

Semantic Externalism and Self Knowledge: Privileged Access to the World

Submitted for the degree of PhD
Sarah Ann Sawyer
King's College London
September 1997

Acknowledgements

I would like to thank David Papineau for a multitude of comments and suggestions on all aspects of the thesis.

I would also like to thank Christoph Hörl and Mark Sainsbury for comments on the final draft, and Sean Crawford for extensive discussion of the nature of *de re* thought.

Special thanks go to my parents for providing a solid basis from which to work.

Abstract

The thesis is concerned to defend the compatibility of two plausible claims about the mind; semantic externalism and privileged access. It is further concerned to demonstrate one important implication of the conjunction of semantic externalism and privileged access, an implication which forces the rejection of the dichotomy between knowledge of one's mind and knowledge of one's world.

Chapter one is a presentation of semantic externalism. Chapter two is a presentation of the claim of privileged access. The claim of privileged access is formulated in response to the following question. How can a subject have privileged access to the contents of her thoughts given that her thoughts depend essentially on contingent facts about her world of which she could have empirical knowledge only?

Chapter three is concerned with the following implication, the consequent of which is *prima facie* absurd. If the contents of the mind depend essentially upon contingent facts about the world, knowledge of the semantic contents "within" can yield knowledge of the world "without". Chapter four is a defence of the consequent. It is argued that the apparent absurdity of non-empirical knowledge of the world arises from a failure to embrace the full force of semantic externalism. We can have privileged access to the world.

Chapter five is an examination of the nature of *de re* thought. Whether the content of a *de re* thought is to be understood as object-dependent or as object-independent will determine the extent to which we can have privileged access to the world: that is, whether we can have privileged access to the objects of our *de re* thoughts as well as to general facts about our world.

Chapter six focuses on the implications of the thesis for external-world scepticism.

Contents

Chapter One: Semantic Externalism **p.**

6

- 1.1 Introduction
- 1.2 Putnam's argument
- 1.3 Burge's proposal: semantic externalism and semantic internalism
- 1.4 Burge's argument
- 1.5 The acquisition of terms
- 1.6 Is water really H₂O?
- 1.7 Conclusion

Chapter Two: Travelling Cases **p.**

30

- 2.1 Introduction
- 2.2 Knowledge of one's thoughts
- 2.3 Comparative content and bare content
- 2.4 The Content Sceptic's Argument
- 2.5 Relevant alternatives
- 2.6 (RA) or (RA')?
- 2.7 The privileged access claim
- 2.8 A formalisation of the privileged access claim
- 2.9 Unconscious mental states and past mental events
- 2.10 The Argument from Memory
- 2.11 The temptation of the arguments
- 2.12 Conclusion

Chapter Three: The Argument from Privileged Access

p. 64

- 3.1 Introduction
- 3.2 The McKinsey recipe
- 3.3 The first crucial issue: "necessarily depends upon"

- 3.4 The second crucial issue: the proposition E
- 3.5 Brown's elaboration of the McKinsey recipe
- 3.6 Tye and McLaughlin
- 3.7 Conclusion

Chapter Four: Privileged Access to the World **p. 90**

- 4.1 Introduction
- 4.2 Compatibilist responses
- 4.3 Instance arguments reconsidered
- 4.4 A necessary causal connection
- 4.5 What should we say about Vatbrain?
- 4.6 Conclusion

Chapter Five: Singular Thought **109**

p.

- 5.1 Introduction
- 5.2 *De dicto* and *de re* attributions
- 5.3 Burge's distinction
- 5.4 Object-dependent thoughts
- 5.5 Against Burge
- 5.6 Deluded subjects
- 5.7 The Two List Argument
- 5.8 Instance arguments
- 5.9 Conclusion

Chapter Six: Scepticism **140**

p.

- 6.1 Introduction
- 6.2 Putnam's argument
- 6.3 Some qualifications and a proof
- 6.4 Brueckner's problem
- 6.5 The remains of scepticism
- 6.6 Conclusion

References
154

p.

Semantic Externalism

No man's intentional mental phenomena are insular. Every man is a piece of the social continent, a part of the social main. (Burge, 1979, p. 87)

1.1 Introduction

The writings of Descartes gave voice to an assumption which has guided much philosophical enquiry since. The assumption is that of *methodological solipsism*. Methodological solipsism has it that no psychological state presupposes the existence of any individual other than the subject to whom the state is ascribed. Thus a psychological state can remain fixed across various logically possible environments. In Descartes, one of the ways in which this manifests itself is as the logical possibility that for any psychological state P, it is logically possible for a subject in a world consisting of nothing other than the non-physical subject to be in P. Nowadays, with the current dominance of one or another form of physicalism, the thought is expressed differently. The assumption manifests itself rather as the logical possibility that for any psychological state P, it is logically possible for a subject with no prior causal contact with an external physical environment to be in P. Supposed examples of such a subject would be that of an accidental replica, or a brain-in-a-vat.

At first sight, relational states are not, on this view, psychological states. That is, such states as being jealous, jealousy being a two-place relation, presuppose the existence of an object as well as a subject, and hence violate the assumption of methodological solipsism.¹ One could draw a distinction between “narrow” psychological states, which

¹ There is a popular (perhaps default) view according to which all psychological states are *relations to propositions*. This view is compatible with the assumption of methodological solipsism as long as a proposition is not thought of as an individual.

respect the assumption of methodological solipsism, and “broad” or “wide” psychological states, which do not. If relational states *are* to be construed as narrow psychological states, they would have to be construed such that the supposed object of the relation falls within the bounds of the subject; the most obvious way to do this would be to take the object of the relation as a perception, or image.² The assumption would then encompass relational states.³ The assumption of methodological solipsism gives rise to the possibility of radical external-world scepticism, since it is consistent with belief in an external world that no such world exist.

The body of philosophers most fervently committed to the assumption of methodological solipsism are marked by their commitment to the project of a scientific psychology.⁴ Their general line of thought runs as follows. The purpose of a scientific psychology is to formulate psychological laws: that is, laws under which the psychological states of subjects can be subsumed. Such laws must be apt for the explanation and prediction of the actions of those subjects. If scientific psychology is to be a viable project, it would be plausible to classify psychological states according to causal powers, and plausible that these causal powers in turn be classified in a way which is independent of the particular context in which the subject of the psychological state happens to be located. That is, content is to be individuated by causal powers, and causal powers are to be individuated narrowly.⁵ The meaning

² It would then be possible to account for phenomena such as “paranoid jealousy”, where one is jealous of a non-existent rival.

³ There is a further distinction between psychological states and *factive* states, such as that of knowing that one is in England. Such factive states cannot be reconstrued in terms of relations to further mental items, since it is precisely their relation to external states of affairs which distinguishes them as the particular factive states they are. This is not to rule out the possibility of knowledge of internal items.

⁴ For the most ardent advocate of this view see Fodor (1980). See also Stich (1978).

of a psychological state will therefore be context-independent.⁶

It would be wrong, however, to think that a commitment to the assumption of methodological solipsism follows from a commitment to the project of a scientific psychology. The line of thought outlined above exemplifies a prior, perhaps metaphysical commitment to methodological solipsism; prior that is to any engagement with the discipline. One could argue that this gets things the wrong way round, and that the correct approach would rather be to examine theories from scientific psychology, and so determine whether there is *de facto* a commitment to methodological solipsism. This approach is evident in a philosophical debate issuing from David Marr's theory of vision. Thus Tyler Burge and Martin Davies have argued that Marr's theory of vision delivers broad perceptual contents, while Gabriel Segal has argued that the perceptual contents delivered are narrow.⁷

In his "The Meaning of 'Meaning'", Hilary Putnam poses a challenge to the assumptions underlying what he takes to be the traditional theory of meaning. The traditional theory of meaning has it that the meaning of a word in a language is determined solely by the psychological state of the subject who uses that word. This challenge in turn prompted Burge to mount a direct attack on the assumption of methodological

⁵ According to this line of thought, two states will be classified as having the same causal powers if they cause the subject to move in the same way. This is, however, a complex issue. It is possible both for the bodily movements of two subjects to be identical while their actions are diverse, and for the actions of two subjects to be identical while their physical movements differ.

⁶ For an attempt to combine semantic externalist intuitions with the motives driving the assumption of methodological solipsism see Fodor (1994).

⁷ For Marr's theory of vision see Marr (1982). For arguments against its being individualistic see Burge (1986b) and Davies (1991) and (1992). For arguments in favour of its being individualistic see Segal (1989b) and (1991).

solipsism.⁸ Putnam's argument concerns linguistic meaning, while Burge's concerns psychological content. According to Putnam, the meanings of a subject's *words* are not determined solely by facts about the individual subject. According to Burge, the meanings of a subject's *thoughts* are no more determined solely by facts about the individual subject than the meanings of her words. The meaning of a psychological state is precisely *not* context-independent. My aim in this chapter is to set out the arguments presented by Putnam and Burge, and to draw attention to certain features which are of particular relevance for the thesis as a whole. The chapter is not intended primarily as an endorsement of the claim that context is constitutive of meaning, whether linguistic or psychological. Rather, it is intended to serve as a preliminary statement of precisely what the claim amounts to. The concluding section presents an overview of the contents of the thesis, and explains how the chapters which follow relate to each other.

