. 294.

The problem stems from the fact that one might not *know* that “water” and “H₂O” in fact refer to the same stuff. If one *treats* the terms “water” and “H₂O” differently, if one believes that they do refer to different stuff, then it seems that each of these terms will play different roles in a rationalising explanation of that person’s behaviour – the fact that it was in fact H₂O in Oscar’s glass, might be irrelevant in explaining why he decided to drink from the glass – if he was unaware that the stuff he calls “water” actually is made of H₂O. McLaughlin and Tye expressly claim that we have no privileged access to the singular content of our thoughts. They say there is:

no implication that one can know *a priori* that one’s thought that *p* has a singular content or that it has a content into which a singular proposition enters as a component.

And further:

Were an incompatibilist to insist that privileged access [to our thoughts] should be understood in such a way as to imply that we have privileged access to whether our occurrent thoughts have associated singular propositions, then the proper response would be to reject the absurdly strong privileged access thesis.²⁸

From the above quotes, it seems McLaughlin and Tye think that we have privileged access to our thoughts only when typed in such a manner that does not reflect the singular content of such thoughts. Since we cannot know *whether* our thoughts have singular content, on their view, if we are to type our thoughts in such a way that they can be knowable to us, then we must type them in terms of something other than their singular content. What the singular content of our thoughts is, they claim, is not knowable to us. To maintain that we can have privileged access to the singular contents of our thoughts is “absurdly strong”.

On reflection, we can agree that self-knowledge is concerned with our mental states individuated in the manner that will be important for rationalising explanation. As I

²⁸ *Ibid*, p. 295.

will argue in Sections III and IV, one of the reasons we want self-knowledge in the first place is because our ability to function rationally and to critically engage with our thoughts requires the ability to know our own thoughts. To offer a rationalising explanation is to offer an explanation in terms of reasons, beliefs and desires. Yet for a mental state to genuinely give me a reason to perform an action, in the manner a rationalising explanation would require, then I must recognise that mental state as something that gives me a reason, and this means I must recognise it as something possessing *content*. If I am not acting *because* that mental state gives me a reason to act then offering a rationalising explanation will not be an appropriate explanation of my behaviour. To behave rationally is to recognise reasons as reasons. What makes a mental state a reason for an action is its content – what it represents in the world. To offer a rationalising explanation is to presuppose the agent recognises reasons as reasons – and this requires some kind of awareness of content. To borrow an example from Terence Horgan, if I fetch a beer from my fridge, a rationalising explanation of this act will offer the fact that I desired a beer, and believed there to be one in the fridge, as an explanation. Now if I did not myself recognise my mental state as *representing* that there was a beer in my fridge, then offering a rationalising explanation will no longer seem appropriate. If I do not recognise that my having this mental state *means* that there is a beer in my fridge, then we would more accurately explain my action mechanistically, in terms of cause and effect, rather than offering a rationalising explanation. Horgan says this of our intuitions on this matter:

Our common-sense belief about our own actions – a belief with enormous force and vivacity, by virtue of the phenomenology of our own agency ... – is that the *content* of the operative belief and desire has an ... immediate kind of causal/explanatory relevance to the action.²⁹

Generally speaking, then, for a rationalising explanation to be appropriate requires that I take my mental state to have semantic or representational content; I have a reason to go to

²⁹ Horgan, (1991, p. 88) Emphasis mine.

the kitchen only *because* I understand my mental state to represent that there is a beer in my fridge.

But according to externalism, the content of some of my mental states will be singular. An externalist, recall, claims that certain thoughts are object- or kind-dependent; that is; that the content of certain kinds of thoughts are partly dependent upon certain features of the subject's environment. Hence, the externalist maintains that certain of our thoughts have singular contents. But this raises the question: how then can McLaughlin and Tye deny that we have privileged access to the singular contents of our thoughts? Since knowledge of the content of our thoughts seems crucial to the appropriateness of a rationalising explanation, this implies that we only have access to some *other* non-singular content that such thoughts possess. How is that something an externalist can accept? Burge describes his externalism, which he labels *anti-individualism* rather than externalism, as follows:

anti-individualism is... a thesis... about intentional mental states. It is about the constitutive or individuating conditions of propositional attitudes, perceptual states, the *having* of concepts, the applications of demonstratives, and so on.³⁰

Externalism for Burge thus applies to actual attitudes, to the very manner in which we take the world to be; if environmental conditions are different, for Burge, then so too may be a person's *attitude*, the *very way that person actually takes the world to be*. The rationalising explanation of a person's behaviour, his attitude towards the world, must for Burge be externalistically individuated – any sort of “individualism” about mental states he rejects. Thus, if Oscar decides to drink a glass of water, and Toscar decides to drink a glass of twater, Burge holds that there is in fact a *different* rationalising explanation of Oscar's drinking a glass of water, than there is from the rationalising explanation of Toscar's drinking a glass of twater. These are separate actions for Burge, with separate

³⁰ Burge, (2003A, p. 350).

reasons behind them. Thus, to type thoughts in as fine-grained a way for rationalising explanation is to type them in terms of their genuine representational (singular) content.³¹

Clearly then, McLaughlin and Tye must disagree. But then how can they maintain a commitment to externalism and self-knowledge whilst maintaining that we have no privileged access to the singular contents of our thoughts? An obvious way of responding to this worry, would be for McLaughlin and Tye to commit to a two-factor theory of thought content or of thought types. If a thought is a singular proposition, then *that* it is a singular proposition is not knowable *a priori* to a person thinking that thought. What is knowable, this response says, is some other content that the thought has, or its mode of presentation, *what* I take its content to be. This is not an implausible thing to commit McLaughlin and Tye to; for they say:

It is fairly common for philosophers who maintain that thoughts have singular propositions as their contents to maintain two-factor theories of thought types. On this view, it is not true that no two thought types can have exactly the same content. What is true, on this view, is, rather, that no two thought types can have exactly the same content and same mode of presentation of that content. Some philosophers, however... embrace a two-factor theory of *content*, with singular propositions as sometimes one of the two factors and a mode of presentation as the other³².

In this way then, McLaughlin and Tye can comfortably offer a two-factor theory as an explanation of how we can explain the subject knowing *something* about her thought, when she cannot know its singular content. So, for McLaughlin and Tye, when typing thoughts in the way required for rationalising explanation, requires we type them in terms of their mode of presentation, rather than their singular content.

To hold a two-factor theory of thought types, as the name suggests, is to hold that in distinguishing one thought type from another, there are two factors that must be taken

³¹ For other philosophers who endorse the above kind of claim about content, see Brewer (2000), McDowell (1986) and McCulloch (1995). Brewer, as we will see in the next section, argues directly that the external-content determining factors must be reason giving.

³² *Ibid*, p. 294.

into account; first, its singular content, and second, its mode of presentation, how the subject presents the thought to herself. If either of these are different, according to this view, the thought type will be different. To speak of a two-factor theory of thought type is to embrace something very like the distinction between “wide” and “narrow” content. Like a two-factor theory of content, there are two content-determining aspects; its singular or wide content, and its mode of presentation, which can be understood in much the same way as narrow content.³³

However, it is important to note that they do not expressly maintain that this is the *only* way to explain what we have privileged access to; it is merely “fairly common”. My committing them to this position is thus undersupported. Yet if they are not committed to a two-factor view, they must explain how we can know our thoughts *without* knowing what the singular contents of our thoughts are. For this, as I’ve shown, is what they are explicitly committed to. It is not clear how they could precede in this – since they expressly maintain that we have no privileged access to *singular* content, and yet think that we can nonetheless know our thoughts typed in the way required for rationalising explanation. Typing them in that way, I argued, is to type them in terms of content³⁴.

In this way, it seems *either* they require a two-factor theory of some kind, or they need some explanation of how typing thoughts in the way required for a rationalising explanation is not to type them in by means of some semantic or representational content, or even by the mode of presentation. Embracing a two-factor theory here seems the only feasible option.

But to respond to the incompatibilist in this way does not answer the question. Incompatibilists are not addressing two-factor theorists. If one maintains that thoughts have “wide” and “narrow” contents, and that we have privileged access only to “narrow” contents, then the problems the incompatibilist raises will obviously not apply – since both of these focus on the “wide” nature of content to cause difficulties. The Achievement Problem argues that the possibility of unnoticed environmental changes

³³ McLaughlin and Tye do themselves not understand the distinction between two-factor theories of thought types and two-factor theories of content, understanding the distinction to be merely “verbal”. See McLaughlin and Tye (1998 p. 294). As such, I think my conflating mode of presentation with narrow content, as I do above, is legitimate in this context.

³⁴ Or at the very least to type them in the way we take the content to be; by its narrow content, or mode of presentation. I accept that these would suffice for rationalising explanation.

threaten our ability to know our thoughts. This clearly, can only apply to thought content that depends upon the environment. If we are talking about thought content that does not depend upon the environment in this way, this possibility would be no threat to knowledge. The Consequence Problem can only work if thought content *implies* particular environmental conditions. Once again, this can only apply to thought content that does indeed have these implications. The incompatibilist wants to know how a theory of content that has these implications can be compatible with self-knowledge. To say we can know only the contents of our thoughts that do not have these implications, then, is to *agree* with the incompatibilist, rather than answering their question. To say we know our thoughts only when described in such a way as to leave out their wide or representational content does not answer the incompatibilist concern. That we can know our thoughts in that way, the incompatibilist would accept as obvious. However, then the question is how do we maintain compatibilism *without* evoking a two-factor theory. Since McLaughlin and Tye do not think we can have privileged access to the “wide” or singular content of our thoughts, their response is no use in answering this question. What is needed then, if this avenue of response to the Consequence Problem is to succeed, is an explanation of how the sort of privileged access we have to our mental states can be understood, given it does not entail we know whether or not our thoughts are singular propositions, without resorting to a two-factor theory. In Section V, I will argue that Burge provides a far more plausible answer to this question. On Burge’s view, as I will show, we can indeed have privileged access to the singular contents of our thoughts.

However, even allowing the resort to a two-factor theory in this way does not remove difficulties – this resort, as Michael McKinsey points out, is not enough to rescue McLaughlin and Tye from violating *either* externalism, or self-knowledge.³⁵ McKinsey criticises McLaughlin and Tye’s recourse to a two-factor view by pointing out that a two-factor view maintains that a thought’s content is determined by *both* wide and narrow content. Both are important, on this view, in individuating a thought. Thus, McKinsey claims, if McLaughlin and Tye maintain we can know only the narrow content, and not know the wide, then it seems we do not know the actual content of the thought we are having; since both are essential to that thought having its content. According to two-

³⁵ McKinsey (2001).

factor theories, a thought's content has two essential properties; its wide content, and its narrow content, or mode of presentation. Not knowing one would be sufficient to undermine our knowledge of the overall content. McKinsey says:

For on the two-factor view, a sentence like [I am thinking that water is wet] will ascribe a type T of thought that either involves both factors, or one of the two. If T involves both the singular proposition [that water is wet] and a mode of presentation of that proposition, or T involves just the singular proposition alone, then it follows that [I am thinking that water is wet] says or logically implies that [I have] a thought whose content is the singular proposition [that water is wet]. But of course, McLaughlin and Tye explicitly deny this consequence.³⁶

McKinsey's point is that even on a two-factor view there is no escaping the fact that singular propositions enter into the content of a thought – the content of our thoughts, on this view is made up of *both* singular content and mode of presentation, or “narrow” content. Thus, to know the content of our thoughts would imply that we know the singular contents as well, since the content of that particular thought, which we are said to know, contains singular content. And if this is knowable, he says us, then the Consequence Problem ought to succeed. McLaughlin and Tye's denial that singular content is knowable in a privileged way thus denies we can know the content of our thoughts, even if we allow for two-factor theories.

The problem is that according to McLaughlin and Tye, what we can know about our thoughts is only – even on a two-factor view – one of two essential content determining aspects. The other content determining aspect is not knowable. If they are committed to self-knowledge, then they must deny that the “wide” content plays a role in the determining of our thoughts. This denies externalism. If they are committed to externalism, then they concede that we cannot know part of our thought's content. But if we cannot know anything about a feature of our thought that determines its content, then, that implies we cannot know what our thought's overall content is. This denies self-knowledge.

³⁶ McKinsey (2001, p 40).

What McLaughlin and Tye need, in order to remain committed to both self-knowledge and externalism, is an argument that denies that we need to rule out alternative possibilities of different singular contents in order to know the contents of our thoughts. It is difficult to see how they could argue for this, whilst maintaining, as they do, that we lack knowledge of the singular contents of our thoughts *because we cannot rule out the possibility of illusory singular contents*.

In Section V, I will show that Burge has available precisely this kind of argument, and so I will postpone further discussion of this possibility until then. Yet, as I shall show, on Burge's view, or the view that I will attribute to Burge, the singular contents of our thoughts are indeed knowable. Conceding that we cannot know the singular contents of our thoughts, as McLaughlin and Tye do, is conceding too much.

(II) Brewer's Principle of Acquaintance

Another possible response to the Consequence Problem is to deny that C2 (I am thinking that water is wet) is knowable to anyone who does not already know the conclusion. Thus the Consequence Problem does not provide us with any new knowledge of the world. In Brewer (2000), Bill Brewer puts forward this solution. The thought "water is wet", he argues, is impossible to have without the subject having "demonstratively based knowledge"³⁷ that water exists.

The argument runs as follows. Brewer argues first that all externalist requirements are a consequence of the following principle, which he thinks is a "version of Russell's Principle of Acquaintance"³⁸. The Principle (henceforth A) claims:

A persons' capacity to make determinate reference to certain objects and kinds in belief depends upon his having *demonstratively based knowledge* about them.³⁹

Brewer then goes on to show that if this is true, the Consequence Problem dissolves, because Brewer's principle, A, contends that it is only possible to have a belief *about*

³⁷ Brewer, 2000 p. 421.

³⁸ *Ibid*, p. 421.

³⁹ *Ibid*, p. 421.

water, if we have demonstratively based knowledge of water. So, he claims, we cannot come to new, non-empirical, knowledge of the world this way, as knowing C2 is only possible for someone with the wherewithal to arrive at knowledge of its conclusion⁴⁰. This knowledge, Brewer suggests, will come from something like reliable perception, or testimony, or something of the kind.

Now Brewer's argument for A runs as follows. The crucial claim for A is that:

externalist relations are necessarily *reason-giving* relations, constituting a source of demonstratively expressible knowledge; where by this I mean reason-giving *from the subject's point of view*, rather than from the perspective of some external theorist.⁴¹

Now of course externalism is the view that it is the causal/perceptual relations between subject and world that give certain of his beliefs, or thoughts, their content – that is, what makes it possible for us to say that this belief is about *this* thing, and not the other thing. Now Brewer imagines we conceive of these relations as not reason-giving in his sense, but merely causal. Now if a person could have two possible contents for his beliefs, *x* and *y* then if these contents are not reason-giving then “a person's causal-perceptual relations with the things around him give him no more reason to believe that *x* than to believe that *y*”⁴².

If the conditions that lead to Oscar's belief that water is wet are merely causal, and not-reason-giving then the following will be true. In this case, Oscar would have been given no more *reason* to believe that water is wet, than he has to believe that twater is wet. Thus he does not and could not have any reason to believe that water is wet *as opposed* to twater is wet – precisely because the circumstances that determine the content of his beliefs do not give him a reason to pick one over the other. As a result, Brewer says, if this were the case, then believing that water is wet and believing that twater is wet are identical to Oscar – since neither belief is different for any reason, he will treat them as precisely the same. From this observation, Brewer claims:

⁴⁰ *Ibid*, p. 417.

⁴¹ *Ibid*, p 421.

⁴² *Ibid* p 422.

Hence the supposedly content-determining role of [Oscar's] environmental embedding is empty. For there is nothing more, or less, to the content of a belief than the way the subject takes the world to be. Thus, if the proposed causal-perceptual relations in which a person stands to certain mind-independent things are not reason-giving relations, then they contribute *nothing* to the determination of specific worldly truth-conditions for his empirical beliefs about such things... the content-determining relations between a person and certain things ... which are posited by the content-externalist are necessarily reason-giving relations.⁴³

To maintain that the relations that fix the content of a belief are not reason-giving leads to the following situation:

A person believes that *p* and does not believe that *q*, even though he has, and could have... no reason to believe that *p* as opposed to believing that *q*... It follows from this that the theorist in question is committed to a conception of belief content which is more discriminating than the subject's own understanding of the contents of his beliefs. For the theorist is obliged to distinguish *p* and *q*, even in the face of the fact that they are absolutely on par as far as the subjects actual or possible reasons for, or against, endorsing them are concerned... This is surely unacceptable by anyone's standards. For the content of a belief is precisely... the way the subject takes things to be.⁴⁴

So now Brewer has claimed that (from externalism) a person's causal/perceptual relations to the things in the world around him determine the content of some of his beliefs, and they do this by giving the person *reasons* to believe that *p* (that water is wet) rather than *q* (that twater is wet). The only way this could be possible is if this person could first be said to know that there is water around him – to know there is water, and not twater. From the fact that he is embedded in his particular environment, Oscar must be able to

⁴³ *Ibid*, pp. 422 – 423.

⁴⁴ *Ibid*, p. 424.

know that water exists, and this gives him a reason to believe that water is wet, and not that twater is wet. Note, again, that on Brewer's account, nothing contributes to Oscar's justification for his belief that he believes water is wet other than reasons Oscar has for believing water is wet. Oscar can know what he believes, presumably, by consulting what he has reason to believe, so I take it that this account does not threaten privileged access in any way.

Now the first point that must be noted is Davies'⁴⁵ – that is that even if this is right, it still must be explained why this argument could itself not add *further* warrant to my belief that there is water in the world. Because even if it is possible to know C2 only if we already are in a position to know C3, it would still seem implausible to cite the argument used in the Consequence Problem as a further warrant for C3. So it must be explained how it is that someone knowing C1 and C2 could not – without falling back on the knowledge he's acquired through perception, or testimony, etc. – validly conclude that C3. So this solution does not solve all the problems unless used in conjunction with something like Davies' Limitation Principle, which will be discussed next, or some other solution to the Consequence Problem. If this solution is to work, then there is a lot more that needs to be explained.

As shall become clearer later, there is also a deeper concern with this solution, insofar as it commits itself to answering the Achievement Problem in the same general way that Warfield does. I will deal with this strategy at a later point, for now, I shall restrict my argument to establishing that Brewer must be committed to a Warfield-like solution.

Consider once more the slow-switching case endorsed by the Achievement Problem. As this example demonstrates, while being slowly switched between Earth and Twin Earth, Oscar would *not* be able to genuinely know that there is water in his environment. That much seems clear – a person being switched in that way cannot know whether there is H₂O or XYZ around him at any time. The reliable perception, which, on Brewer's account, is supposed to give Oscar knowledge of water in his environment, is in this particular case no longer reliable, and thus insufficient for knowledge. In that case, from Brewer's response to the Consequence Problem, Brewer must conclude that Oscar

⁴⁵ Davies (2000, p. 388).

cannot know that he is thinking that water is wet; since that is only possible, he argues, for someone able to know that there is water in his environment. Since, on anyone's account of perceptual knowledge, the reality of a slow switch can be understood to undercut our ability to know that there is water in our environment, Brewer must concede that if a slow switch were actual, Oscar would not have the wherewithal to arrive at the conclusion "there is water in my environment". At this point, Brewer has two options. First he can claim that in this case, Oscar nonetheless can know that he is thinking that "water is wet". But then this serves as a counter example to his conclusion that it is possible only for a subject to think that "water is wet" if she possesses the wherewithal to arrive at the conclusion "there is water in my environment". This would undermine this as a possible solution to the Consequence Problem. Alternatively, Brewer must claim that in such a case, Oscar would in fact not know that he is thinking that water is wet. And as I will show in Section IV, the prospects of this stance are not good.

(III) Davies' Limitation Principle

The third kind of response to the Consequence Problem is that of Martin Davies. Davies argues that while one may know C1 and C2, knowing these claims does not give us warrant to believe the conclusion. This is because Davies endorses the following Limitation Principle on the transfer of epistemic warrant. Now Davies's argument is that cases where warrant is not transmitted from premises of a valid argument to its conclusion will be "the analogue, within the thought of a single subject, of the dialectical phenomenon of begging the question"⁴⁶. The account of begging the question that Davies relies upon is Frank Jackson's. According to Davies, this claims that:

Given argument to given conclusion is such that anyone... who doubted the conclusion would have background beliefs relative to which the evidence for the premises would be no evidence.⁴⁷

⁴⁶ Davies (2000, p. 394).

⁴⁷ *Ibid*, p. 396.

I will not get into the intricacies of how Davies applies this explanation of begging the question to his Limitation Principle, as they are not relevant to the criticisms I will be making. It suffices, for my purposes, merely to say that from his assessment of begging the question, as well as from fine-tuning it to deflect counter examples Davies develops the following two principles:

First Limitation Principle (multi-premise version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, the warrant for that premise counts as a warrant only against the background of certain assumptions and acceptance (i) of those assumptions and (ii) of the warrants for the other premises cannot be rationally combined with doubt about the truth of the conclusion.

And:

Second Limitation Principle (multi-premise version):

Epistemic warrant cannot be transmitted from the premises of a valid argument to its conclusion if, for one of the premises, acceptance (i) of the assumption that there is such a proposition for the knower to think as that premise and (ii) of the warrants for the other premises cannot be rationally combined with doubt about the truth of the conclusion.⁴⁸

If correct, this does solve the Consequence Problem. If we are to accept premise C1, then, *if we were to doubt C3*, we would have to have background beliefs that would suggest that there is no proposition for “water is wet” for me to possibly think. If I were to accept the proposition that “if I am thinking that water is wet, then my environment contains or did contain water” then the only way in which I would be *capable* of doubting the conclusion of the argument, (that I *am* in an environment that contains or did contain water) would be if I doubted that I “water is wet” is a proposition I am capable of thinking. There is no other rational way to accept C1 and not accept C3. So acceptance of the assumption that there is such a proposition for me to think as C2 and acceptance of

⁴⁸ *Ibid*, p. 412.

the warrant for C1 cannot be rationally combined with doubt about C3. Therefore, the Consequence Problem, if used dialectically, would be begging the question, and it is such an argument where warrant for the premises cannot be transmitted to its conclusion.

Bill Brewer has responded to earlier versions of these principles by arguing that they have the consequence of making inferential knowledge of any necessary truths impossible, which is absurd as it

places logic, mathematics, and, on many conceptions, philosophy, in a very poor position epistemically speaking. Something must have gone badly wrong with [Davies' Limitation Principles].⁴⁹

Davies' response was to reformulate his Limitation Principles in such a way as to make them immune from such counter-examples. Whether or not Davies has succeeded in this, I will not get into here. Rather, I will argue that *either* Davies denies that non-perceptual knowledge is closed under known entailment, which would indeed place logic, mathematics and philosophy in a very poor epistemic position, *or* he is committed to something like Brewer's Principle of Acquaintance (A).

Now, in the previous chapter, I remarked that it appears that this Limitation Principle denies that knowledge is closed under known entailment; that if I genuinely know that p , and I genuinely know that if p then q , then I will be in a position to genuinely know that q . Davies, however, denies that this is indeed a consequence. Upon examination it becomes clear that Davies is right about this, however, as I will show, if Davies wants his solution to be consistent with the idea that knowledge is closed under known entailment (henceforth, "closure"), then he is committed to a Brewer-like account of self-knowledge, as discussed in Section II.

Closure in the above argument, and a purely a priori self-knowledge (in the sense that I can know I am thinking that "water is wet" without being in a position to know that there is water in my environment) are incompatible, if Davies is right. Either option, I will argue, is problematic.

⁴⁹ Brewer (2001, pp. 420 – 421).

So how is Davies' position compatible with closure? The idea that anybody who knows that p , and also knows that if p then q , is able to know q , is compatible with Davies' claim that warrant for C1 and C2 is not transmitted to C3, because closure is not committed to the claim that it is the warrant we have for believing that p , and the warrant we have for believing that if p then q , that gives warrant to believe that q . Closure does not require that it be the premises themselves supplying the warrant for q ; it might be that the warrant we have to believe that q is from an entirely different source. That would not violate closure; all that is required for closure is that anybody who has warrant to believe p and if p then q will also have warrant to believe q . Our warrant to believe q need not be generated out of our warrant to believe those two premises. So there is no immediate incompatibility with Davies' solution and closure.

The problem emerges when we consider from what source our warrant to believe a conclusion like C3 might stem. Now the Consequence Problem, recall, is a *reductio ad absurdum* against the compatibility of externalism and (a priori) self-knowledge, because of its absurd conclusion that we might have *a priori* knowledge of the external world. So it seems that if we want to deny this absurd conclusion, we must say that the warrant we have to believe C3 is based on something we know *a posteriori* – and not *a priori*. The warrant for C3 must be grounded in some empirical observation, and not simply a priori theorising.

That much is obvious. But it implies – if Davies wants to maintain closure obtains – that it is *impossible* for us to know both C1 and C2 *without* also having made some empirical observation on the basis of which we can know C3. How might we explain this? Premise C1 seems like it would be knowable, if at all, on possible worlds where there was no such thing as water. All that is required for it is for the sort of armchair reflection that led philosophers like Burge to devise their externalist accounts. Anybody who can reason like that can know C1, and it would be bizarre to suggest that being able to reason like that is only possible on worlds that contain samples of water.

To deny that one can know C1 without some kind of reason to *a posteriori* warrant to believe that there is water in their environment is absurd. Davies then, if he is to accept closure, must maintain that C2 is knowable only to people who are already in a position to know C3. But this, clearly, is precisely the position Brewer argued for,

discussed above. Thus, whatever problems there are with Brewer's account, will affect Davies, if Davies wishes to accept closure. If Davies retains a commitment to closure, then he accepts, like Brewer, that one cannot know that one is thinking that water is wet, unless one can know that there is water in one's environment. As remarked above, this implies that the reality of a slow-switch would undermine self-knowledge. In this next section, I will take issue with this position.

But Davies will only be committed to this if he accepts closure. But what if we deny closure? This also is not something that we can deny unproblematically, given the context of the debate. The challenge the Consequence Problem raises only makes sense in the first place if we accept closure. If we deny this, then the Consequence Problem is all too easily dissolved – for *of course* we cannot know C3 a priori, since knowing C3 does not follow from knowing C1 and C2. Without closure, we would have no special reason to expect people knowing C1 and C2 to be in a position to know C3. To accept that knowledge is not closed is to deny the relevance of Davies' Limitation Principle to the solution of the Consequence Problem. Since Davies understands questions about the transmission of epistemic warrant to be distinct from questions about closure, then it seems that he will need independent grounds for establishing that knowledge is not closed under known entailment, which he has not expressed, or committed himself to in this paper. Nor is it a solution that many protagonists in this debate would be comfortable with. Insofar as people think that the Consequence Problem is worth engaging with at all, it is safe to assume that they also think that accept a closure principle – at least for the *a priori* knowledge presented in each of the argument's premises. If we are to deny closure for the *a priori* knowledge presented in the Consequence Problem's premises, then, clearly, the Consequence Problem must fail. For knowing, if we deny closure here, knowing that I am thinking that water is wet, and knowing what this entails (that I am in an environment that contains or did contain water) does not allow me to legitimately infer that I am indeed in an environment that contains or did contain water. So, any philosopher who takes the Consequence Problem seriously must assume that closure applies in this case. Closure, to such philosophers, must seem plausible. This then, would not help Davies provide a plausible solution to the Consequence Problem.