1.2 Putnam's argument

Putnam identifies two traditional claims about meaning and the mind: first, that knowing the meaning of a term is just a matter of being in a certain psychological state, and second, that the meaning of a term (its intension) determines its extension. Putnam challenges the cotenability of these claims, arguing that they are not jointly satisfied by any notion, and *a fortiori* not by a theory of meaning. Putnam argues, by means of the (by now all too familiar) thought experiment set out below, that psychological state does not determine extension, and hence that one of the traditional claims must be rejected. "Cut the pie any way you like, 'meanings' just ain't in the *head*."⁹

This is how Putnam presents his argument for the claim that psychological state does not determine extension. We are to suppose that somewhere in the galaxy there is a planet

⁸ Burge (1979).

⁹ Putnam (1975b) p. 227, original emphasis.

we shall call Twin Earth. Twin Earth is exactly like Earth in all but a few respects. People on Twin Earth even speak a language with the same syntax and at least mostly the same semantics as English, a language which could be regarded as a dialectical version of English. One of the respects in which Twin Earth differs from Earth is that the liquid people on Twin Earth call “water” is not H₂O but a different liquid whose chemical formula, being very long and complicated, is abbreviated simply as XYZ. XYZ is supposed to be indistinguishable from water at normal temperatures and pressures, and, further, to fill the same role on Twin Earth as H₂O does on Earth. That is, it quenches thirst, fills the lakes and seas, and so on.

Putnam claims that if a spaceship from Earth were to visit Twin Earth, the travellers would at first suppose that the water-like stuff they encounter, and which the locals call “water”, *is* water, and hence that “water” has the same meaning on Twin Earth as on Earth. However, Putnam claims, this supposition will be corrected when it is discovered that the water-like stuff on Twin Earth is in fact XYZ. The case will be symmetrical for visitors from Twin Earth; they will at first suppose that their term “water” has the same meaning as the syntactically identical term in English, a supposition which will be corrected when they discover that “water” on Earth is H₂O.

Note that there is no problem about the extension of the term ‘water’. The word simply has two different meanings ... : in the sense in which it is used on Twin Earth, the sense of water_{TE}, what we call ‘water’ simply isn’t water; while in the sense in which it is used on Earth, the sense of water_E, what the Twin Earthians call ‘water’ simply isn’t water. The extension of ‘water’ in the sense of water_E is the set of all wholes consisting of H₂O molecules, or something like that; the extension of ‘water’ in the sense of water_{TE} is the set of all wholes consisting of XYZ molecules, or something like that. (Putnam, 1975b p. 224, original emphasis)

Putnam’s further claim is that the meaning of “water” both on Earth and on Twin Earth has remained constant over time.

Considering a typical speaker, Oscar₁, from Earth in 1750, and his counterpart, Oscar₂, from Twin Earth, Putnam writes,

You may suppose that there is no belief that Oscar₁ had about water that Oscar₂ did not have about 'water'. If you like, you may even suppose that Oscar₁ and Oscar₂ were exact duplicates in appearance, feelings, thoughts, interior monologue, etc. Yet the extension of the term 'water' was just as much H₂O on Earth in 1750 as in 1950; and the extension of the term 'water' was just as much XYZ on Twin Earth in 1750 as in 1950. Oscar₁ and Oscar₂ understood the term 'water' differently in 1750 *although they were in the same psychological state*, and although, ... , it would have taken their scientific communities about fifty years to discover that they understood the term 'water' differently. Thus the extension of the term 'water' ... is *not* a function of the psychological state of the speaker by itself. (Putnam, 1979 p. 224, original emphasis)

The traditional conception that meaning what one does should be classified as a state of mind is thus brought under pressure. Oscar₁ and Oscar₂ are psychologically identical, and yet their counterpart linguistic terms "water" differ in meaning. One of the traditional claims about meaning must be discarded: either knowing the meaning of a term is not just a matter of being in a certain psychological state, or the meaning (intension) of a term does not determine its extension. Putnam opts to discard the former, and retain the latter. Knowing the meaning of a term according to Putnam then, is not just a matter of being in a certain psychological state "narrowly construed".

Putnam's positive proposal is that a theory of meaning should be seen as composite, essentially incorporating the external world to which the subject bears a contingent relation. The meaning of a linguistic term is, according to this view, constituted by the set composed of a "narrow" psychological state together with an external relational fact. That Putnam accepts some notion of narrow content is

evident in his assertion that Oscar₁ and Oscar₂ may be supposed to be *duplicates in thought*.¹⁰

Perhaps the easiest way to understand Putnam's positive proposal is by drawing attention to a parallel with an approach to indexical terms introduced by David Kaplan.¹¹ Consider the sentence "I am thirsty". If Oscar₁ and Oscar₂ were to utter that sentence, they would thereby have expressed different propositions. The extension of an indexical term depends upon the context of utterance. The extension of the indexical term "I", for instance, depends upon the identity of the speaker; the extension of the indexical term "here" depends upon the location of the speaker. Nevertheless, the utterances of Oscar₁ and Oscar₂ clearly have *something* in common. Kaplan calls this common element "character".¹² On this view, then, two terms are said to have the same character if they effect the same mapping of utterance plus context onto truth-conditions. For standard indexical terms, such as "I", "here" and "now", the extension is clearly not a function solely of the narrow psychological state of the subject. Furthermore, as Putnam notes, such a theory has never been suggested for indexical terms.¹³ If Putnam is right, for a natural kind term the extension is

¹⁰ While I take Putnam's commitment to some form of narrow content to be evident, he does appear to have some qualms about the assumption of methodological solipsism. Given that his paper was programmatic, I do not think this is particularly surprising, but his ambivalence is confirmed in Putnam (1996).

¹¹ Kaplan (1980).

¹² Perry (1977) introduces the notion of psychological "role" which serves essentially the same purpose as Kaplan's notion of character.

¹³ It should be remembered that the issue under consideration here is that of *linguistic* indexical terms. For an argument to the effect that the meaning of a psychological indexical is a function solely of the psychological state of the subject see Evans (1981). There is some reason to think that such indexical components of thought can be neither eliminated in favour of, nor reduced to, non-indexical components. For a convincing explanation of this see Perry (1979).

similarly not determined by the psychological state of the subject.

Putnam writes,

Our theory can be summarized as saying that words like 'water' have an unnoticed indexical component: 'water' is stuff that bears a certain similarity relation to the water *around here*. Water at another time or in another place or even in another possible world has to bear the relation [same-liquid] to *our* 'water' in order to be water. (Putnam, 1975b p. 234, original emphasis)

This analogy between indexical expressions and natural kind terms is by no means perfect. The following disanalogy is surely important. The *extension* of indexical type-expressions is *context-dependent*, and yet indexical type-expressions are nevertheless taken to have a constant *meaning*. Hence contextual factors do not enter into the meaning of indexical expressions. The indexical expression "I" means the same when uttered by Oscar₁ and Oscar₂, even though the extension of each utterance differs. With respect to natural kind terms, Putnam's line appears to be the following. The extension of a natural kind expression is context-dependent, and *therefore* natural kind type-expressions do not have a constant meaning. Contextual factors *do* enter into the meaning of natural kind terms.

Using this disanalogy as an argument against Putnam's claim would, however, be unfair to Putnam, who himself draws attention to this fact.¹⁴ Putnam's claim is that, for indexical type-expressions and for natural kind type-expressions alike, *psychological state does not determine extension*. In both cases we must choose one of two possible options, each of which correspond to the rejection of one or other of the two doctrines underlying the traditional theory of meaning as characterised by Putnam. To choose the first option would be to maintain that the type-expressions have the same meaning despite a possible difference in extension; thereby giving up the doctrine that meaning determines

¹⁴ Putnam (1975b) pp. 245-6.

extension. To choose the second option would be to maintain that difference in extension is *ipso facto* difference in meaning; which is inconsistent with the doctrine that knowing the meaning of a term is a matter of being in a certain psychological state. It is surely right to point out that the first option embodies the standard approach to indexicals, whereas Putnam's preferred option with respect to natural kind terms is the second.¹⁵ However, this does not damage the parallel Putnam wishes to draw. The parallel holds to the extent that in neither case does psychological state determine extension.¹⁶

The adoption of such a composite theory of meaning allows that the investigation of the mental can remain a scientific investigation along the lines proposed by for example Jerry Fodor and Stephen Stich. The proper concern of a scientific psychology, with its commitment to individuation by causal powers, must, according to them, be *narrow* psychological facts, and, since Putnam's positive proposal admits the notion of narrow content, the project of a scientific psychology is left intact. Nevertheless, if Putnam's proposal is accepted, the discipline has turned out to be concerned with items which are other than originally envisaged. Scientific psychology cannot deal directly with *meaning*, since,

¹⁵ Putnam also draws a parallel between natural kind terms and rigid designators. This is perhaps a better analogy, since for rigid designators, difference in extension makes for difference in meaning. Burge's contention (1982) is that Putnam's assimilation of natural kind terms *both* to indexicals *and* to rigid designators is actually inconsistent. This may well be right, but, again, only once one has committed oneself to the first possible option for indexicals. The parallel with indexicals is surely to be drawn prior to any such commitment. It might be thought that indexicals are indeed rigid designators. Consider the following line of thought. If Fred says "I am tall", what he expresses is true in any world *w* if and only if Fred is tall in *w*. Thus his utterance of "I" rigidly designates himself. This assumes sameness of subject across different contexts.