There are also good independent reasons to consider the denial of closure dubious. Keith DeRose, for instance, calls the claim that I may know that p , know that p entails that q , but yet not know that q “the abominable conjunction”, and describes it as “intuitively bizarre”⁵⁰. Indeed, a prominent proponent of non-closed knowledge, Fred Dretske, admits that the support for the claim that perceptual knowledge is not closed under known entailment derives from the fact that, if it *were* closed, in his view, scepticism would be true, and scepticism must be avoided. Dretske says, when considering why we should accept this position:

One possible reason to abandon K-closure [the thesis that perceptual knowledge is closed under known entailment] is that its denial is not just *a* way to avoid scepticism, but the *only* way. This reason won’t appeal to the skeptic, of course, but, if we could make a case for it, it might carry weight with those who find scepticism as “bizarre” or “abominable” as the rejection of K-closure.⁵¹

This suggests that even Dretske admits that the idea that perceptual knowledge is closed under known entailment is itself compelling; so compelling in fact, that the best reason Dretske finds to abandon it is to show that if we accept it we are forced into denying an equally compelling position (that scepticism is false).

However note that Dretske is committing himself only to denying the claim that *perceptual* knowledge, knowledge gained via perception, is closed under known entailment. This would be sufficient to defeat the scepticism Dretske wants to avoid. In this way, denying closure for non-perceptual knowledge, knowledge gained through *a priori*, armchair reflection, would have all the intuitive bizarreness with none of the anti-sceptical benefit. Given self-knowledge is of this non-perceptual sort, as is the knowledge we can have of the truth of externalism, this suggests making a case for the denial of closure in the premises of the Consequence Problem would be considerably harder than making a case for the denial of closure for perceptual knowledge. This I take to offer

⁵⁰ DeRose (1995, p 28).

⁵¹ Dretske (2003, p. 112).

good reason to avoid, if possible, the claim that closure is violated in the premises of the Consequence Problem.

However, more pressingly, the denial of closure in non-perceptual knowledge is not *only* something to be avoided if possible, but is also a genuinely absurd conclusion. The reason philosophers like DeRose might consider the denial of closure “intuitively bizarre” is not just that it seems compelling. If closure did not hold, then we would effectively have undermined the epistemic support for a large proportion of the things we think we know. A great deal of philosophy operates with that assumption of closure in mind. Philosophical argument proceeds by putting forward premises, and coming to conclusions on the basis of those premises, in the manner that the Consequence Problem does. If we are to deny the link between knowing such premises and knowing the conclusions they generate, then have undermined a great deal of philosophical conclusions.

Indeed, it is not just inferential knowledge in philosophy, or – as Brewer thought – of necessary truths, but *all* inferential knowledge that is undermined if we deny closure. Inferential knowledge proceeds by inferring a conclusion on the basis of a known entailment; and this is impossible if we deny closure of this sort obtains. This is clearly absurd, not only would philosophy, logic and mathematics be in a poor epistemic position as a result, but virtually all forms of human inquiry. Such a conclusion is thus unacceptable; if this is indeed a consequence of Davies’ Limitation Principles, then, like Brewer, we should conclude that something must have gone badly wrong with it.

The less problematic route for Davies to take here, thus, would be to commit himself to a Brewer-like explanation; to claim that closure is not violated, and that rather it must be true that anyone knowing that they think “water is wet” must be in a position to know that water exists in their environment. As I mentioned in the previous section, this position is committed to the claim that the reality of a slow-switch must undermine self-knowledge. The prospects of this avenue will be discussed in this following section.

(IV) Warfield and the Relevant Alternative Approach

In this section, I will examine Warfield's argument for why the Achievement Problem only undermines knowledge for the person being slowly switched – without this threatening the self-knowledge of someone *not* being slowly-switched. This is a position I have argued that Brewer must be committed to, and Davies as well – if Davies does not wish to deny closure⁵².

Warfield's approach is to demonstrate that the slow-switching possibility that the Achievement Problem relies upon can only establish that Oscar, *if* being slowly-switched, cannot know that he is thinking that water is wet. It does not demonstrate, Warfield thinks, that Oscar cannot know he is thinking that water is wet if he is not slowly-switched. He understands the Achievement Problem to be formulated thus:

- P1 To know that P by introspection, S must be able to introspectively discriminate P from all relevant alternatives of P.
- P2 S cannot introspectively discriminate water thoughts from twater thoughts
- P3 If the Switching Case is actual, then twater thoughts are relevant alternatives of water thoughts
- C1 S doesn't know that P by introspection⁵³

To draw that conclusion, he thinks, is a fallacy, as it does not follow from the premises; since the premises are geared around one particular state of affairs, the conclusion that follows from them cannot apply to states of affairs different from the one in question. He says that the conclusion that Oscar, if slowly-switched, would lack self-knowledge is:

relevant *at most* to the following question:

Q: Given externalism, is it *necessary* that the contents of a thinker's thoughts are knowable to the thinker on the basis of introspection?

⁵² For the rest of this section I will operate under the assumption that Davies does not reject closure, and is thus committed to this position.

⁵³ Warfield (1992, p. 218).

[The] argument has not, I maintain, shown *anything* about the compatibility of introspective self-knowledge and externalism. Rather, granting the assumptions of the Switching Case is logically possible, and all three premises are true, it has shown only that the answer to (Q) is “No”.⁵⁴

Warfield goes on to imagine a reply to his argument, claiming that we require that self-knowledge be in some sense necessary. Warfield responds to this by denying that self-knowledge needs such a requirement.

In response to Warfield, first note that, given the difference in formulations, my understanding of the Achievement Problem will escape the fallacy charge. That argument sets up general criteria for self-knowledge, and so its conclusion does in fact follow from the conclusion. Though in return, I take it Warfield will reject the strong condition it places on knowledge as implausible. I will not get into this matter at this stage. My criticisms will be more general than this sort of rejoinder would allow.

In what follows I will argue that any normal, rational agent will be, at any time, capable of forming correct judgements about what he occurrently believes. The claim that any normal person would get such judgements wrong if the environment were to change, is implausible. Any account of self-knowledge that allows that such cases are possible is thus implausible.

This is indeed the sort of account Warfield, Brewer and Davies are committed to, because they admit that the reality of a slow switch would undermine Oscar’s knowledge that he is thinking that water is wet. Now the Achievement Problem, as we have seen, relies upon the possibility of error to undermine knowledge. Thus, to accept that the Achievement Problem undermines knowledge at all (even if it is just in cases where the possibility of error is actual) is to accept that in such a case, Oscar genuinely does commit an error in his judgement about what he is thinking. He may *think* he believes that water is wet, but in fact he does not believe this – rather, he believes that twater is wet.

⁵⁴ *Ibid*, p 219.

I further accept it as uncontroversial that the environmental changes would not affect Oscar's rationality, or any other part of his everyday functioning. For all intents and purposes, Oscar will behave as a normal rational agent, and the idea that this may change because of he is unknowingly shifting between a place that contains H₂O in its environment to a place that contains XYZ in its environment, is absurd. In this way, according to Warfield, Brewer and Davies, in this case we would have a perfectly rational Oscar incorrectly ascribing occurrent beliefs to himself. His second-order judgements about his occurrent beliefs would thus conceivably be false – hence *our* second-order judgements about beliefs *can conceivably* be false, without any change in our functioning as a rational agent.

In what follows, I will argue that any normal, rational agent must always have available a means of making true judgements about what his occurrent beliefs are. On this basis, I will conclude that Warfield, Davies and Brewer are committed to an implausible account of self-knowledge, and thus their position can be dismissed as a viable response to incompatibilism.

Accepting this sort of possibility of error is to accept the possibility of what Sydney Shoemaker calls a “self-blind” person⁵⁵. That is, a person who is capable of blindness regarding his own mental states, in the sense that he is capable of getting judgements about what he is occurrently thinking wrong.

Shoemaker offers the following argument against the possibility of “self-blindness” for rational creatures such as ourselves. If a person is self-blind in this way, then conceivably it is possible for him to make assertions like those commonly brought up in discussions of Moore's Paradox. The kind of assertion in question is of the following sort “it is raining, but I don't believe it is raining”. Despite difficulties in articulating what exactly is wrong with such an assertion⁵⁶, it is evident that there is something logically improper about anyone sincerely asserting it. I cannot, without some logical failure, genuinely believe occurrently that the proposition “water is wet” is true, at the same time as occurrently believing the proposition “I don't believe that water is wet” is true. This much would appear obvious on reflection to any rational agent.

⁵⁵ Shoemaker (1988, p. 34).

⁵⁶ There is a great deal of literature here, but for two attempts to explain Moore's Paradox see Heal (1994) and Shoemaker (1995).

But if a person is self-blind, or at least self-blind about some of her beliefs, then it ought to be possible for her to not notice that she in fact *does* believe that it is raining. That is, a claim of the sort, “I believe it is raining”, should in principle be something that she can be wrong about. She ought to be able to assert “I don’t believe it is raining” when in fact this is something she *does* believe.

But if she can be wrong about this, then she ought to be able to hold something like “it is raining” (from her observations of the environment) while simultaneously holding that she does not believe that it is raining (from her examining the contents of her web of beliefs – whatever that entails – and being mistaken).

Shoemaker considers the possibility of a self-blind person (George) who due to his self-blindness becomes capable of asserting such statements. Assuming that George’s self-blindness does not constitute a failure of rationality, it seems that upon asserting such a proposition, George would immediately notice that there is some logical impropriety in asserting it. As mentioned earlier, so much should be obvious to any rational agent. How would a rational person respond to this impropriety? I take it, as Shoemaker does, that a rational agent would seek to establish consistency between what he takes himself to *believe* and what he takes to be *true*. It is this discrepancy that generates the impropriety, and thus requires elimination. That much, too, I take as obvious on reflection to any rational person. A rational person would also note that the most direct way of doing this would be simply to alter his method of forming beliefs about his beliefs (whatever that is) so that he forms his second-order beliefs in correspondence with how he would respond to claims about the world. That is, George would note that to avoid the logical impropriety, he would need, when considering questions like “do I believe it is raining?” to modify his answer so that its answer is the same as his answer to the question “is it true that it is raining?” If his answer to the second question is “yes”, then George would understand that his answer to the first question must likewise be “yes”, even if he fails to observe that belief within his framework of beliefs, and has independent reasons to think that this is not something he believes.

At this point, however, Shoemaker notes that George is “beginning to look just like a normal person. There is nothing in his behaviour, verbal or otherwise that would

give away the fact that he lacks self-acquaintance”⁵⁷. This, Shoemaker takes as a *reductio* of the position that a rational person can be self-blind; by following the norms of rationality, we can come to a means of correctly judging what our beliefs in fact are. The presence of rationality insulates us from the logical error embodied in assertions like “it is raining, but I don’t believe it is raining”; and in this insulation, it offers us a reliable means of true second-order judgements.

Without a failure of rationality then, it becomes difficult to see how any person could assert propositions like those of Moore’s Paradox. Yet this ought to be conceivable for people capable of making errors in their judgements about their occurrent beliefs. Possessing rationality, for Shoemaker, is thus sufficient for possessing self-knowledge.

Now as this argument suggests, the question of whether I believe that *p* is transparent to the question of whether *p* is true. And if that is the case, then insofar as I am rational, I will always be able to know whether I believe that *p* by considering whether I think there is good reason to think that *p* is true. If this condition of transparency obtains, then – as far as my *occurrent beliefs* are concerned – extreme failures in rationality aside, I have an infallible means of correcting judging what it is I believe. As such, without severe psychological breakdown, I will not be able to be wrong in the way that Warfield, Brewer and Davis are committed to saying that Oscar can be wrong.

There is good reason, beyond the above considerations, to think that transparency obtains. Richard Moran points out:

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.⁵⁸

When questioned about, say, whether or not I believe Thabo Mbeki is a good president, what is it I consider in giving an answer to this question? I do not direct my attention to some inner realm of psychological facts. Rather my attention moves outward; I consider

⁵⁷ *Ibid*, p. 36.

⁵⁸ Moran (2001, p. 62).

facts about Thabo Mbeki, about his policies and whether or not I consider them to be indicative of a good president or not. Moran claims that questions about my belief that *p* are in this way *transparent* to questions about the truth of *p*. Moran argues that beliefs are transparent in this way because enquiries into my own beliefs are undertaken in a *deliberative spirit*; a process of “deciding and declaring myself on the matter” rather than understanding questions about my belief as “a purely psychological one about the beliefs of someone who also happens to be me”⁵⁹.

These considerations support the idea that, from the first-person perspective, questions of whether I believe that *p*, are closely linked in some way to questions of whether *p* is true. Certainly, I cannot accept that certain claims about the world are *in fact* true, and deny that that I believe those claims. This offers powerful support for Moran’s claim that questions about what we occurrently believe are answered by considering how we would answer to the corresponding question about the world – if beliefs are indeed formed this way, this would nicely explain why transparency is the case.

Yet for my purposes, I do not need to argue that beliefs need be formed that way; all that is required for my argument is that by judging what I take to be true of the world, I can correctly arrive at what I occurrently believe. Insofar as this is available, and knowable to me, I will be able to correctly judge what I occurrently believe.