¹⁶ For further discussion of natural kind terms and the possibility of treating them as having an indexical component see for example Zemach (1976), Mellor (1977), and Sterelny (1981).

following Putnam's line, linguistic meaning incorporates facts external to the subject.

1.3 Burge's proposal: semantic externalism and semantic internalism

Subsequent to Putnam's argument, Burge advocates an alternative more radical proposal. Burge accepts Putnam's primary claim that a difference in meaning can be due to a difference in *context alone*, that the meaning of a linguistic term is not context-independent. However, he maintains that the natural consequence of this observation is that a difference in context can *equally* make for a difference in *thought*. He even expresses some incredulity as to "why Putnam did not draw a conclusion so close to the source of his main argument"¹⁷. Rather than holding onto the claim that, while linguistically different, Oscar₁ and Oscar₂ are alike in their thoughts, Burge rejects entirely the familiar conception of the mental. On this view, the counterparts are treated as differing *psychologically*. Even though what is in the head may be thought of as *causally* relevant, it can no longer be regarded as being of *constitutive* relevance to the meaning of one's state of mind.¹⁸ On the Burgean picture there is no narrow content; all content is broad. The environment permeates to the very heart of individual psychology.

¹⁷ Burge (1982) p. 103.

¹⁸ The issue which concerns the individuation of the causal powers of psychological states is an interesting and contentious one. If Putnam is right, it might seem that the same narrow psychological states could nevertheless differ in their causal powers in virtue of being had by subjects who were related to different physical environments. Psychological state P could cause Oscar₁ to reach for H₂O, and Oscar₂ to reach for XYZ. If such subjects *are* in psychological states which differ in causal powers, they should presumably not be grouped together for the purposes of psychological explanation. For an interesting discussion of these issues, and an argument *denying* any relevant difference in causal powers see Fodor (1987) and (1991).

I will refer to those theories which hold onto some notion of narrow content for the purposes of psychological explanation *semantic internalist* or *individualistic* theories. Those which jettison such a notion will be referred to as *semantic externalist* or *anti-individualist* theories.¹⁹ Semantic internalism and semantic externalism are theses concerned with the status of the contents of psychological states.

The most neutral way to explicate the complex notion of narrow content is in terms of a *supervenience* claim, where this is to be understood as a determination relation, and not as a dependence relation.²⁰ Thus content that supervenes locally on the subject is termed narrow content. How to delineate the subvening base is a moot issue. Should it be thought of as the brain, the brain plus central nervous system, or something even wider? Certainly the base will not extend beyond the bodily confines of the individual.²¹

Just how to characterise the notion of narrow content is the issue which gives substance to one of the most fundamental debates within semantic internalism; whether it should be characterised as non-truth-conditional, akin to Kaplan's notion of character, or rather as everyday, truth-conditional

¹⁹ I will throughout refer to such theories as *semantic* internalism and externalism, to contrast with what I term *epistemic* internalism and externalism. For an account of the theories of epistemic internalism and epistemic externalism see Lehrer (1974), Pollock (1986), and Chisholm (1989). See also Nozick (1981), and Goldman (1986).

²⁰ The importance of understanding supervenience in this context in terms of a determination relation as opposed to a dependence relation can be illuminated by seeing how the notions come apart. It is plausible to hold that the property of being coloured depends upon the physical properties of objects, and hence that there could be no change in the colour without a change in the physical, while maintaining that colour is response-dependent, that what determines that an object is red, for instance, is nothing other than our reaction to it. For a thorough investigation into the notion of supervenience see Kim (1993).

²¹ Perhaps the notion of the individual ought not in this context to be taken for granted. The extent of an individual's psychological phenomena is, after all, precisely what is at issue.

content. Understanding it as non-truth-conditional content is relatively easy, precisely because we have the Kaplan notion of character to which to appeal. However, there is a question as to whether narrow content understood in this way, as non-truth-conditional, is really *content*. Because of this question, some have tried to defend a form of truth-conditional narrow content, arguing as follows.²² The attribution of a thought content is made so as to facilitate the explanation and prediction of a subject's actions. A correct attribution will therefore account for difference and sameness in behaviour. If we consider Oscar₁ two things are apparent. First, he is unable to distinguish between H₂O and XYZ, even after multi-modal interaction. That is, both substances prompt identical forms of behaviour narrowly construed. Second, Oscar₁'s behaviour is indiscriminable from Oscar₂'s behaviour from an internalist perspective. According to this line of thought, it would be theoretically unmotivated to attribute a subject with a content which credits her with discriminatory powers she clearly does not possess. It would be similarly theoretically unmotivated to attribute two behaviourally indistinguishable subjects with *different* contents. Not only is there no *reason* to attribute the subjects with such fine-grained thought contents, it would be wrong to do so. Rather, in the case of Oscar₁ and Oscar₂ the subjects should be attributed a content which has in its extension *both* H₂O *and* XYZ. *This* is the attribution which best explains Oscar₁'s behaviour with respect both to water and twin water, and equally explains the behaviour of Oscar₂. This is not, as some have suggested, to attribute Oscar₁ or Oscar₂ with disjunctive contents.²³ Many of our terms have extensions which range over different types of object; "animal", for example. That a category of objects *can* be further subdivided provides no basis on which to attribute a subject who is unable to discriminate between the subdivisions within the broader category a thought-content which is specific to one of the subdivisions. While this

²² For a thorough exposition of the view see Segal (1989a), (1989b) and (1991). See also Crane (1991).

²³ See for example Davies (1991) and (1992), and Egan (1991).

issue is an important one, the debate is internal to semantic internalism, and since our concern here is with semantic *externalism*, we shall dwell on it no further. To reiterate, semantic externalism and semantic internalism are theses concerned with the content of psychological states. The former admits some form of narrow content; the latter does not.

According to these definitions, it is clear that Putnam must be classified as a semantic *internalist*. He is, after all, committed to some form of narrow content. His contribution was not, therefore, the introduction of semantic externalism. Rather, it was to show that psychological state does not determine extension. Semantic externalism came with the recognition, by Burge, of the implications of Putnam's thesis for the philosophy of mind. Putnam has since confessed that although he appreciated the fact that denying that meanings are in the head must have consequences for the philosophy of mind, at the time of writing "The Meaning of 'Meaning'", he was unsure as to just what those consequences were. Putnam has subsequently acknowledged that the notion of narrow content ought to be jettisoned, showing an allegiance to semantic externalism. He writes,

In Burge's view, my attempt in "The Meaning of 'Meaning'" to hold a place open for a notion of "narrow content" and for "narrow mental states" represented a confusion on my part, and I have come to believe that he is right. (Putnam, 1996 p. xxi)

My primary concern will be with semantic externalism – a thesis about the constitutive conditions for the contents of one's psychological states. To this extent I focus on the thesis presented by Burge, the argument for which is set out in section **1.4** below.

1.4 Burge's argument

Putnam's thesis as presented so far is a thesis about natural kind terms; terms which pick out natural kinds in the world,

demarcated as it would be by the best scientific taxonomy.²⁴ However, Putnam's argument, as he states, is not restricted to natural kind terms, but works equally for other types of words. Examples offered by Putnam include "elm", "mackerel", and "aluminium". Again, these examples turn on the supposition that items which fall within the extension of a given term have a shared underlying structure, whether that be genetic, biological, or molecular. In not all cases will the physical environment be the contextual factor which accounts for the divergence in meaning between the counterpart terms of counterpart subjects. Burge's argument places greater emphasis on the claim that *social context* is constitutive of content, and not external to it; on the role of the linguistic community. The extent of the argument is therefore best brought home by Burge.

The argument has an extremely wide application. ... We could have used an artifact term, an ordinary natural kind word, a color adjective, a social role term, a term for a historical style, an abstract noun, an action verb, a physical movement verb, or any of various other sorts of words. (Burge, 1979 p. 79)

Burge's argument, referred to in the above quote, is presented as a three-step thought-experiment. He writes,

Suppose first that:

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

²⁴ The notion of a natural kind is not a clear one. See for instance Wilson (1982), with whose line of argument I am in broad agreement.

he has developed arthritis in the thigh. (Burge, 1979 p. 77)

On reporting his fear to his doctor, the patient is surprised to find out that he cannot have arthritis in his thigh, since arthritis is specifically an inflammation of joints. He accepts the doctor's word, and relinquishes his belief. The second step of the thought-experiment consists of a counterfactual supposition. We are asked to conceive of a situation in which the patient proceeds from birth through the same course of physical events, up to and including the time at which he first reports his fear to his doctor. In the actual case, "arthritis" as used in his community does not apply to ailments outside joints. In the counterfactual case, however, we are to imagine that "arthritis" as used by the community *does* apply to ailments outside joints. Burge summarises the second step as follows.

The person might have had the same physical history and non-intentional mental phenomena while the word 'arthritis' was conventionally applied, and defined to apply, to various rheumatoid ailments, including the one in the person's thigh, as well as to arthritis. (Burge, 1979 p. 78)

The final step is given as an interpretation of the counterfactual situation. Since differences in *extension* make for differences in *meaning*, the word "arthritis" in the counterfactual community, according to Burge, does not mean *arthritis*: it does not apply to only inflammation of the joints. The crucial pair of suppositions is as follows: a) that in the counterfactual case it would be incorrect to ascribe any content clause containing an oblique occurrence of the actual term "arthritis"; and b) that such differences in extension of counterpart expressions in that-clauses make for differences in the *content* of the psychological states thereby attributed. Burge in effect argues that although Putnam is right to maintain that meanings are not in the head, any difference in context which makes for a difference in linguistic meaning will

in turn carry over into the realm of the mental, and make for a difference in psychological content.

It might be objected that Burge's inference from a) to b) expresses a commitment to the following contentious claim: noun phrases that embed sentential expressions in mentalistic idioms provide the content of the mental state or event in question. It might appear that without the assumption of that claim, there would be no way to infer from the difference in that-clauses in the attributions of propositional attitudes to Oscar₁ and Oscar₂ to the claim that the contents of their psychological states similarly differed.²⁵ However, I think it is reasonable to concede that "differences in the *extension* – the actual denotation, referent, or application – of counterpart expressions in that-clauses will be semantically represented, and will ... make for difference in content."²⁶ Hence, if it would be incorrect in the counterfactual situation to ascribe any content clause containing an oblique occurrence of the term 'arthritis', it would be incorrect in the counterfactual situation to attribute any psychological states which had a content containing the notion of arthritis.

In the counterfactual situation, the patient lacks some – probably *all* – of the attitudes commonly attributed with content clauses containing 'arthritis' in oblique occurrence. He lacks the occurrent thoughts or beliefs that he has arthritis in the thigh, that he has had arthritis for years, that stiffening joints and various sorts of aches are symptoms of arthritis, that his father had arthritis, and so on. (Burge, 1979 p. 78)

In conclusion,

The upshot of these reflections is that the patient's mental contents differ while his entire

²⁵ For an argument which claims just this see Loar (1988).

²⁶ Burge (1979) p. 75. This is not to say that differences in that-clauses always make for a difference in the contents of the psychological states attributed. It should be allowed that Pierre and Peter say the same thing by their respective utterances of "This sentence is false" and "Cette phrase est fausse".

physical and non-intentional mental histories, considered in isolation from their social context, remain the same. (Burge, 1979 p. 79)

The traditional conception of *psychological facts* as facts that hold independently of the subject's relation to the external world is brought under pressure.

The argument trades upon the possibility of attributing a mental state or event whose content involves a notion that the subject understands only partially or incompletely. There is an interesting issue as to just how much understanding is required before we are willing to attribute a mental state or event whose content contains a notion that the subject partially understands; as to how much ignorance an attribution will tolerate. No uniform answer will be forthcoming.²⁷ However, I agree with Burge that the cases in which we are willing to so attribute are common enough for this issue not to be to the point. As so often, context takes the lead, and the attribution will depend on the interests of the attributer and the audience. Seldom is an attribution made in isolation; attributions are made in bundles, with explanations and qualifications to bolster them.

In addition, the fact that the argument trades upon such a possibility should not be taken to restrict the thesis of semantic externalism to the claim that social context is constitutive *only* of those contents which involve such a partially understood notion. Rather, "*even those propositional attitudes not infected by incomplete understanding depend for their content on social factors that are independent of the individual, asocially and non-intentionally described*"²⁸. It is, after all, a contingent matter that a subject understands a given notion as well as she does.

The argument presented by Burge illustrates the following point. The traditional conception according to which psychological facts are facts which hold independently of the

²⁷ For a more detailed discussion of this issue, intended as a defence of the view that partial understanding is commonplace see Burge (1979) pp. 92-103.

²⁸ Burge (1979) pp. 84-5, original emphasis.

subject's relation to the external world fails to acknowledge the essential, constitutive contribution to meaning of the *social* environment as well as of the *physical* environment. Burge's focus on social factors emphasises the important point that the truth of semantic externalism does not depend on the plausibility of Twin Earth cases. Even were one adverse to Twin Earth thought experiments, semantic divergence is common enough between actual communities for the thesis of semantic externalism to get a hold.²⁹

1.5 The acquisition of terms

Putnam's aim is not to question whether or not one can have knowledge of one's linguistic terms, but rather to argue that whatever knowledge a subject *does* have of her terms cannot be explained purely by appeal to the psychological state of that subject. Clearly, neither can such knowledge be a matter of knowing everything about the extension of that term. To say that the subject who could not distinguish water from XYZ did not know the meaning of the word "water" would be, according to Putnam, to confuse lack of scientific knowledge with lack of linguistic competence.³⁰

There is, then, a question as to how, in accord with the new theory of meaning according to which meaning is partly determined by factors external to the subject, a subject can *acquire* a term: how she can come to understand it, to know what it means. With a term such as "water", whose meaning is fixed partly by facts about the physical environment, the question is this: how can a subject get her mind around the nature in question, though ignorant of the science that demarcates it? With a term whose meaning is fixed partly by social factors, the question is rather: how can a subject know

²⁹ While it is true that Burge's thesis of semantic externalism is argued for on the basis of a thought experiment, the thought experiment to which he appeals does not involve the dubious assumption that there could be duplicate natural kind substances which appeared and functioned in exactly the same way as actual natural kind substances.

³⁰ Putnam (1996) p. xvi.

what a term “X” means, whilst having only a partial understanding of X or Xs? Naturally enough, the theory suggests an answer which is two-fold. In response to the first question, it is suggested that one would appeal to the ways in which exemplifications of the nature have figured in the subject’s cognitive and practical dealings with the world. According to this view, in learning the word “water”, one is in effect learning that the word refers to *that* stuff. The reference of a natural kind term is, perhaps, fixed by an ostensive indication of a paradigm instance. It is clearly consistent to hold that indexicals frequently play a part in the *introduction* of natural kind terms, even though the natural kind terms themselves do not partake in the indexicality of their introducers.

It is important to emphasise that it is consistent with the initial assumption of methodological solipsism to hold that as a matter of fact we acquire terms and concepts in much the same way as is being advocated here; that is, by cognitive and practical dealings with the world we inhabit only contingently. Even faithful individualists can hold a causal theory of language-acquisition. However, it is definitive of semantic internalism that it allow for the *logical* possibility that a subject could have had no such dealings with the world, and yet have just the same terms and concepts. For individualism, the constitutive focus is not on the nature of the actual world, as it is for semantic externalism.

For a term whose meaning is determined partly by social factors, knowing the meaning of that term must be understood rather as a matter of participating successfully in what Putnam calls the “linguistic division of labor”. In effect, this is nothing more than an appeal to experts to fix the meanings of one’s terms. The intention of the subject is that the meaning of her term, and hence its extension, coincide with that of the experts.

Every linguistic community ... possesses at least some terms whose associated ‘criteria’ are known only to a subset of the speakers who acquire the terms, and whose use by the other speakers

depends upon a structured cooperation between them and the speakers in the relevant subsets. (Putnam, 1975b, p. 228)

This is also known as linguistic deference.³¹ Putnam's claim about the acquisition of a term is as follows.

Whenever a term is subject to the division of linguistic labor, the 'average' speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state *certainly* does not fix its extension; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension. (Putnam, 1975b p. 229)

In the case of a term whose meaning depends purely on social factors, the communal standards will have to be conceived as realised in the detailed practice of an actual subcommunity whose members count as experts. In the case of natural kind terms, on the other hand, there need be no members of the community who *actually* know the stuff's chemical structure. All that is required is that there *could* be experts.