Let me clarify the argument, thus far. It is never possible, without a rational failure, for us to be able to assert propositions like “it is raining, but I don’t believe it is raining”. As far as occurrent beliefs are concerned, the discrepancy between how we take the world to *be*, and how what we take ourselves to *believe* about the world, does not arise. To be the sort of person capable of asserting “it is raining, but I don’t believe it is raining” thus would require a substantial failure in rationality. Yet it appears that Oscar, insofar as he can be mistaken about what he occurrently believes, ought to be capable of asserting such propositions. This suggests that there is something wrong with any account of self-knowledge that allows that we can be mistaken about our occurrent beliefs without rational failure.

A related point is that what we believe seems to be transparent to what we think is true. But this means that insofar as we are rational, and capable of forming judgements

⁵⁹ *Ibid*, p. 63.

about the world, we ought to always have available an incorruptible ability to judge correctly about what our occurrent beliefs are. All it requires is that we recognise ourselves to accept certain things about the world to be true. If we can understand what we *take* to be true of the world, we can understand what it is we occurrently believe. And the slow-switching case ought not to impact on Oscar's ability to understand how he takes the world to be. It would be implausible to suggest that an unnoticed change in Oscar's environment impacts on his ability to accept certain propositions as true, and *know* he accepts them as true. If he can still answer negatively or positively to questions like: "do you believe that water is wet?" then he is capable of taking the world to be a certain way; he is capable of *avowing* certain things about the world. Thus, if he can truthfully answer "yes" to that question, and notice himself doing it, then he can know how he takes the word to be, and thus he can reason his way to a correct judgement about what he occurrently believes.

Thus, if all this is correct, then, as Shoemaker says, rationality must be sufficient for the ability to correctly ascribe second-order beliefs to oneself. Insofar as he is rational, Oscar thus has an infallible means of arriving at true judgements about whether he believes "water is wet". If he thinks the proposition "water is wet" to be true of the world, then he can correctly judge that *right now* he believes that water is wet. Hence the position that Warfield, Brewer and Davies make must be false – since if undergoing a slow switch Oscar *could not* on their view, correctly judge that he is thinking that water is wet.

Burge also supports the above position. Like Shoemaker, he argues against the possibility of Oscar being rational, but incorrect, when he claims "I am thinking that water is wet". Burge argues that there is a necessary connection between judgements about our thoughts, and the thoughts themselves. Like Moran, he says that a rational agent will consider judgements like "I believe that water is wet" and "water is wet" from "the *same* points of view"⁶⁰. This argument, which will be discussed in the following section, claims that second-order beliefs cannot be wrong without a failure of rationality. When I make a second-order judgement about what I believe (say, when I consider whether I believe "water is wet") it is done from the same point of view that I adopt when

⁶⁰ Burge (1996, pp. 257 -258).

I wonder whether the claim “water is wet” is true. In this way, like Moran, Burge thinks that questions of what I believe are closely related to questions about what I think is true.

What this means, however, is that when I say “I am thinking that water is wet” this must be true, unless I violate the norms of rationality. In considering whether or not this statement is true, for Burge, I will follow the same considerations that would justify an answer as to whether or not “water is wet” is true. I can only take the considerations to be sufficiently persuasive for me to accept “I believe water is wet” to be true, if I also take the conditions to be sufficiently persuasive for me to accept “water is wet” to be true. In this way, I cannot hold “water is wet” to be true, but not be able to, on reflection, come to the conclusion that “I believe that water is wet.” And similarly, I cannot claim “I believe that water is wet” without also holding the claim “water is wet” to be true. In this way, Burge argues that claims about what I am occurrently thinking must be *immune from error* – they will always be true, barring a failure of rationality. (In the next section we will see more precisely how this is true even in the case of a slow-switch.).

However, this applies only to occurrent beliefs; it is silent on the issue of whether or not, or how, we are correct in our judgements about what we have believed in the past. Yet that is sufficient for my purposes. Warfield, Brewer and Davies are committed to the claim that we can be mistaken in our judgements about our occurrent beliefs; according to them, Oscar may be wrong when he says that “I believe that water is wet”. Thus the more limited focus on occurrent beliefs will suffice, since it is they are committed to a particular position regarding occurrent beliefs position. If Oscar is rational, then he ought not to be capable of error in the formation of his beliefs about his occurrent beliefs. Warfield, Brewer and Davies are committed to the claim that Oscar can in fact be mistaken in this regard, if his environment is set up in a particular way. However claiming a slow switch undermines Oscar’s rationality is implausible. As such, I take these arguments to constitute good reason to reject this view.

In the next section, I will consider Burge’s defence of compatibilism, the defence I find the most plausible.



(V) Burge, Immunity from Error and Critical Reasoning

The Achievement Problem, recall, as I formulated it, proceeds as follows:

- A1) If I know that I believe that water is wet, then I must know that I do not believe that twater is wet.
- A2) I do not know that I do not believe that twater is wet.
- A3) Therefore, I do not know that I believe that water is wet.

Recall further that in justifying the first premise, I appealed to Stroud's condition for knowledge, a condition that appears to follow the manner in which our everyday ascriptions of knowledge work. The condition was:

if somebody knows that something, p , he must know the falsity of all those things incompatible with his knowing that p (or perhaps all those things he knows to be incompatible with his knowing that p).⁶¹

Understood like this, the role of the slow-switching case was to support the second premise of the above argument – the possibility of a slow-switch raises a possibility of error, it demonstrates a state of affairs that we cannot rule out, that would if actual undermine Oscar's knowledge that he is thinking that water is wet. Burge's response to this problem is to reject the second premise of the above argument altogether – Burge argues that the slow-switching case does *not* actually demonstrate a possible state of affairs wherein Oscar would lack self-knowledge.

Burge's defence of this claim rests in his argument, mentioned briefly in the previous section, for the claim that the point of views of second-order judgements ("I am thinking that water is wet") and first-order thoughts ("water is wet") is in fact the same point of view. Burge responds to the Achievement Problem by pointing out that the second-order thought: "I am thinking that water is wet" is immune from error. That is, whenever you make a claim about what you believe in that sort of way, it is entirely

⁶¹ Stroud (1984A, pp. 29 -30).

impossible for that claim to ever come out as false. In this way, then, Burge concludes that the negative force raised by the possibility of a slow-switch, is dissipated. And thus the Achievement Problem does not generate a genuine threat to self-knowledge.

To provide some more detail, his argument runs as follows: *cogito*-like (judgements about what I am currently thinking) are never wrong, because thinking the second-order thought (“I am thinking that water is wet”) is possible to do only whilst thinking the corresponding first-order thought (“water is wet”) so that the concept (“water”) expressed in the first-order is conceptually redeployed to the second-order. In this way, whatever the term “water” means in the first-order thought, it will mean the same thing in the second-order thought as well. This, for Burge, is because first- and second-order judgements operate from the same point of view.

When we think about what we are thinking, according to Burge, and as demonstrated earlier, Moran, what we do is think about what is true about the world. Thus, if I ask the question: “do I think that water is wet?” my answer will be informed by the same sorts of considerations that I would look to if I was wondering “is it true that water is wet?” If I decide that I do believe that water is wet, that means that I have also decided that it is true that water is wet – and that involves thinking the proposition “water is wet”. In this way, to answer in the affirmative to the second-order question is to think, at that moment, the proposition “water is wet”. In this way, since the proposition “I am thinking that water is wet” is only possible to truly believe if one *simultaneously* thinks “water is wet”, since these judgements happen at the same time, whatever elements that contribute to the content of the one proposition, will contribute likewise to the content of the other proposition. The concept “water” then, must always be the same at both levels. So, for Burge, whilst we might not be able to tell whether we are expressing one particular concept by our term “water” or another distinct concept, we can still be said to know that we are thinking that water is wet. So, the possibility of a switch does not impact on our ability to know what we are thinking. Burge puts this as follows:

Imagine a case of slow switching between actual-home and twin-home situations. In the former situation, the person may think “I am thinking that water is a liquid”. In the latter situation, the person may think “I am thinking that twater is a

liquid”. In both cases the person is right and fully justified as ever. The fact that the person does not know that a switch has occurred is irrelevant to the truth and justified character of these judgements.⁶²

So, on Burge’s account, even though we cannot tell whether we are having water-thoughts or twater-thoughts, we can still be said to know what we are thinking, when we say: “water is wet”.

So we have immunity from error, which thus withstands the Achievement Problem – which, as I explained, relies upon the assumption that alternative possibilities raise a possibility of error. What Burge is arguing, then, is that despite the possibility of a slow switch, Oscar *can* know the falsity of all things incompatible with his knowing that water is wet. According to Stroud’s requirement, the second premise in a standard sceptical argument offers a state of affairs, which we do not know to be false, which would be incompatible with our knowing that p. However, if Burge is right, then the second premise of the Achievement Problem does not present us with such a state of affairs. There is no possible way, save severe mental breakdown, for Oscar’s assertion: “I am thinking that water is wet” to ever come out as false.

However, it is at this point worth noting that questions of a judgement’s *truth*, are distinct from questions of a judgement’s *epistemic status*. To say that a belief is true, even infallibly true, does not alone entail that it must be construed as knowledge. These might be true for non-epistemic reasons – there might be some brute fact of the matter making it true, but yet not something we are justified in believing. Burge, however, acknowledges this point.

On his view, it is also not accidental that these claims are never false. As Burge argues in Burge (1996), if these claims were only accidentally correct, and not correct in a way sufficient for genuine knowledge, we would not be able to function properly as critical reasoners – and of course we do function as critical reasoners. He says:

if reflection were connected to the truth of our *cogito*-like judgements in an accidental or non-knowledge yielding way, the reason-guiding and rational-

⁶² Burge (1988, p. 122).

coherence-making functions of rational review [of our own thoughts] would be broken. Since part of our entitlement to reflective judgements about our judgements about our attitudes derives from their functions in critical reasoning, the entitlement itself would be undermined.⁶³

Critical engagement with our own beliefs and desires makes those beliefs and desires more reasonable. Now this, Burge argues, is possible only because our judgements about our thoughts are not only reliably true, but true in a non-accidental manner. If they were only accidentally true, then following those judgements would not make our beliefs and desires more reasonable – if one forms a belief on the basis of something only accidentally true, the modified belief would not have a better justificatory status than the original belief. However, in critically reflecting on our beliefs, and modifying them on the basis of that reflection, we clearly do improve the justificatory status of our beliefs. To deny this is to deny that critical reasoning has any epistemic value. If we want to hold that critical reasoning has epistemic value, we must concede that we possess self-knowledge. And the fact that critical reasoning has epistemic value, seems undeniable.

In this way, there seems to be no good reason to assert that we would lack knowledge in a slow-switch – our judgements would be true (because they are immune from error), and they would be true non-accidentally (because we are in fact critical reasoners). What more could we want for a belief to count as knowledge? Thus, even Stroud's strict requirement is satisfied – the possibility of a slow switch does not undermine our knowledge of our mental states in the same way that the possibility of the bird outside my window being a goldfinch, undermines my supposed knowledge that it is a canary outside. If Burge is right, then with *cogito*-like judgements, cases like the canary/ goldfinch case do not ever occur. This is Burge's answer to the sceptical challenge the Achievement Problem offers.

Burge has never to my knowledge responded directly to the Consequence Problem. Nonetheless, I think we can draw out his position merely from an examination of his solution to the Achievement Problem. In what follows, I will offer a response to the Consequence Problem that I think Burge has available to him – given his response to the

⁶³ Burge (1996, p. 250).

Achievement Problem. This is not an argument of Burge's, but merely the best argument I think Burge can make, given his solution to the Achievement Problem.

In responding to the Achievement Problem, note that Burge has offered an account of how Oscar can know that he is thinking that water is wet, even though he may not be able to distinguish times when he is thinking that "water is wet" from times when he is thinking that "twater is wet". Oscar will not, on Burge's view *notice* the slow-switch, and will be unaware that the concept he picks out when he utters "water" might be changing. Thus, it seems, Burge agrees with McLaughlin and Tye, that we need not know so much about our concepts to be able to infer anything specific in the environment. We don't need to know whether our concepts are dependent on the environment, or in what way they are, to know what we are thinking, if Burge is right. About this he says:

to self-ascribe thoughts in the way expressed by that-clauses, one has to understand the thoughts one is referring to *well enough to think them*. One need not master anti-individualism, much less have an empirical mastery of the conditions that have established the identity of the thoughts one thinks.⁶⁴

In this way, it seems that Burge agrees with McLaughlin and Tye. Yet Burge need not, and I think would not, commit himself to the two-factor view endorsed by McLaughlin and Tye. There is an alternative explanation available to him. McLaughlin and Tye, recall, thought that the actual singular content of our thoughts was not something we could ever have privileged access to, yet there is another description of that thought under which we could be said to know it in a privileged way. However, this appeared to imply that we *cannot* know what our thoughts are, described in terms of their content, which itself implies that we do not know what the contents of our thoughts are. In explaining what it is about our thoughts that we can be said to know, McLaughlin and Tye appealed to a two-factor theory of thought content or of thought type, in order to maintain privileged access. To provide more detail on this matter, as McLaughlin and Tye understand it; we cannot have privileged access to our thoughts individuated in terms of

⁶⁴ *Ibid*, p. 244. Emphasis mine.

their content *because* we cannot rule out the possibility that they are not empty concepts (like would be the case if we lived on a Dry Earth, where we were collectively hallucinating that there was a clear, non-toxic liquid in the oceans and taps, etc.) This kind of possibility undermines our knowledge of our thoughts, individuated by their singular content. However, explaining it this way, requires them to posit a description of our terms under which they can be understood by us – and it was this aspect that required, problematically, a two-factor theory – McLaughlin and Tye need this to posit something else about our thoughts that we have access to. This was how they explained how it is that we can know what our thoughts are without knowing what exactly they refer to.