This then leads on to the question as to how the roles of the physical and the social environment in individual psychology are to be understood to be related. I propose to place the emphasis on the role of the social community. Thus, I maintain that the physical environment has its constitutive significance for individual psychology only as mediated by the social environment.³² While I agree that water is H₂O, and

³¹ Contrary to the claims of Putnam, Fodor has since remarked, "What philosophers call 'linguistic deference' is actually *the use of experts as instruments*; not Marxist division of labor in semantics but capitalist exploitation in epistemology". Fodor (1994) p. 36. As he notes, this doctrine has roots in Smart, who writes "... even a color-blind person can reasonably assert that something is red, though of course he needs to use another human being, not just himself, as his 'color meter'". Smart (1962) p. 172.

³² This concurs with Burge's views. See for instance (1982) p. 102.

hence that XYZ is not water, I do not agree that the external factors which make this so are purely facts about the physical environment. To maintain that the demonstrative *that* stuff refers to H₂O independently of the intentions of the community would be to maintain that our natural kind terms cut nature at its joints. I believe this assumption is incorrect. The way we group items, even that we do so by appeal to internal structural properties, is largely dependent on contingent facts about human psychology and the *de facto* progression of science. What a subject means by a word is determined by the correct use for that word in her community, and its correct use in her community is as a word for *that* stuff: but the demonstrative gets its reference from the intentions of the social community, which are in turn shaped by the relevant contingencies.

1.6 Is water really H₂O?

Putnam has claimed that “[the] extension of ‘water’ in the sense of water_E is the set of all wholes consisting of H₂O molecules, or something like that”³³. Putnam’s approach displays a commitment to an essentialist semantics of a kind found in Saul Kripke³⁴, according to which the reference of a word such as “water” is determined by internal structural properties. Such an essentialist semantics has been challenged by Mark Wilson, Barbara Malt and Noam Chomsky.³⁵ Maintaining that water is H₂O requires an explanation of the following two facts. First, substances which are largely H₂O may be called something else; and second, what we call “water” is in the main not pure H₂O. Examples in the first category would be tears and tea; examples in the second, almost everything we in fact call “water”.

Putnam’s essentialism is less radical than Kripke’s, according to which natural kind terms are non-descriptive, and have no meaning in the Fregean sense. Putnam, in

³³ Putnam (1975b) p. 224.

³⁴ Kripke (1972) and (1980).

³⁵ Wilson (1982), Malt (1994), Chomsky (1995).

addition to his argument that psychological state does not determine extension, proposed a theory for the description of word meaning which would include not only a specification of the extension of the given term (H₂O for “water”), but also “syntactic markers” (“mass noun; concrete”), “semantic markers” (“natural kind; liquid”) and a “stereotype” (“colorless; transparent; tasteless; thirst-quenching; etc.”).³⁶ This does not, however, amount to an explanation of the two relevant facts.

The following explanation has been offered by Barbara Abbott.³⁷ The first thing to note is that substances such as beer and lemonade are largely water *precisely* in virtue of consisting largely of H₂O molecules. To deny that water is H₂O would be to deny this basic fact. What of the claim that such substances are not called “water”? Abbott explains this by appeal to a distinction between what something is called and what it is. This amounts to my claim above that the categories according to which we choose to classify items are in general chosen because of contingent facts about our interests. However, that dogs and cats are classified as “dogs” and “cats” does nothing to undermine the fact that they in fact all fall into the same category, “animal”. Thus, that we do not refer to tears as “water” simply reflects a special interest we have in distinguishing that type of water from other types of water. The issue is not whether we tend not to call tears “water”, but whether we deny that tears are water; that is, whether we say, “not water” of them, which, I take it, we do not.

What of the claim that what we call “water” is in the main not pure H₂O? Abbott explains this by appeal to,

... a very general and natural type of vagueness in our use of linguistic expressions, and therefore one which need not and should not be incorporated separately into our account of the meaning of each word and phrase. For example, when someone asks to borrow our car we are not

³⁶ Putnam (1975b) p. 269.

³⁷ Abbott (1997).

required to remove from the glove compartment the maps, box of Kleenex, registration, etc. - i.e. everything that is not part of the car strictly speaking. It is true that Portia was able to get Shylock on this type of technicality, but had the legal advisor been anyone else, Antonio probably would have been a goner. (Abbott, 1997 p. 317)

While I am sympathetic to this line of reasoning, there is one important fact which has yet to be accounted for. There are substances such as “heavy water”, which, despite the fact that they contain no H₂O molecules, are nevertheless classified as water. It is not clear that Abbott’s explanation can account for this fact. For this we must rely on the claim I made above that the physical environment has its constitutive significance for individual psychology only as mediated by the social environment. This claim can be invoked as an explanation as to why substances such as heavy water are classified as water. The important point is that water is, as Putnam put it, “the set of all wholes consisting of H₂O molecules, *or something like that*”³⁸.

1.7 Conclusion

The physical and social environment do not merely surround the subject, providing a context within which her propositional attitudes can be assessed. Contextual facts inextricably permeate the field of psychological investigation, even when what is under study is the psychological organisation of an individual. (Pettit and McDowell, 1986 p. 14)

This is the claim of semantic externalism. My concern here has not so much been to defend the claim, as to present it. I am, however, in broad sympathy with it; particularly the claim that *social* context is constitutive of content. The thesis is presented as an investigation into certain consequences of the claim. Specifically, my focus rests on the epistemological question of how, assuming the truth of semantic externalism, a subject can be said to know the meanings of her

³⁸ Putnam (1975b) p. 224, emphasis added.

psychological terms in a direct and authoritative manner. If contextual facts “inextricably permeate the field of psychological investigation”, the possibility of direct access to one’s psychological states is brought under pressure. I call those who aim to reconcile semantic externalism with privileged access *compatibilists*, and those who believe them to be irreconcilable *incompatibilists*.³⁹ I maintain that semantic externalism is consistent with privileged access to one’s thoughts, and am therefore a compatibilist.

The structure of the thesis is as follows. The initial charge of incompatibility between semantic externalism and privileged access is set out in chapter two. The arguments discussed there appeal to “travelling cases”, according to which subjects are switched between relevantly differing environments without their knowledge (or consent). The arguments demand an explanation of two salient facts; that is, an answer to two questions. First, how can a subject have privileged access to the contents of her thoughts given that she would be ignorant of a change in her body of concepts, the concepts she possessed, were she to be switched between environments of the differences between which she would be equally unaware? Second, how can a subject have privileged access to the contents of her current thoughts, given that switching her to a new environment would result in her being wrong about the contents of those very same (now past) thoughts? This latter question demands an account of memory consistent with semantic externalism. I answer both questions, and argue that semantic externalism is consistent with privileged knowledge of the meanings of one’s psychological terms. Further, I argue that such access to one’s psychological states is privileged and authoritative, and offer an account of the authority accorded.

In chapter three I consider a subsequent argument, the *Argument from Privileged Access*, intended to discredit the proposed compatibility of semantic externalism and privileged access set out in chapter two. The argument takes the form of a *reductio ad absurdum*, and runs as follows. If

³⁹ I first came across this terminology in Boghossian (1997).

semantic externalism and privileged access were compatible, one would have the means by which to gain knowledge of substantive empirical facts by introspection and conceptual analysis alone. One cannot gain knowledge of empirical facts by introspection and conceptual analysis alone. Therefore, semantic externalism and privileged access are incompatible. I accept and defend the claim that if semantic externalism and privileged access are compatible, one will have the means by which to gain knowledge of substantive empirical facts by introspection and conceptual analysis alone.

In chapter four I argue that, contrary to appearances, this consequence should be embraced as a natural consequence of semantic externalism. That is, I deny the force of the *Argument from Privileged Access*. I maintain that the consequence appears absurd only if one has not fully embraced the force of semantic externalism. To accept semantic externalism is to reject the assumption of methodological solipsism. Since the constitutive nature of the mind is not as we thought, it should be no surprise that introspection isn't either.⁴⁰

In chapter five, I address the issue of *de re* thought. I consider two opposing theories, each of which is defined by its answer to the following question: is the content of a *de re* thought to be understood as object-dependent or object-independent? If the content of a *de re* thought is object-dependent, it will be possible to gain non-empirical knowledge of the existence of the particular object or objects upon which a given *de re* thought depends. On this view, there will be no relevant asymmetry between *de dicto* thoughts and *de re* thoughts. If, on the other hand, the content of a *de re* thought is object-independent, no such non-empirical knowledge will be available. If this view is correct, there *will be* an asymmetry between the inferences which can be drawn from one's supposed *de dicto* thoughts

⁴⁰ Introspection is taken intuitively as being the means, if any, by which a subject has privileged access to the contents of her mind.

and the inferences which can be drawn from one's supposed *de re* thoughts.⁴¹

Finally, in chapter six, I focus directly on an anti-sceptical argument presented by Putnam.⁴² The argument aims to show that from the assumption of a causally-constrained theory of reference, such as semantic externalism, the proposition that I am a brain-in-a-vat is self-refuting and necessarily false. I defend the argument, and relate it to the conclusions reached thus far.

As mentioned above, while I am broadly sympathetic to semantic externalism, the thesis is not intended as a direct endorsement of it. The aim here is to show that the rejection of the assumption of methodological solipsism has wide-reaching implications, in particular for the notion of introspection, and for the alleged distinction between mind and world. If the subject is not the sole determiner of her mind, introspection can no longer be seen as something purely internal to the subject. With these thoughts in mind, let us proceed.

⁴¹ Burge is at pains to emphasise that semantic externalism is a theory concerned with the oblique occurrence of terms in propositional attitude attributions, which carries over into psychological content ascription. He emphasises that it is orthogonal to any issue about *de re* belief, which he treats in a very different manner. For more on this see chapter five below.

⁴² Putnam (1981) chapter one.

Travelling Cases

2.1 Introduction

Descartes' writing provides a classic example of a position which presupposes that at least some propositional mental events can be known by the subject in a direct, authoritative, non-empirical manner. If a subject knows a given propositional mental event in this non-empirical way, I will refer to that subject as having *privileged access* to that thought.⁴³ I take it to be an intuitively compelling presupposition that we do indeed have privileged access to at least some of our thoughts. However, the recent emergence of semantic externalism, according to which a subject's set of possible and actual thoughts is dependent upon, and restricted by, relations that subject bears to her environment, has been taken to threaten this intuitively compelling presupposition.

In this chapter I will discuss two arguments designed to bring out the alleged tension. I have grouped the arguments together because they both appeal to "travelling cases", according to which a hypothetical subject is unknowingly switched between different environments. The environments are assumed to differ in imperceptible ways, ways which nevertheless suffice to cause a change in the concepts that subject possesses and hence in the thoughts it is possible for her to entertain. The differences between environments are *semantically relevant*. It is this type of imperceptible switching which in both arguments brings out the alleged

⁴³ The term "thought" as I use it should not be understood as a Fregean thought. Rather, "thought" should be understood as a synonym for "propositional mental event". Hence, two subjects cannot have the same thought, but can have thoughts with the same content. Similarly, a subject cannot be said to have the same thought at different times, but can have two thoughts with the same content at different times. When a subject knows a propositional mental event, she knows the content of her thought.

incompatibility between semantic externalism and privileged access.

I will refer to the first argument as the *Content Sceptic's Argument*. It demands an answer to the following question: how can a subject have privileged access to the contents of her thoughts given that she would be ignorant of a change in her body of concepts, a change in the concepts she possessed, were she to be switched between environments of the differences between which she would be correspondingly unaware? To answer this question is to provide an account of privileged access consistent with semantic externalism. I argue that on the semantic externalist view, privileged access is correctly understood as an externalist phenomenon, and can not be accounted for in terms of evidence available to the subject. The most natural way to interpret this notion of "evidence available to the subject" is as phenomenological evidence. I argue that however one thinks of phenomenology, that is, whether or not one thinks of it as locally supervenient, it cannot provide evidence for introspective knowledge. I then provide a positive account of privileged access.

I will refer to the second argument as the *Argument from Memory*. It demands an answer to the following question: how can a subject have privileged access to the contents of her current thoughts given that, were she to be imperceptibly switched to a relevantly different environment, she would thereby come to have false beliefs about the contents of those very same (now past) thoughts? An answer to this latter question demands an account of memory consistent with semantic externalism. I agree that semantic externalism opens up a new form of scepticism about the contents of one's past thoughts. However, I argue that this does nothing to jeopardise the privileged access one has to one's current thoughts, even if the content of those thoughts purports to refer to a past event.

By providing an answer to both questions, I will have shown that neither the *Content Sceptic's Argument* nor the *Argument from Memory* establishes that semantic externalism is incompatible with privileged access.

2.2 Knowledge of one's thoughts

I maintain that the following three claims correctly capture the nature of the relation between a subject and *at least some of* her propositional mental events.

1. *Self Knowledge*: a subject can have knowledge of (at least some of) her thoughts.
2. *Privileged Access*: the knowledge a subject has of (at least some of) her thoughts is non-empirical.⁴⁴
3. *First Person Authority*: there is a presumption in favour of a subject's claims to self knowledge.

Whether or not true self-ascriptions should be classified as knowledge will be discussed in section **2.8** below, where I formalise the privileged access thesis, and again in chapter four, where the issue becomes pivotal. For present purposes, since nothing immediate hinges on the issue, I will talk as if true self-ascriptions were instances of knowledge. For the moment, then, the important claims are the second claim, the claim of privileged access, and the third claim, the claim of first person authority.

What of the claim of privileged access? There is an asymmetry between the knowledge a subject can have of her own thoughts, and the knowledge she can have of another's thoughts. Typically, Susan knows what she is thinking in a way in which others do not. To find out what Susan is thinking, others will have to engage in some form of empirical investigation; they will have to watch her behaviour, or listen to her utterances. No such observation or interpretation is necessary if Susan wants to find out the content of at least

⁴⁴ In the characterisation of privileged access, I have chosen to use the term "non-empirical" as it seems the most neutral way to characterise the asymmetry between the way in which a subject can know her own thoughts and the way in which she can know the thoughts of another. By non-empirical knowledge, I simply mean knowledge that can be had without recourse to observation, or external perception.

some of her thoughts. If she has a current belief that time is of the essence, say, she can know that she is thinking that time is of the essence without having to observe what she says or what she does. In fact, it is plausible to maintain that it is precisely *because* she has that particular belief that she says what she does and acts as she does. A subject's knowledge of a her thought is arguably not a conclusion based on an inference from empirical evidence. Wittgenstein expressed the asymmetry between the knowledge a subject can have of her own thoughts, and the knowledge she can have of another's thoughts thus. The criteria for the attribution of a belief to another are her words and actions, whereas for a self-attribution there are no such criteria.⁴⁵ This is the claim of privileged access.

What of the claim of first person authority? When a speaker asserts that she has a belief, a desire, a fear or an intention, there is, typically, a certain presumption that she is correct, a presumption that does not attach to her ascriptions of propositional mental events to others. That there is this presumption is the claim of first person authority. If the claim of first person authority is to be accepted, an explanation of the asymmetry between the authority accorded to attributions of attitudes to our present selves and attributions of the same attitudes to other selves must be forthcoming. What accounts for the authority accorded first person present tense claims, and denied second or third person claims?⁴⁶

Characterisations of first person authority often invoke the claim of privileged access. They point to the fact that self-attributions are not normally made on the basis of evidence or observation, that it does not normally make sense to

⁴⁵ Wittgenstein (1953) §377. While the underlying thought here is essentially the same, criteria and evidence should not be conflated. For the importance of the distinction see for example McDowell (1982).

⁴⁶ It is important to note that first person authority is accorded to a subject's claims about her *present* propositional mental events. The relevance of this restriction is particularly evident in section **2.9** below, where I discuss a form of scepticism concerning a subject's claims about her *past* propositional mental events.

question why a person attributes herself with the propositional mental events she does. However, an explanation of the asymmetry central to first person authority cannot be given purely by pointing to this distinction concerning the existence or absence of an evidential basis for the ascription of propositional mental events. I agree with Donald Davidson when he says that first person authority is not explained by the fact that self-attributions are not based on evidence, since, “claims that are not based on evidence do not in general carry more authority than claims that are based on evidence, nor are they more apt to be correct”⁴⁷. What would be required in addition is an account of why in this particular case, the case of attributions of thoughts to one’s present self, the lack of evidence supported the correctness of the attribution claim; and if this could be given, the initial appeal to the lack of evidence as an explanation of the asymmetry would be rendered redundant.⁴⁸

It is at this point worth reflecting briefly on the writings of Gilbert Ryle⁴⁹. Ryle, famously, denies the claim of privileged access. That is, he maintains that the knowledge a subject has of her own thoughts is of a piece with the knowledge another has of her thoughts. Self-attributions are based on behavioural evidence, evidence which is equally available to others as it is to the subject. Even while Ryle denies the privileged access claim, he nevertheless maintains the claim of first person authority. Any account of first person authority he offers could not, therefore, commit the mistake mentioned above in connection with Davidson. Ryle can make no appeal to the claim that self-attributions evidence-independent in

⁴⁷ Davidson (1984a) p. 103.