Burge, on the other hand, need not explain the situation like this. On what I take Burge's view to be, the possibility of emptiness (that is, the possibility of "water" being an empty concept like "phlogiston") does not undermine Oscar's ability to know that he thinks *water is wet*, when he does, and *twater is wet* when he thinks that. That is, Burge does not concede that I cannot know the singular content of my thought because I cannot rule out the possibility of emptiness. Burge concedes this is a possibility I cannot rule out. However, on my understanding of his view, ruling out this possibility is not *required* for me to know the content of my thought. In this way the difference between the explanations is thus: McLaughlin and Tye retain higher requirements for knowing the contents of our thoughts (the need to rule out the possibility of emptiness) but posit another description of our concepts under which these requirements are met. Thus, they limit what it is that we have access to. Burge, alternatively, need not posit another description of our thoughts described in terms of properties we can have access to, but can deny that we need to rule out the possibility of emptiness in order to properly know the content of our thoughts. Burge can limit the epistemic requirements, without altering what it is we have access to. One is, on his view, not required to rule out the possibility of emptiness, or even know very much *about* the content of our thoughts in order to know what it is. In the next chapter, I will return to these lower epistemic requirements in some detail. I will argue that they are in fact deeply-counter intuitive and thus implausible; knowing so little *about* the content of my thought, according to our own general intuitions about knowledge, is not sufficient for knowledge *of* content. But I will leave this discussion until the next chapter.

How does Burge establish that these lower epistemic requirements apply? Again this is not something Burge has expressly argued for, yet I think he does have an answer available to him. We have seen, in his discussion of the contextually self-verifying nature of second-order thoughts, that there is an immunity from error in *cogito*-like judgements – that is, judgements about what I am thinking. Further, we have also seen, given his discussion of critical reasoning, that these second-order judgements need not only be true, but be true *non-accidentally*. Thus, we have a substantive account of knowledge – we have something that is both always true, and non-accidentally so. Why then, should we have any reason to suppose that what has been described here is anything less than knowledge? What more should we require? It seems, given that these requirements are met, it would be unreasonable, or unrealistic to demand anything further of Burge’s account of self-knowledge. Yet we have also seen that explained in this way, Oscar need not distinguish his water-thoughts from (if he has them) his twater-thoughts, or from his dwater-thoughts (the thoughts he’d be thinking if slowly switched to a Dry Earth, where he hallucinated that there was a liquid in the taps and oceans). This is because Oscar *cannot* distinguish his beliefs in this way, and yet we have been given good reason to consider his beliefs knowledge. Thus the fact that Oscar cannot tell the difference between these thoughts is irrelevant. (Again, I will revisit this argument in more detail in the next Chapter).

Thus, Burge can then explain also why the Consequence Problem is not a genuine problem for self-knowledge. In this way, Burge has an argument that can defeat both incompatibilist challenges. Accepting Burge’s account thus appears the best response currently available to the compatibilist. However, in the next chapter I will argue that vital to this response is a transcendental argument for self-knowledge, and that, in this context, this transcendental argument can offer us no good reason to think that our judgements about our thoughts count as knowledge rather than mere belief *if externalism is true*. As such, despite its apparent success, Burge’s argument against the incompatibilist ultimately fails.

Chapter 3

In the last chapter, I argued that Burge, of the compatibilists discussed, has the best response to the incompatibilist. However, I will argue that Burge's argument does not, in fact, provide us with a good reason to think externalism and self-knowledge are compatible. This, I shall argue, is because Burge resorts to a transcendental argument to establish that we must have self-knowledge, and this does not give us a good reason to consider our judgements about our thoughts as knowledge, rather than mere belief, *if externalism is true*. And, as I said in Chapter 1, to provide an explanation of how externalism and self-knowledge are compatible, Burge would need to give us a reason to think that our judgements about our thoughts count as knowledge, rather than mere belief. As such, Burge does not provide the explanation that is required in order to adequately answer the incompatibilist challenge.

Burge's argument does establish that the access we *do have* to our mental states must be sufficient for knowledge. This question, however, must be understood as distinct from the question of whether or not the sort of access externalism implies we have, is sufficient for knowledge. There may be independent good reason to accept externalism – but at the same time, as I will show, there is also independent good reason to think that the sort of access to the content of our mental states that externalism implies, is insufficient for knowledge. The premises of Burge's transcendental argument for compatibilism could just as easily be used by the incompatibilist to offer a *reductio* of externalism. I will argue that without providing another argument for why we ought to accept his low epistemic requirements for knowledge of content, Burge has not provided a reason to accept that our judgements about our thoughts count as knowledge, rather than mere belief, if externalism is true. To provide a reason for this would entail showing that we have *better* reason to accept externalism than we have to deny his low epistemic requirements for knowledge of content, which is at this point not at all obvious. In this way, the truth of compatibilism cannot be established by Burge's present arguments alone.

In Section I, I revisit the claim I made in Chapter 1 about how a proper response to the incompatibilist entails an explanation of how externalism and self-knowledge are

compatible, which in turn consists in offering us good reason to think that our judgements about our thoughts count as knowledge rather than mere belief, if externalism is true. In light of this I will examine the sort of answer Burge is giving to this question.

In Section II, I will explain how Burge's answer to this question relies upon a transcendental argument, a type of argument that Stroud has influentially argued lacks dialectical force against a sceptic. However, I will also show that given what Burge's transcendental argument attempts to prove, these general criticisms arguably do not apply.

In Section III, I show that even if these general criticisms do not apply to Burge, his argument nonetheless lacks dialectical force against the incompatibilist. As I said above, this argument suffices to show that the access we do have to our mental states is sufficient for knowledge, but without a further argument, this is not an explanation of how *externalism* can be understood to be compatible with self-knowledge.

In Section IV, I start to examine the implications of this conclusion. Here I argue that we cannot treat the conclusion that externalism and self-knowledge are incompatible as merely a paradox. Rather, if we cannot establish compatibilism, then we ought to give up either self-knowledge or externalism.

In Section V, I argue that we ought to give up externalism if we were to be unable to provide an adequate explanation of compatibilism. However, I argue that this has not been established conclusively in this dissertation; we also have to examine in closer detail the prospects of Burge's low epistemic standards. The question that must be answered, then, is whether we ought to accept Burge's low epistemic requirements, or abandon externalism. Which of these is the better alternative, I will leave as an open question. And on that point, I will conclude this dissertation.

(I) Burge and Explanation

As I argued in the first chapter, in raising the Achievement and Consequence Problems, the incompatibilist does not seriously mean to doubt that self-knowledge is possible. What the incompatibilist is asking for, I argued, is an explanation of how self-knowledge can be understood to be compatible with externalism. This sort of explanation, I further

argued, amounts to answering the following question: “why is it that our judgements about our thoughts count as knowledge, rather than mere belief, if externalism is true?”

Thus, to point out that it is the case that we do or must possess self-knowledge, without explaining how it is that we have it *if externalism is true* is not to answer the question. For if we had good reason to believe that externalism was indeed incompatible with self-knowledge, the claim that we *must* possess self-knowledge could be taken as a *reductio* of externalism. That may answer a question about whether we have self-knowledge, but not a question about how it is compatible with externalism. And the two questions must be understood as distinct.

As I mentioned in the last chapter, we must also distinguish questions of truth, from questions of justification, warrant, or knowledge. To say that a belief is *true* is not the same as to say that it is justified, much less than that I know it. One is a claim about truth, while the other is a claim about epistemic status, and these claims must not be conflated. As William P. Alston says:

We can quite properly be interested in what epistemic status [a] belief has, and what gives it that status, without ever having any doubts as to its truth⁶⁵.

To explain how a proposition must be true, is not to explain how it enjoys the epistemic status that it has (whatever that happens to be). Questions of truth and questions of epistemic warrant must be understood to be distinct; to explain how a proposition is *true* is not to explain how we have epistemic warrant to believe it.

Now recall that Burge’s solution to the Achievement Problem was to remove the possibility of error it required. However, by itself, this establishes only that the proposition “I am thinking that water is wet” cannot be false. This by itself, then, is not an explanation of how this can be something we can *know*. To establish that this is knowledge requires another argument. If Burge only established that our *cogito*-like judgements were never false, this would not necessarily explain why we should think that this is in fact knowledge.

⁶⁵ Alston (2005, p. 197).

Burge is aware of the need to say more in this regard. In Burge (1996), he sets out to explain the warrant we have for our beliefs more thoroughly than pointing to their contextual self-verifying nature. He says:

noting that [a *cogito*-like thought] is self-evidently self-verifying ... would not capture fully what is involved in its epistemic status⁶⁶.

In Section III, I will argue that Burge's explanation of how Oscar can *know* he is thinking water is wet is inadequate. Whilst Burge does explain how claims like "I am thinking that water is wet" cannot be false, he does not give us good reason to take it to be *knowledge*, if externalism is true.

Burge, recall, goes on to ground the entitlement we have to our own mental states in our capacity for critical reasoning. Burge argues that being rational agents, capable of rationally and critically engaging with our own thoughts, requires that we have an entitlement to make judgements about our mental states, and that these judgements normally be correct in a non-accidental way. On the basis of such considerations, Burge concludes, regarding the entitlement we have to our mental states:

The entitlement remains constant under possible unnoticed variations in environmental circumstances or cognitive content. For it does not depend on checking whether or judgements meet certain conditions. It depends on the judgements' being instances of a kind essential to critical reasoning. Critical reasoning presupposes that people are entitled to such judgements. Since we are critical reasoners we are so entitled⁶⁷.

Thus Burge offers a more substantive account of the epistemic entitlement or justification basis of self-knowledge by grounding self-knowledge in critical reasoning. Burge can show that critical reasoning requires that our judgements about our occurrent thoughts be

⁶⁶ Burge (1996, p. 241).

⁶⁷ *Ibid*, pp. 262 – 263.

true, and be true in a non-accidental manner. What more should we require in an account of self-knowledge?

At this point, one may be tempted to understand such a story as a good reason to think of our judgements about our thoughts as knowledge rather than belief – and as such, providing the required explanation. Burge’s explanation runs as follows: any critical reasoner must possess self-knowledge. Thus, we have good reason to think that critical reasoners necessarily possess some reliable belief forming mechanism, so that our beliefs about our mental states is a reliable guide to what our mental states really are like.

However, in the next two sections, I will argue that this in fact *does not* provide us with good reason to consider our judgements knowledge rather than belief – because the argument Burge uses is transcendental, which, I hope to show, cannot be employed to successfully in this context. However, as I will argue, this argument is not based on Stroud’s argument that transcendental arguments require a prior means of bridging the gap between mind and world.

(II) Transcendental Arguments: General Problems

What, in the first place, is a transcendental argument? In his introduction to the collection *Transcendental Arguments: Problems and Prospects* Robert Stern says:

The first, and perhaps most definitive feature, is that these arguments involve a claim of a distinctive form: namely, that one thing (X) is a necessary condition for the possibility of something else (Y), so that (it is said) the latter cannot obtain without the former.⁶⁸

Understood like this, then, Transcendental Arguments might proceed something like this:

- T1) X is a necessary condition for Y
- T2) Y obtains,
- T3) Therefore, X obtains

⁶⁸ Stern (1999, p. 3).

If this is right, then it looks like Burge's argument is indeed transcendental; it proceeds, as the above quoted passage from Burge demonstrates, roughly as follows:

- BT1) Self-knowledge is a necessary condition for critical reasoning
- BT2) We are critical reasoners
- BT3) Therefore, we must have self-knowledge

That Burge's argument here is transcendental, is in fact noted by Stern in his introduction – Stern explicitly cites the above argument as a “less well-known” transcendental argument, which has been debated less than the other influential recent uses of these kinds of arguments – namely the transcendental arguments of Davidson, Putnam and Searle⁶⁹.

Burge, we have seen, using this transcendental argument to give us good reason to think of our judgements about our mental states as knowledge, and not mere belief, as the sceptical argument the Achievement Problem uses suggests. However, objections raised by Stroud against the use of transcendental arguments in general, suggest that transcendental arguments lack dialectic force against the sceptic, as they presuppose a prior, and controversial, means of bridging the gap between mind and world that would, if true, offer an independent refutation of scepticism. If these objections apply to Burge, then his argument, as a response to the Achievement Problem, would not have succeeded in giving us reasons to think of our judgements about our thoughts as knowledge, rather than mere belief. And therefore, would not have succeeded in providing the required explanation. I will go into these objections in more detail momentarily, but first I will make some cursory points about Burge's employment of a transcendental argument in the debate regarding self-knowledge and externalism.

It is interesting that Burge uses a transcendental argument here – especially given the prominence in the debate of the Consequence Problem. Like Burge's argument, the Consequence Problem follows the form Stern associates with transcendental arguments. Observe:

⁶⁹ *Ibid*, p. 1.

- C1) If I am thinking that water is wet, then I must be in an environment that contains or did contain water.
- C2) I am thinking that water is wet.
- C3) Therefore, I am in an environment that contains or did contain water.

Clearly, then, the above argument form is transcendental; it begins by considering that some state of affairs (my thinking that water is wet) is only possible if certain conditions are met (I am in or have been in a watery environment) and concludes, given that such a state of affairs is real (I am indeed thinking that water is wet) that its conditions thus must be actualised (So I must be, or have been, in a watery environment). However, this conclusion is understood in this argument to be *absurd*; one cannot, the incompatibilist thinks, establish that water exists in my environment by means of such an argument. The fact that transcendental arguments are prominently used in this context as a *reductio* ought perhaps to immediately raise some suspicions about their legitimacy. I will return to the links between transcendental arguments and the *reductio* arguments later on.

For the moment, however, I will articulate Stroud's influential criticisms of transcendental arguments, in order to see whether they would apply to Burge's argument. The move of a transcendental argument that, since Stroud, is considered problematic, is the move from a psychological fact, which is essentially what critics understand premise T1 to be, to a non-psychological fact – T3.