⁴⁸ Davidson has provided an account of first person authority which does not invoke the privileged access claim. According to Davidson, if we do not assume a subject knows her thoughts, then we cannot begin the process of radical interpretation. See Davidson (1984a). This explains why the presumption is needed, but it is not clear that this in itself provides a justification for it. For more on the notion of radical interpretation see Davidson (1973), (1974) and (1976).

⁴⁹ See Ryle (1949).

providing an account of first person authority, since it is his contention that self-attributions *are* based on empirical evidence. Ryle's positive account of first person authority appeals rather to the fact that a subject spends more time in her own company than anyone else does, and hence has more evidence – empirical evidence – on which to base predictions about her behaviour.

I am sympathetic to the view that much of our self knowledge is not achieved directly, but relies on evidence similar to that required for knowledge of the thoughts of others. I agree, then, that there are instances of self knowledge which cannot be characterised by the privileged access claim. In addition, I agree that there is much that we do not know, or that we misconstrue, about our own minds. However, Ryle's thesis is too strong. There is a class of propositional mental events, albeit a severely restricted class, to which we do have privileged access, and about which we are authoritative. That there are many exceptions does nothing to jeopardise the truth of this claim.⁵⁰

In section **2.8** I formalise the privileged access claim and offer an account of first person authority. The account emerges from an examination of the two arguments which provide the focus for the present chapter. Before turning to these, however, I would like to make explicit an initial qualification to the account of privileged access.

2.3 Comparative content and bare content

Kevin Falvey and Joseph Owens draw a distinction between two different kinds of privileged access:

KC: An individual knows the contents of his occurrent thoughts and beliefs authoritatively and directly (that is, without relying on inferences from observation of his

⁵⁰ In fact, I think that the class of propositional mental events to which we do have privileged access is more restricted than is often thought. The account of privileged access offered in section **2.7** below reflects the limited sense in which I am willing to endorse the claim of privileged access.

environment). Call this kind of knowledge *introspective knowledge of content*. (Falvey and Owens, 1994 p. 107)

KCC: With respect to any two of his thoughts or beliefs, an individual can know authoritatively and directly (that is, without relying on inferences from his observed environment) whether or not they have the same content. Call this kind of knowledge *introspective knowledge of comparative content*. (Falvey and Owens, 1994 p. 109-10)⁵¹

KC deals with occurrent thoughts only, whereas **KCC** deals with both occurrent and non-occurrent thoughts.

Falvey and Owens argue that the incompatibility between semantic externalism and privileged access obtains only when privileged access is understood in the latter sense – as introspective knowledge of comparative content. They argue that there is no such incompatibility if one has the former meaning in mind. According to Falvey and Owens this is as it should be, since there are reasons independent of semantic externalism for thinking that **KCC** is false.

The tension between knowledge of comparative content and semantic externalism is illustrated by the following scenario. Consider Susan, who, we are to suppose, has been switched back and forth between Earth and Twin Earth without her knowledge. Suppose that once she is linguistically embedded on Earth she is asked whether, when she utters the sentence “water is a liquid”, she expresses the same thought-content she would have expressed by uttering that sentence, syntactically individuated⁵², one year ago. What will she answer?

She will presumably say yes, but if she was on Twin Earth last year, then she will be mistaken. And nothing that is available to her through introspection alone will reveal her mistake to her. She will not learn the truth until she investigates

⁵¹ The terms **KC** and **KCC** are mine.

⁵² To assume that Susan utters a token of the same type of sentence semantically individuated would be to beg the question. Tokens of the same type of sentence semantically individuated would of course express thoughts with the same content.

her environment. If externalism is true, the principle embodied in [**KCC**] is false. (Falvey and Owens, 1994 p. 111-2)

Their example involves a dispute between Benson Mates⁵³ and Alonzo Church⁵⁴ about whether the following two sentences mean the same thing.

- a) Nobody doubts that whoever believes that Mary is a physician believes that Mary is a physician.
- b) Nobody doubts that whoever believes that Mary is a physician believes that Mary is a doctor.

Mates believes they express different thoughts; Church disagrees. And this even though each believes that the thought he expresses when he utters either of these sentences is the thought expressed by the sentence in the public language, English. Still, “[one] of them is wrong, but whoever it is, it is implausible in the extreme to suppose that his error is due to introspective failure.”⁵⁵ The grounds on which the decision is made cannot be purely introspective, Falvey and Owens argue. Rather, the information needed to determine whether the two thoughts are the same will come via an empirical investigation into the external world, the nature of our linguistic practice, the semantic theories that best represent that practice, and so on.

So far as we can tell there is nothing in commonsense psychology to suggest that introspection alone provides all we need to ground judgements of sameness and difference in the contents of our propositional mental states. Even if one knows what one is thinking at a given time, and knows what one is thinking at a later time, it may be necessary to know something about one’s environment in order to know whether these two thought contents are identical. (Falvey and Owens, 1994 p. 113)

⁵³ Mates (1952).

⁵⁴ Church (1954).

⁵⁵ Falvey and Owens (1994) p. 113.

In accord with this analysis, I maintain that the claim of privileged access is correctly understood as a claim about a subject's knowledge of the content of an occurrent thought only. It cannot be extended to cover instances of knowledge of difference or sameness of the content of different thoughts. As will become clear in due course, what is wrong with **KCC** is that it requires privileged access to the content of non-occurrent thoughts. This will be further developed over the next two sections.

Bearing this in mind, let us turn to the first argument for the incompatibility between semantic externalism and privileged access, the *Content Sceptic's Argument*. The argument aims to prove that there is no class of thoughts to which a subject has privileged access.

2.4 The Content Sceptic's Argument

The *Content Sceptic's Argument* can be extrapolated from the following passage from Falvey and Owens.⁵⁶

[I]s it a consequence of [semantic] externalism that when Susan thinks the thought she would express using the words 'water is a liquid', she does not know directly and authoritatively that she is thinking that water is a liquid? It might seem that the answer is yes, by virtue of the following reasoning. In order for Susan to know that she thinks the thought that water is a liquid, she would have to know that her thought involves the concept *water* rather than the concept *twater*. But, by hypothesis, there is nothing in her experiential history that provides her with the conceptual resources necessary to discriminate between these two concepts, and hence she has no introspectively available evidence that her present thought invokes the one concept rather than the other. Therefore, she cannot know by introspection alone that she is thinking that water is a liquid. She will have to examine her

⁵⁶ Henceforth, XYZ will be referred to as "twin water", or simply "twater". Correspondingly, the relevant concept had by people on Twin Earth will be referred to simply as *twater*.

environment in order to know the content of her thought. (Falvey and Owens, 1994 pp. 113-4)

Burge expresses the same sceptical worry as follows.

How can one individuate one's thoughts when one has not, by empirical methods, discriminated the empirical conditions that determine those thoughts from empirical conditions that would determine other thoughts? (Burge, 1988 p. 653)

The argument purports to show that if semantic externalism is true, Susan must undertake some form of empirical investigation in order to know what thought she expresses by the words "water is a liquid". That is, since things would seem the same to Susan on Earth as on Twin Earth, she must conduct some form of empirical investigation into the structure of the liquid before her if she is to know the content of her current thought. This ensures that Susan *does not know* (with privilege or otherwise) the content of her current thought. She is on this occasion denied self knowledge. In addition, the fact that gaining knowledge of the content of her thought would require empirical investigation, is a fact which denies Susan privileged access to the content of that thought.

Since it is plausible to maintain that if the argument is correct there is no thought to which the argument would not apply, we appear to be forced to give up the claim of privileged access. In addition, if self knowledge invariably depends upon an empirical investigation of the environment, someone else may well be in a better position than Susan to know the content of any of her given thoughts, which contradicts the claim of first person authority: there are no grounds on which to base a presumption in favour of a subject's claims to self knowledge.⁵⁷

⁵⁷ I am not here assuming that the denial of privileged access entails the denial of first person authority. As noted in section 2.2 above, Ryle maintains the latter while denying the former. However, there is an important difference between Ryle's thesis and the *Content Sceptic's Argument*. Ryle could consistently maintain the claim of first person

Extrapolating, we get the following form of argument, the *Content Sceptic's Argument*.

(pi) If Susan knows she's thinking that water is a liquid, her evidence is such that it would rule out the possibility that she is thinking that *twater* is a liquid.

(pii) Susan's evidence is not such that it would rule out the possibility that she is thinking that *twater* is a liquid.