The problem is that the sceptic may respond to T1 by claiming that it only *seems* that X is a necessary condition for Y – that this is in fact merely a fact about how we see the world, rather than a fact about how the world is. Or, if we define Y in such a way that X is required for it, then the sceptic may deny that Y, defined in that way, genuinely obtains. The problem is that it seems that saying that X is necessary for Y is a conceptual claim – but, in general, one of the things the sceptic doubts is whether or not our concepts or thoughts about the world genuinely correspond to what is in fact in the world. In order to establish that it follows from our *thinking* that X is necessary for Y, to the fact that X is necessary for Y, Stroud thinks that some version of the verificationist principle is required – some principle that confines questions of what is meaningful to questions of

what is empirically knowable. On this view, statements about unknowable things are meaningless. To establish that any claim X really is necessary for Y (call this claim 'S') what is required is to show that conceptual link between X and Y must apply to the real world – that given what X and Y *mean* it must be true that X and Y are genuinely, in the real world, connected the way we think they are. But this implies a connection between something being meaningful, and something being true. But to say that is to endorse a form of verificationism. As Stroud says:

Any opposition to scepticism on this point [denying that S really is true, rather than merely making sense to us without being true] would have to rely on the principle that it is not possible for anything to make sense unless it is possible for us to establish whether S is true, or alternatively, that it isn't possible for us to understand anything at all if we know only what conditions make it look for all the world as if S is true, but which are still compatible with S's falsity. The conditions for anything's making sense would have to be strong enough to include not only our beliefs about what is the case, but also the possibility of our knowing whether those beliefs are true; hence the meaning of that statement would have to be determined by what we can *know*. But to prove this would be to prove some version of the verification principle, and then the sceptic will have been directly and conclusively refuted.⁷⁰

If we dispute the sceptic's claim that T1 is merely a psychological claim on any other ground other than endorsing verificationism, we will run into similar problems. This can be understood when we consider what must be done to establish the first premise as a non-psychological fact. To establish the first premise as a non-psychological fact is to establish it as a very strong modal claim – about how things must necessarily be in the world order for Y to obtain. The issue is that it seems the first premise is established by means of an *a priori* analysis of our own thoughts and concepts – our own way of looking at the world. The question, in essence, thus becomes: how can we make a strong, modal claim about the world, if the only evidence we use for this claim is based on

⁷⁰ Stroud (1968, p. 24).

investigation of our thoughts and concepts of the world? A sceptic, as Stroud points out, can very plausibly maintain that despite the strong conceptual support for the first premise, this is nonetheless simply based on our – perhaps unavoidable – particular way of looking at the world, rather than on the way the world really is. In this case, then it seems we have a gap between how we understand the world to be, or our concepts, and how the world really is. Some independent means of bridging this gap is required for the transcendental argument to work. As Robert Stern says, in this way, transcendental arguments lack *dialectical* force against the sceptic – that is, they presuppose either verificationism, which by itself would refute scepticism, or some other means of closing the gap between concepts and world that would in itself offer an independent refutation of most forms of scepticism⁷¹. And if that is the case, then the transcendental argument offers no new solution to scepticism.

So the reason that the above argument lacks dialectical force is because we can assume the sceptic to oppose the move from psychological fact (T1) to non-psychological fact (T3). Yet it is not clear that the incompatibilist would resist the move from how we think about the world, to how the world really is. Why would the incompatibilist oppose this? The incompatibilist is not a sceptic; she does not rely upon this sort of general gap between mind and world in making her argument. Indeed, if the incompatibilist is indeed committed to scepticism, that may be construed as hurting her position, as I will discuss in Section IV. And insofar as the incompatibilist can plausibly be understood to accept that what we *think* is required for self-knowledge is what is in fact required, then, in this context, Burge can legitimately assume that there is no general gap here. Given the differences in what the sceptic accepts from what the incompatibilist accepts, these general criticisms thus cannot apply to Burge's argument.

In this way, Burge can legitimately use a transcendental argument here; we have thus far no reason to suppose that Burge's argument indeed lacks *dialectic force* against the incompatibilist, and thus can then be seen as providing a reason to think of our judgements about our thoughts as knowledge.

⁷¹ Stern, (1999, pp. 6 – 7).

However, there is another reason to dismiss it as an adequate explanation of why our judgements about our thoughts count as knowledge rather than belief, if externalism is true. In the next section, I will put forward my argument for this claim.

(III) Burge's Transcendental Argument: Specific Problems

To clarify my claim, let me return to what was missing from Burge's contextually self-verifying argument against the Achievement Problem.

I noted above that pointing out that the claim "I am thinking that water is wet" is immune from error does not, by itself, give us a reason to think we *know* what we are thinking. This is not an explanation of the epistemic status of that belief, or how it can be considered knowledge.

Why might we consider the claim "I am thinking that water is wet" *not* to count knowledge, if externalism is true? Or, more specifically, why might we consider the claim "I am thinking water is wet" not to count as knowledge of *content* if externalism is true? As I made clear in Chapters 1 and 2, what we require in an account of self-knowledge is that it delivers knowledge of the content of our thought – if it offers only knowledge of something other than that thought's content, then this is an inadequate account of self-knowledge. In this way, what is required here is knowledge of a particular feature of that thought – I not only need to know *something* about this thought, I need to know its content. In this way, there is in fact a particular feature of my thought that needs to be knowledgeably available to me. Now, as I will show, it seems far more plausible to claim that I do *not* know the content of my thoughts, if externalism is true, given our common and strongly held intuitions about the requirements for knowledge. Rather, it seems, if externalism is true then I can only know some other feature of it.

As I noted in Chapter 2, this claim is supported by McLaughlin and Tye. They point out that we can only know our thought when we type them in terms of certain properties they have – not all properties. This implies we cannot know all properties of our thoughts. We cannot for example, know (at least not in the manner required for self-knowledge) our thoughts when typed in terms of the brain states they correspond with.

This seems quite clear – imagine my thinking “water is wet” happens to be regularly correlated with C-fibres firing in my brain. This relation is thus a property my thought does have. But if I am asked whether I am thinking a thought that regularly occurs at the same time as C-fibres firing in my brain, I might in fact reply I am not (assuming I have some other, mistaken opinion). This clearly indicates that I do not know my thoughts, when typed in terms of the relations they have to my physical brain states. And from this, it seems obvious that we can conclude that I do not know this particular property of my thoughts. To suggest I *know* my thought “water is wet” regularly corresponds with C-fibres firing in my brain in this case would seem ludicrous. I do not have privileged access to this feature of my thoughts – I have privileged access only to another feature of my thoughts.

Now, in a similar way, Oscar might not know that when he thinks about “water” he is in fact thinking about a substance with a particular chemical composition. If we ask Oscar whether he is thinking about a stuff made up of H₂O, he might reply that he is not. Yet according to externalism, “H₂O” is still part of the propositional content of Oscar’s thought, when Oscar thinks about “water”. In this way, it looks like Oscar does not know his thought, when typed in terms of its *singular* content. This is precisely what McLaughlin and Tye concluded –and they further thought that any claim that this is something we could know, was “absurd”. Now, as we noted in the above example, if we cannot know a thought when it is typed in terms of certain properties, this implies we cannot know that it has these properties. This, then, implies we cannot know the singular contents of our thoughts.

Yet, as I argued in Chapter 2, externalism implies that the content of certain thoughts is singular. From all this then, it seems that we cannot know the content of our thoughts if indeed externalism is true.

Now Burge, I suggested, would claim that we can indeed know the singular content of our thoughts. This is because Burge endorses lower epistemic standards for knowledge of content than do McLaughlin and Tye. To reiterate, McLaughlin and Tye think I do not know the singular content of my thought because its singular content may be different from how I take it to be (in the sense that I might not take the content of “water” to be “H₂O”, in the manner described above). As a result of these sorts of

possibilities of error, I do not, on McLaughlin and Tye's view, genuinely know what singular content it has.

I also argued that Burge would deny this. He does not require that we be able to rule out such possibilities in order to know the content of our thoughts. Hence his epistemic requirements for knowledge of content are lower than McLaughlin and Tye's. On Burge's view, Oscar can know the *content* of his thought, even when H₂O is in fact part of this content. In support of this, to repeat, he says:

to self-ascribe thoughts in the way expressed by that-clauses, one has to understand the thoughts one is referring to *well enough to think them*. One need not master anti-individualism, much less have an empirical mastery of the conditions that have established the identity of the thoughts one thinks.⁷²

Burge's explanation of self-knowledge thus is committed to the following position: I do not need to have discovered the chemical composition of water in order to know what I am thinking about when I'm thinking about "water". Yet H₂O is the singular content of my water-thoughts, and this singular content, on Burge's view, must be knowable. As opposed to McLaughlin and Tye, who think that the fact that I may have mistaken beliefs about what the singular content of our thoughts means we cannot know this content, Burge's account entails that we can know the singular contents of our thoughts ("water") when I might in fact have many mistaken beliefs about what the term "water" is supposed to refer to. If someone asks Oscar whether his thought has singular content, he might reply that it has none; even so, Oscar, on Burge's account, *knows* the singular content of his thought.

But this seems intuitively bizarre. We do not – and should not – consider Oscar to be able to know that his thought corresponds with C-fibres firing in his head. This is because if we described the thought in that manner to Oscar, he might well honestly deny that it is in fact the thought he is thinking. Given his honest denial that he is thinking a thought corresponding with C-fibres firing in his brain, it seems the property of being

⁷² Burge (1996, p. 244). Emphasis mine.

regularly correlated with C-fibres in his brain, is not a property of his thought that Oscar knows.

Similarly, since Oscar might honestly deny that he is thinking about H₂O, and since “H₂O” forms part the content of that thought, it makes sense to deny that Oscar can know his thought, when typed in terms of its singular content. Which implies, as we have seen, that the actual content of his thought, is not a property of his thought that Oscar can know either.

This intuition that *these* sorts of possibilities of error need to be ruled out in order for something to count as knowledge, does not stem from an intuition like the one underpinning the sceptical challenge. *Alternative* possibilities are not at issue here. All this intuition stems from, is the intuition that if Oscar does not seem to *believe* that “H₂O” forms part of the content of any proposition he thinks, then this cannot be something he *knows*. Yet, this is something he is required to know, if both self-knowledge is true, and externalism is true. “H₂O” is vital component of the content of Oscar’s water-thoughts, and Oscar, for all intents and purposes gives us no reason to suppose he believes this. As such, it does seem strongly counter-intuitive to suggest that Oscar can know what the actual content of his water-thought is. If Oscar seems to know very little *about* the singular content of his thoughts, in the sense that he might completely misdescribe it, then it seems bizarre to suggest that he knows *what* the singular content of his thought is. It would be far more reasonable, in such a case, to conclude that Oscar only has an epistemically privileged access to some other feature of such thoughts – not their singular content. But from this it appears to follow that Oscar does not know the *singular content* of his thoughts – in the same way that it seems clear that Oscar does not his thought corresponds with C-fibres firing in his head. If our knowledge of content is answerable to the same requirements as our knowledge of the relational properties of our thoughts, then it would seem undeniable, in this case, that Oscar does not the content of this thought. Neither singular content, nor the relation it bears to his brain seems to be known to Oscar, according to these requirements.

Knowledge of content, thus, must be seen, on Burge’s view, to have much lower, or at least very different requirements than our requirements for other forms of

knowledge⁷³. How can we accept such low requirements for knowledge of content? What reason does Burge give for why we should accept these standards as sufficient for knowledge?

The best reason for this will emerge from Burge's transcendental argument. Burge has argued that without rational failure we must possess self-knowledge. Thus, since there are in fact cases where we cannot achieve requirements higher than the standards discussed above, we might conclude that these lower epistemic requirements must be sufficient for knowledge of content. Since we must possess self-knowledge, and because these cases of error sometimes occur, it follows from those considerations that to know I am thinking "water" I need not know much *about* its content, in order to be said to know what content it has.

Burge, then proceeds in this way; in arguing for these standards, he establishes that we do and must possess self-knowledge. He then argues that the Achievement and Consequence Problems establish that we cannot achieve higher standards than these, if externalism is true. He then concludes that the lower standards, the ones we do achieve, must be sufficient for knowledge. This then, would be Burge's argument for the lower requirements:

- A) We must have self-knowledge because we are critical reasoners.
- B) We can only achieve lower epistemic requirements, according to externalism.
- C) Therefore, lower epistemic requirements must be sufficient for knowledge.

Note, at this point, that the conclusion does not follow from the argument's premises, since there is nothing in the above argument to say that we do only achieve the lower epistemic requirements. This can be modified as such:

- D) We must have self-knowledge because we are critical reasoners.
- E) We can only achieve lower epistemic requirements, according to externalism.
- F) Externalism is true.

⁷³ Burge could supposedly, deny this and claim that all forms of knowledge need only meet these requirements. But if this is the case, then we have no good reason to deny that Oscar knows that his thought correlates with C-fibres firing in his brain. But to say that Oscar might know this, clearly, would be absurd.

G) Therefore, lower epistemic requirements must be sufficient for knowledge.

However, even with this modification, there is no reason to suppose that the incompatibilist need accept, or would want to accept, this argument. The suggestion that Oscar might know what the singular content of his thoughts are in instances where he may be mistaken in an assessment of what the content of his thought is, might be taken to be intuitively bizarre, to the same degree that philosophers like DeRose find the claim that knowledge is not closed under known entailment intuitively bizarre. It certainly would be utterly bizarre to suppose that Oscar genuinely knows that his thought has the property of being correlated with C-Fibres firing in his brain, after all. This would be at least as bizarre as the denial of closure. Yet, it seems, Oscar may in some cases have as much access to his thoughts, typed in terms of their singular content, as he does to his thoughts typed in terms of the relations they hold to his brain. Burge's lower epistemic requirements clearly cannot plausibly apply to knowledge of the relations our thoughts have to our brains. To claim this would be intuitively bizarre. So why should we think they are less bizarre when applied to knowledge of content? If I believe that I am thinking about any clear, tasteless liquid when I think about "water", then it seems I do not know that I am in fact thinking about H₂O. Yet H₂O is the singular content of my water-thoughts, and this, on Burge's account, is knowable without any empirical research.

And as we saw in the last chapter, the supposed "bizarreness" of the denial of closure, as Dretske acknowledges, has led some philosophers to treat any argument in favour of that conclusion to be a *reductio* of the position endorsed; it is too counter-intuitive, they think, to take seriously.

Thus a similar claim might be made about the lower epistemic standards endorsed by Burge. The claim that I know the singular content of my thought in cases where I believe that this thought has different content, seems as counter-intuitive as the claim that perceptual knowledge is not closed under known entailment. The position Burge maintains is in this way intuitively bizarre, and difficult to take seriously.

Burge's support for these lower epistemic requirements, we noted, stems from his claim that if they are insufficient for knowledge, *then we lack self-knowledge*. This then might be considered analogous to Dretske's argument, quoted earlier:

One possible reason to abandon K-closure [the thesis that perceptual knowledge is closed under known entailment] is that its denial is not just *a* way to avoid skepticism, but the *only* way. This reason won't appeal to the skeptic, of course, but, if we could make a case for it, it might carry weight with those who find skepticism as "bizarre" or "abominable" as the rejection of K-closure.⁷⁴

Like Dretske, Burge endorses a counter-intuitive position on the basis that it seems that only in endorsing this position can self-knowledge can be had – assuming the truth of externalism.

Yet this gives us some reason to think that externalism is false. The situation is as follows; if we accept externalism, it seems we must accept one of two counter-intuitive positions; *either* we lack self-knowledge, or we must accept Burge's lower epistemic standards as sufficient for knowledge. Both seem highly counter-intuitive. In Chapter 1, I said that if it could be shown that accepting externalism leads to intuitively bizarre consequences, this would diminish the support we have for externalism. As Boghossian points out, the support for externalism derives largely from the intuitions brought out by the Twin Earth thought experiments. The support of externalism is that it appears the only way we can explain the fact that Oscar and Toscar will be having different thoughts; the conclusion that they *are* having separate thoughts, is supported by our intuitive responses to the thought experiment. As Boghossian says: philosophers accept externalism "because their intuitive responses to a certain kind of thought experiment... appear to leave them little choice"⁷⁵. If externalism could be shown to have intuitively bizarre consequences, this would substantially diminish the primary support for the position.

That we must accept externalism, in this situation, is no longer so obvious. We could consider the "absurdity" of the lower epistemic requirements sufficient to show that externalism *must be* false, or we could consider the support for externalism sufficient to establish that the lower requirements *must be* true. The argument that the compatibilist

⁷⁴ Dretske (2003, p. 112).

⁷⁵ Boghossian (1998, p. 273).

treats as a transcendental argument for the lower requirements, the incompatibilist treats as a *reductio* against externalism. It is by no means clear which would be the better way to proceed.

However, if this is the position Burge is in fact in, then he has not yet provided us with a reason to think that, if externalism is true, our judgements about our thoughts will count as knowledge as opposed to mere belief. How we respond to the considerations Burge raises seems largely a matter of intuition. Why should we accept that his lower epistemic requirements are sufficient for self-knowledge rather than accept that externalism is in fact false? At present, there seems no means of deciding this question. As such, we have *not* been given a good reason to think of our judgements about our thoughts as knowledge rather than belief, if externalism is true.

To provide such a reason, what must be established is that it is more reasonable to endorse the lower epistemic requirements than to abandon externalism. To do this he must establish that these consequences are in fact not intuitively bizarre – or not as bizarre as the denial of externalism. To establish that, however, would require not only an argument for externalism, but also a thorough examination of our intuitions regarding requirements for knowledge, and showing that our common understanding of knowledge does not commit us to anything stronger than these lower requirements in this context. Yet, to do that, I take it, would require a separate argument.

In this way, it becomes clearer why Burge's transcendental argument for self-knowledge cannot adequately answer supply the reason we are looking for. Transcendental arguments proceed by offering premises that we must accept, extracting the necessary conditions for the truth of these premises, and then concluding that these necessary conditions must be true. In this particular case, however, it isn't clear whether it would be more reasonable to deny one of the premises to avoid the conclusion, rather than accept the premises and be committed to the conclusion. An alternative manner of responding to Burge's argument for the lower epistemic requirements would be to combine Burge's considerations with the intuition that his conclusion must be false, to argue that one of his premises (externalism is true) must be false. Whether we should treat his premises as a transcendental argument for the lower epistemic requirements, or as a *reductio ad absurdum* of externalism is not yet established. As such, we have not

been given a good reason to think that our judgements about our thoughts count as knowledge, rather than mere belief, if externalism is true. Burge's transcendental argument thus offers no explanation of how self-knowledge is possible if externalism is true.

(IV) Implications: Incompatibilism as Paradox

I have argued that the incompatibilist has been asking for is an explanation of how self-knowledge is compatible with externalism – and I argued that this requires us to supply a reason to suppose our judgements about our thoughts count as knowledge, rather than mere belief, if externalism is true. I have also argued that thus far no such reason has been given.

From what I said above, a reason could be given if it could be established that Burge's low epistemic requirements are in fact not as intuitively bizarre as they first appear. I do not intend to make this argument here – as it stands, I have no idea what a plausible argument for this would look like.

Rather than offering this answer, in this section I will consider the implications of what I have established. Does a lack of this kind of explanation entail that we give up externalism or self-knowledge? If so, which? Or do we treat this situation rather like a paradox – puzzling, but not to be taken seriously?

If a case could be made for either of the following claims, then I think we would have reason to treat this conclusion as a paradox. The first is that the considerations in favour of incompatibilism are the same considerations that favour Cartesian scepticism, so that we cannot consistently take incompatibilism seriously without taking scepticism seriously. The second is that these same problems will prevent someone who is not an externalist being able to explain how self-knowledge is possible. In what follows, I will examine and dismiss the case for each of these claims. In the next section, I will examine whether we ought to give up externalism or self-knowledge.

I will begin by showing why these claims, if correct, should lead us to consider incompatibilism as simply a paradox. If we could show that the incompatibilist is simply

bringing the same considerations that underwrite scepticism to bear on self-knowledge, then this might give us a reason not to take the outcome seriously. Stroud says this:

I would grant – indeed insist – that philosophical scepticism is not something we should seriously consider adopting or accepting.⁷⁶

If, like Stroud, we feel that scepticism about the external world ought not to be adopted, or accepted, and if the reasons that underwrite scepticism underwrite incompatibilism, then we would have some reason not to take incompatibilism seriously. If these considerations are indeed analogous, then consistency requires we treat one the same as the other.

Alternatively, if it could be shown that the problems raised by the Achievement and Consequence Problems would apply also to philosophers who deny externalism, then these challenges would not have established that externalism and self-knowledge are indeed incompatible; merely, rather, that self-knowledge is hard to explain. If it is not only the externalist who has trouble responding to these problems, then we have no reason to think that it is in virtue of his externalism that he faces difficulties in explaining self-knowledge. As such, the lack of a robust compatibilism here need not suggest that there is a problem with holding externalism to be true for a philosopher committed to self-knowledge. Rather, there may simply be a problem with accepting self-knowledge.

There are certainly similarities between the standard sceptical challenge and the Achievement Problem – as I remarked in Chapter 1. Both arguments rely upon the possibility of error that we cannot rule out, to undermine knowledge of the sort in question. The sceptical challenge argues that because we cannot rule out the possibility that our perceptions are in some way deceived, nothing we accept on the basis of our perceptions can count as knowledge. The Achievement Problem argues that because we cannot rule out the possibility of a slow switch, we cannot accept that any belief we hold about an occurrent belief or desire where a kind-dependent term like “water” is part of its propositional content, can count as knowledge. Both arguments, in this way, commit themselves to the same strict epistemic requirements.

⁷⁶ Stroud (1984B, p.1).

Yet it is certainly unclear that anyone convinced by the Achievement Problem need be committed to scepticism, or *vice versa*. There are options available in response to the sceptic, that are not available in response to the Achievement Problem. Also, there are means of arguing against the Achievement Problem, which cannot be used against scepticism.

In Chapter 2 I argued that one cannot refute the Achievement Problem on the grounds that the slow-switching case is not a *relevant possibility*. I argued that this position is committed to an implausible account of self-knowledge; namely, that it is possible for a rational agent to be mistaken regarding his occurrent beliefs. This is implausible, because from the first-person perspective, questions of our occurrent beliefs are transparent to questions of truth – so, if I can assess what it is I think is true of the world, then I can know what I occurrently believe. Hence, if I am questioned about whether I believe that *p*, I can answer correctly, by considering whether or not I think *p* is true. In this way, any person capable of assessing what he thinks is true of the world has available a means of correctly judging what he occurrently believes. A change in his surrounding environment cannot be understood to detract from this.

There is no such limitation in responding to the sceptic – it is *not* implausible to maintain that my beliefs about the external world can be right or wrong, depending upon the environment around me. Indeed, this seems quite likely; in a normal world, when I form beliefs based on sense experience, they will for the most part be correct. If the world around me is as I take it to be, then my belief that the sun is shining will be *true*. If the world around me is in fact an illusion created by an evil demon or a mad scientist, then my belief that the sun is shining will be *false*.

In this way then, there are prospects available for refuting scepticism that are not available for refuting the Achievement Problem; one can run an argument claiming that I need only rule out *relevant alternatives* to my knowing that the sun is shining, in order to know that the sun is shining, or any one of the various contextualist solutions available⁷⁷. Yet, because these options all admit of a context in which my apparently well-justified belief that the sun is shining is false, without severe psychological breakdown, these

⁷⁷ For some examples, see DeRose (1995) and Dretske (2003).

solutions can offer no analogous response to the Achievement Problem. One need not be committed to scepticism if one accepts the Achievement Problem.

This would also work the other way. Earlier in this Chapter, I argued that the general concerns regarding transcendental arguments as responses to scepticism, need not apply to a transcendental argument for self-knowledge. The problem with Burge's argument was that it was committed to an implausible understanding of the epistemic requirements for knowledge of content, so that it was not clear whether it would be more reasonable to treat the argument as a *reductio* of one of its premises (externalism is true) rather than accept it as a transcendental argument for its conclusion. However, a transcendental argument using premises that are more clearly undeniable could in principle provide an acceptable solution to the Achievement Problem.

Finally, *even if* anyone endorsing the Achievement Problem was committed to scepticism, the incompatibilist could always rely upon the Consequence Problem to generate worries. As I argued in Chapter 2, Burge's account seems the best available response to the Consequence Problem. This response claims that I do not need to know much about the content of my belief, in order to know what its content is. So it does not follow from my knowing that I am thinking that "water is wet" to my being able to conclude that there is water in my environment. As I argued in this Chapter, the reason Burge can offer us for why we ought to consider this lower epistemic requirement knowledge, rather than belief, is based on his transcendental argument, and this argument does not provide us with a good reason to accept these low requirements.

As such, *if* the Achievement Problem were indeed too similar to scepticism to be taken seriously, the incompatibilist nonetheless has the Consequence Problem available. There is no obvious link between the Consequence Problem and external world scepticism at all, so there is no good case to be made for why one cannot accept the Consequence Problem without accepting scepticism. It is thus no great loss even if the incompatibilist were to be forced to abandon the Achievement Problem.

It is also not clear how either the Achievement or Consequence Problems could apply to a non-externalist. As I remarked in Chapter 2, it is only upon the assumption the contents of our thoughts can be *wide* that either the Achievement or Consequence Problems threaten self-knowledge. The Achievement Problem argues that possible

unnoticed environmental changes threaten our knowledge of the content of our thoughts. But if our thought content is *not* taken to depend upon the environment in the way externalism maintains, then possible unnoticed changes in the environment cannot threaten self-knowledge. If thought content is taken to be narrow, however, then it would not change if the environment were to change. Hence the sort of possibility of error raised by the Achievement Problem cannot apply to anyone who thinks that thought content is not dependent upon the environment.

The Consequence Problem serves as a *reductio* of self-knowledge only if we assume that having certain thoughts entails that certain environmental conditions obtain. If we deny that thoughts are dependent upon the environment in the way externalism maintains, then it would not make sense to suggest that if I am thinking that “water is wet”, then I must be in an environment that contains or did contain water. This only makes sense if we maintain that my thought content depends upon my physical environment. If we deny this, the Consequence Problem, too, will not pose a problem for self-knowledge.

There may be other problems that threaten our ability to know our thoughts, regardless of whether or not externalism is true. However these problems, whatever they are, must be seen as different from the Achievement and Consequence Problems. If one is not an externalist, one can plausibly maintain *both* that we possess self-knowledge, and that the arguments offered by the Achievement and Consequence Problems are correct.

In this way then, the two claims I considered which could conceivably support our treating incompatibilism as merely a paradox, are incorrect. In the absence of any other plausible reasons, I conclude that we have no reason to think the correct response to incompatibilism is to treat it like a paradox. In the next section, I will consider whether we ought to give up self-knowledge or externalism as a result of an unanswered incompatibilism. In this section I will also conclude this dissertation.

(V) Implications: Self-Knowledge and Externalism Revisited

In this section, I argue that of the two, externalism ought to be given up rather than self-knowledge, on the grounds that scepticism about self-knowledge seems at least as

difficult to take seriously as, if not more so than, scepticism about the external world. Externalism, while plausible, is not as undeniable as self-knowledge.

However, I do not recommend that we accept the falsity of externalism just yet. I also argue that this is only the position we should take *given the unavailability* of a response to incompatibilism. It is certainly not the case that no reason can be given for why we should accept our judgements about our thoughts to count as knowledge. This can be done; what must be established is that Burge's lower epistemic requirements for knowledge of content are not as implausible as they first appear. There may indeed be some independent good reason to accept this claim – nothing I have said entails that no such reason exists, just that, as it stands, it is not at all clear (at least not to me) what this reason could be.

As I said in Chapter 1, that we possess self knowledge is highly plausible intuitively speaking; that I am thinking that the sun is shining, and that I *know* I am thinking this seems as undeniable, if not more so, than the claim that I have hands, and that I *know* I have hands. The denial of self-knowledge is at least as unpalatable as the denial that I have knowledge of the external world.

In Chapter 2, I examined Burge's argument for why critical reasoning requires self-knowledge. If this argument is right, and I think it is, then if we are to give up self-knowledge, then we must give up the idea that we possess the ability to reason critically. To deny that we are capable of critical reasoning, is to deny the legitimacy of our epistemic practices of examining our beliefs, and altering them as we see appropriate. Critically engaging with our beliefs, it would follow, adds nothing to the epistemic status of these beliefs.

It must be accepted that arguments threatening the legitimacy of such practices cannot be taken seriously outside the philosophy seminar room. Just as arguments claiming we have no justification to hold any of our beliefs about the external world do not cause us to drop or suspend such beliefs, claims that we do not know, or have no justification for our beliefs about what we are thinking could not lead us to cease forming beliefs on this basis. Even if we were to accept these arguments are correct, it would be a bizarre response for anybody to cease to make claims about what they are thinking, in response to such arguments.

The fact that scepticism is in this manner hard or impossible to take seriously has led some philosophers to treat the sceptical argument less seriously. As Stroud says:

Many would dismiss scepticism and defend not taking it seriously on the grounds that it is not a doctrine or theory any sensible person would contemplate adopting as the truth about our position in the world. It seems to them too frivolous or perverse to concentrate on a view that is not even a candidate in the competition for the true or best theory as to how things are.⁷⁸

Stroud, as we have seen, agrees that we should not seriously accept scepticism, but thinks that it is still important to deal with scepticism, as the argument it uses has – he thinks – no obvious answer. This, Stroud thinks, is troubling, not because it means we ought to give up the idea that we have knowledge of the external world, but rather that it is not at all obvious *how* we can have it. The task in responding to the sceptic, on Stroud's view, becomes one of explaining how we have self-knowledge – *whether* we have it, is no longer at issue. We can assume that we do, but to explain *how* we do requires an answer to scepticism.

Though, even if this is the case, scepticism presents us with the embarrassing situation of us having an argument we cannot fault⁷⁹, with a conclusion we cannot – or even *should* not – believe. This situation is to be avoided – which is why philosophers like Dretske have accepted “abominable” and “bizarre” theses as a means of finding fault with its premises.

If we were to take the Achievement and Consequence Problems to undercut self-knowledge, we would be faced with the same embarrassing situation. But in response to these problems we have an easier and less embarrassing option available – deny externalism.

I will not go into the strengths and weakness of externalism here. I remarked in the first chapter that the general form of externalism discussed here derives its

⁷⁸ Stroud (1984B, p. 2).

⁷⁹ Assuming that we cannot fault it. There will be many, naturally, who do fault it for some reason or another. Yet nonetheless, there will be a great deal of perfectly rational people who have read skepticism, accepted the argument as correct, but find themselves unable to take the conclusion seriously.

plausibility from the Twin Earth thought experiments – externalism seems the best way of accommodating our intuitive response to those sorts of cases. Nonetheless, the cost of abandoning externalism here is not as great as the cost of abandoning self-knowledge; the falsity of externalism does not threaten our ability to critically reason, nor is it a conclusion we cannot believe outside the seminar room – in the way that we cannot believe the falsity of self-knowledge outside the seminar room.

As I also remarked, the intuitive support for externalism is diminished if indeed externalism can be seen to have intuitively bizarre consequences – like *either* we lack self-knowledge, or we have to accept Burge’s low requirements for knowledge of content. Given these consequences, the case for externalism loses some of its force. In this way, it seems best to agree with McLaughlin and Tye when they say, and I repeat:

we find [self-knowledge] sufficiently compelling that if Twin Earth externalism is incompatible with this privileged access thesis, that is a powerfully compelling reason for rejecting Twin Earth externalism. Indeed, were we ourselves to come to think that the incompatibilist view is right, we would reject Twin Earth externalism.⁸⁰

Yet it is not clear where this would leave us, in our views regarding mental content. In this dissertation I have not examined whether we can generate any independent reason for accepting that mental content be narrow in the way required for self-knowledge. The considerations I have mentioned here simply suggest that given the lack of any good reason to accept Burge’s low epistemic requirements for knowledge of content, *if* we want a plausible account of self-knowledge, then it seems as we require that mental content be understood to be narrow. This then would be an argument for narrow content from self-knowledge. But to make a successful case for narrow content would require more than this. If we can make no sense of mental content construed narrowly, then this will diminish the plausibility of this argument. If mental content can only be plausibly understood to have wide content, accepting narrow content on the basis of an argument for self-knowledge would be problematic.

⁸⁰ McLaughlin & Tye (1998, p. 289).

Further, and as I have said, externalism and self-knowledge are not the only competing options here – there is also the possible option that we accept Burge’s low epistemic requirements for knowledge of content. This would allow us to keep both self-knowledge and externalism. But again it is not clear whether we can generate any independently plausible grounds for this – these requirements simply appear to be totally different from the requirements we place on every other kind of knowledge; the fact that I know so little about my mental content, on Burge’s view, threatens the plausibility of his claim that I know what it is. Knowing so little about anything else uncontroversially implies that I do not know what it is. Why should this not be the case regarding content? In this way, giving us reason to accept these requirements will require a substantial amount of work.

In this way, I argue that it is not yet clear that our best response to the arguments I have provided would be to abandon externalism. That is an avenue that should be explored, but what must also be explored is whether we can generate any plausible account of knowledge of content that is consistent with Burge’s lower epistemic requirements. But it seems our best option will be to accept one of these two avenues – *either* we accept a theory of narrow content, *or* we accept the lower epistemic standards for knowledge of content. Which of these is best will depend upon the plausibility of these prospective accounts. It is to the examination of such accounts, I submit, that this debate should proceed.

Bibliography

- Alston, William, P. 2005. *Beyond "Justification": Dimensions of Epistemic Evaluation*, Cornell University Press, New York.
- Boghossian, Paul A. 1989. "Content and Self-Knowledge" in *Philosophical Topics* 17, pp. 5 – 26. My page numbers refer to the reprint in *Externalism and Self-Knowledge*, Peter Ludlow and Norah Martin, ed. CLSI Publications, Stanford (1998).
- . 1998. "What the Externalist Can Know A Priori" in *Knowing Our Own Minds*, Crispin Wright, Barry C. Smith and Cynthia Macdonald, ed. Clarendon Press, Oxford.
- Brewer, Bill. 2000. "Externalism and A Priori Knowledge of Empirical Facts" in *New Essays in the A Priori*, Paul Boghossian and Christopher Peacocke, ed. Oxford University Press, Oxford.
- Brown, Jessica. 1995. "The Incompatibility of Anti-Individualism and Privileged Access" in *Analysis* 55 pp. 149 -156.
- Brueckner, Anthony. 1990. "Scepticism about Knowledge of Content" in *Mind*, Vol. 99, pp. 447 – 451.
- . 1992. "What An Anti-Individualist Knows A Priori", in *Analysis* 52, pp. 111 – 118.
- Burge, Tyler. 1979. "Individualism and the Mental" in *Midwest Studies in Philosophy* 4, pp. 73 – 122.
- . 1988. "Individualism and Self-Knowledge" in *The Journal of Philosophy* LXXXV, 11 (November). My page numbers refer to the reprint in *Externalism and Self-Knowledge*, Peter Ludlow and Norah Martin, ed. CLSI Publications, Stanford (1998).
- . 1996. "Our Entitlement to Self-Knowledge" in *The Proceedings of the Aristotelian Society* XCVI. My pages numbers refer to the reprint in *Externalism and Self-Knowledge*, Peter Ludlow and Norah Martin, ed. CLSI Publications, Stanford (1998).
- . 2003A. "Davidson and Forms of Anti-Individualism: Reply to Hahn" in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Martin Hahn and Bjørn Ramberg, ed. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.
- . 2003B. "Epiphenomenalism: Reply to Dretske" in *Reflections and Replies: Essays on the Philosophy of Tyler Burge*, Martin Hahn and Bjørn Ramberg, ed. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.
- Davidson, Donald .1983. "A Coherence Theory of Truth and Knowledge" in *Kant Oder Hegel?* D. Henrich, ed. Stuttgart, Klett-Cotta.
- Davies, Martin. 2000. "Externalism and Armchair Knowledge" in *New Essays in the A Priori*, Paul Boghossian and Christopher Peacocke, ed. Oxford University Press, Oxford.
- DeRose, Keith. 1995. "Solving the Skeptical Problem" in *The Philosophical Review*, Vol. 104, No. 1, pp. 1 - 52
- Dretske, Fred. 2003. "Scepticism: What Perception Teaches" in *The Sceptics: Contemporary Essays*, Steven Luper, ed. Aldershot: Ashgate.
- Fodor, Jerry A. 1995. *The Elm and the Expert*, MIT Press, USA.

- Heal, Jane. 1994. "Moore's Paradox: a Wittgensteinian Approach" in *Mind*, Vol. 103, pp. 5 – 24.
- Horgan, Terence. 1991. "Actions, Reasons, and the Explanatory Role of Content" in *Dretske and His Critics*, Brian McLaughlin, ed. pp. 73 – 101.
- Jones, Ward E. 2000. "Can We Infer Naturalism from Scepticism?" in *Philosophical Quarterly* 50: 201, pp. 433-451.
- McCulloch, Gregory. 1995. *The Mind and Its World*, Routledge, London.
- McDowell, John. 1986. "Singular Thought and the Extent of Inner Space" in *Subject, Thought and Context*, Philip Petit and John McDowell, ed. Oxford University Press, Oxford.
- McKinsey, Michael. 1991. "Anti-Individualism and Privileged Access" in *Analysis* 52, pp. 232 – 237.
- . 2001 "The Semantic Basis of Externalism" in *Meaning and Truth*, J. Campbell, M. O. Rourke, and D. Shier, ed. Seven Bridges Press, New York.
- McLaughlin, Brian and Tye, Michael. 1998. "Externalism, Twin Earth and Self-Knowledge" in *Knowing Our Own Minds*, Crispin Wright, Barry C. Smith and Cynthia Macdonald, ed. Oxford University Press, Oxford.
- Millikan, Ruth. 1989. "Biosemantics" in *The Journal of Philosophy*, Vol. 86, No. 6, pp. 281 – 297.
- Moran, Richard. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton University Press, Princeton.
- Papineau, David. 1984. "Representation and Explanation", in *Philosophy of Science* Vol. 51, No. 4, pp. 550 – 572.
- Putnam, Hilary 1975. "The Meaning of 'Meaning'" in *Philosophical Papers 2* Cambridge University Press, Cambridge.
- . 1991. *Reason, Truth, and History*. Cambridge University Press: Cambridge.
- Shoemaker, Sydney. 1988. "On Knowing One's Own Mind" in *Philosophical Perspectives* 2, *Epistemology*. James E. Tomberlin, ed. Ridgeview Publishing Co.: Atascadero. My page numbers refer to the reprint in *The First Person Perspective and Other Essays*, Cambridge Studies in Philosophy, USA. (1996).
- . 1995. "Moore's Paradox and Self-Knowledge" in *Philosophical Studies* 77, pp. 211 – 228.
- Stern, Robert. 1999. Introduction to *Transcendental Arguments: Problems and Perspectives*, Robert Stern ed. Oxford University Press, New York.
- Stroud, Barry. 1968. "Transcendental Arguments" in the *Journal of Philosophy* (1968). My page numbers refer to the reprint in *Understanding Human Knowledge*, Oxford University Press, New York (2000).
- 1984A. *The Significance of Philosophical Scepticism*, Oxford University Press, Oxford.
- 1984B "Scepticism and the Possibility of Knowledge" in *The Journal of Philosophy* (1984). My page numbers refer to the reprint in *Understanding Human Knowledge*, Oxford University Press, New York (2000).
- 1989. "Understanding Human Knowledge in General" in *Knowledge and Scepticism*, M. Clay and K. Lehrer, ed. Westview Press: Boulder, Colorado. My page numbers refer to the reprint in *Understanding Human Knowledge*, Oxford University Press, New York (2000).

- Van Fraassen, Bas C. 1977. "The Pragmatics of Explanation" in *American Philosophical Quarterly*, Vol. 14, No. 2, pp. 143 – 150.
- Warfield, Ted A. 1992. "Privileged Self-Knowledge and Externalism are Compatible" in *Analysis* 52, pp. 232 – 237. My page numbers refer to the reprint in *Externalism and Self-Knowledge*, Peter Ludlow and Norah Martin, ed. CLSI Publications, Stanford (1998).
- Wright, Crispin. 1991. "Scepticism and Dreaming: Imploding the Demon" in *Mind*, Vol. 100, pp. 87 - 116.
- 1998. "Self-Knowledge: The Wittgensteinian Legacy" in *Knowing Our Own Minds*, Crispin Wright, Barry C. Smith and Cynthia Macdonald, ed. Oxford University Press, Oxford.