(c) Therefore, Susan doesn't know she's thinking that water is a liquid.⁵⁸

If we are to maintain that semantic externalism and privileged access are compatible, we must find fault with this line of reasoning. So what is wrong with the *Content Sceptic's Argument*? The *Content Sceptic's Argument* appeals to two related facts. First, a subject would be unable to distinguish water from various other superficially identical yet structurally distinct substances, such as twin water. Second, a subject would be unable to distinguish the concept *water*, which refers to water, from various other hypothetical concepts, such as *twater*, which would refer to such superficially identical yet structurally distinct substances. These two related facts are by themselves insufficient to demonstrate an incompatibility between semantic externalism and privileged access. The semantic externalist

authority even while denying privileged access, since according to Ryle the empirical facts known by the subject would never be fewer than the empirical facts known by others. This is because the relevant empirical evidence was on the Rylean view behavioural and verbal. In the present case, however, it *is* possible that the subject have less empirical evidence than others for the correct attribution of a thought to herself. With the adoption of semantic externalism, the relevant evidence is no longer just behavioural and verbal; rather, the evidence concerns the underlying structure of substances in the world, structure which cannot be inferred from the superficial qualities of those substances, and the linguistic practices of the community.

⁵⁸ This is similar to the way in which Brueckner expresses the argument in his (1990).

can happily accept them. The ability to distinguish between a water concept and a twater concept would be required for the kind of knowledge captured by **KCC**, knowledge of comparative content, but, as noted above, the fact that one does not have knowledge of comparative content is insufficient to undermine the claim of privileged access. To bring out the alleged incompatibility, the proponent of the *Content Sceptic's Argument* needs to appeal to a further assumption. The requisite assumption is that knowledge that one's thought involves the concept *water* requires knowledge that one's thought does not involve the concept *twater*. That is, the *Content Sceptic's Argument* requires the assumption that knowledge of the content of a thought requires the ruling out of various possible alternative thought-contents. In the next section I will present a challenge to this assumption.

2.5 Relevant alternatives

In the previous section I remarked that the *Content Sceptic's Argument* invokes the following assumption. Introspective knowledge of a given thought requires that the possibility that the subject be entertaining a relevantly alternative thought be ruled out. The underlying assumption here is that knowledge in general requires the ruling out of various *relevant alternatives*.⁵⁹ A relevant alternative is one the mere possibility of which is enough to defeat an actual claim to knowledge. This is a familiar assumption within the context of perceptual knowledge.⁶⁰ Consider, for instance, Alvin

⁵⁹ On the notion of relevant alternatives see for example Dretske (1970), and Goldman (1976).

⁶⁰ I do not here intend that introspective knowledge be assimilated to perceptual knowledge. The analogy is instructive in so far as it brings out a difference which is crucial to the fault in the *Content Sceptic's Argument*. For criticisms of the perceptual model of introspective knowledge see for example Shoemaker (1985) and (1988), Davidson (1987), and Burge (1988) and (1996). For an interesting account of introspective knowledge as analogous to bodily perception see Armstrong (1968) especially pp. 323-38.

Goldman's example of Henry.⁶¹ Henry is driving through the countryside, pointing out barns to his son. Does Henry *know*, on pointing to a particular barn, that it is a barn to which he is pointing? The answer, according to Goldman, will depend upon the existence or absence of relevant alternatives which could serve to discredit Henry's claim to knowledge. If Henry is in an area where the fields are replete with papier-mâché barns, even though Henry is in fact pointing to a real barn, we would not attribute Henry with *knowledge* that it is a barn to which he is pointing, because that he is pointing to a papier-mâché barn is in this situation a relevant alternative.⁶²

Our inclination to deny Henry knowledge that he is pointing to a barn is captured, according to Falvey and Owens, by the following principle of relevant alternatives.

(RA) If (i) q is a relevant alternative to p , and
(ii) S 's belief that p is based on evidence that is compatible with its being the case that q , then S does not know that p . (Falvey and Owens, 1994 p. 116)

The notion of a relevant alternative is explicated as follows: q is a relevant alternative to p if q is a logically possible proposition incompatible with p , and the possibility that q obtains is relevant in the context.⁶³

Take p to be Henry's pointing to a barn. Take q to be Henry's pointing to a papier-mâché barn. Henry's belief that he is pointing to a barn is based solely on the visual appearance of the object to which he is pointing. His evidence for the belief that he is pointing to a barn is therefore compatible with his pointing to a papier-mâché barn. In addition, in the situation described above, q is a relevant

⁶¹ Goldman (1976).

⁶² If, on the other hand, Henry were in an area where there were nothing but real barns, his pointing to a papier-mâché barn would not be a relevant alternative, and his claim to knowledge would not be defeated.

⁶³ Falvey and Owens (1994) p. 116.

alternative to p . Hence, according to **(RA)**, Henry does not know that he is pointing to a barn.⁶⁴

The reason that Henry's claim to knowledge is undermined by his failing to possess evidence ruling out the possibility that the object to which he is pointing is a papier-mâché barn is that, given the high frequency of papier-mâché barns in the area, Henry could easily be deceived into thinking that a papier-mâché barn was a genuine barn. Crucially, if the object *were* a papier-mâché barn, Henry would still believe that it was a genuine barn.⁶⁵ Such reflections lead Falvey and Owens to claim that the plausibility of **(RA)** is grounded in the "more basic principle" **(RA')**.

(RA') If (i) q is a relevant alternative to p , and (ii) S 's justification for his belief that p is such that, if q were true, then S would still believe that p , then S does not know that p . (Falvey and Owens, 1994 p. 116)

It is this principle which they believe "best captures the importance of relevant alternatives in refuting putative claims to knowledge"⁶⁶.

Now think back to Susan. Susan, it will be remembered, believes she is entertaining the thought that water is a liquid. To begin with, let us suppose that Susan occasionally travels back and forth between Earth and Twin Earth. On this supposition, that Susan be entertaining the thought that twater is a liquid becomes a relevant alternative to her entertaining the thought that water is a liquid. The pertinent

⁶⁴ It might be objected that the formulation of **(RA)** depends upon the controversial assumption that perceptual knowledge is evidential, where the evidential base for a perceptual belief is, for instance, a sense-datum. However, I take it that **(RA)** does not embody any such strong thesis. Rather, it serves to pick up on the fact that Henry, for example, has not tried to verify his belief by investigating further; that is, by gathering more evidence.

⁶⁵ For counterfactual theories of knowledge and justification see Dretske (1971) and (1981), Goldman (1976) and (1986), and Nozick (1981).

⁶⁶ Falvey and Owens (1994) pp. 116-7.

question, then, is whether or not her self-ascriptive belief can be counted as an instance of knowledge. What do **(RA)** and **(RA')** tell us about Susan's putative claim to knowledge?

Falvey and Owens assume that Susan's belief that she is entertaining the thought that water is a liquid is based on evidence: evidence that is compatible with her in fact believing that *twater* is a liquid. If their assumption is correct, Susan's belief does not constitute knowledge according to **(RA)**. But, according to Falvey and Owens, this is not to the point, since it is **(RA')**, and not **(RA)**, which "best captures the importance of relevant alternatives in refuting putative claims to knowledge"⁶⁷: and if we consider **(RA')** matters are different.

(RA') has it that for Susan not to know that she is entertaining the thought that water is a liquid, her justification for her self-ascriptive belief must be such that if she were entertaining the thought that **twater** is a liquid, she would still believe that she was entertaining the thought that **water** is a liquid. But reflection on the thesis of semantic externalism reveals that this could never happen. According to semantic externalism, the two environments differ in semantically relevant ways. These differences would therefore be reflected in Susan's thoughts; *all* of Susan's thoughts. Susan's justification for her belief that she is entertaining the thought that water is a liquid is *not* such that, were she on Twin Earth she would still have the belief that she was entertaining the thought that water is a liquid. Indeed, if Susan were on Twin Earth, she *could* not believe she was entertaining the thought that water is a liquid, since she could have no propositional mental events involving the concept *water*. It is simply not true that were she on Twin Earth she would still believe that she was entertaining the thought that water is a liquid.⁶⁸

⁶⁷ *ibid.*

⁶⁸ This assumes that Susan on Twin Earth is embedded in the linguistic community on Twin Earth. As will become clear in due course, this is an important issue.

It is this consequence of semantic externalism to which Falvey and Owens appeal. However, they give no explanation as to why it should be so. Spelling this out will be my task in sections **2.7** and **2.8**. For the moment, it will suffice to note that, according to Falvey and Owens, the concepts available for use in self-ascriptions will be the very same concepts available for use in first-order thoughts. Crucially, the concepts which figure in higher-order thoughts will vary in tandem with the concepts which figure in first-order thoughts. If this is correct, **(RA')** does not refute Susan's putative claim to introspective knowledge. Hence the *Content Sceptic's Argument* does not prove that semantic externalism and privileged access are incompatible.

I agree with Falvey and Owens that "[i]n such a case, the inability of the subject to eliminate a relevant alternative does not entail that the subject is liable to error in her beliefs"⁶⁹. In addition, I agree that this is because the appeal to relevant alternatives made by the *Content Sceptic's Argument* is indeed correctly captured by **(RA')**. However, Falvey and Owens provide no good reason for choosing **(RA')** over **(RA)**, and their claim that the appeal to relevant alternatives made by the *Content Sceptic's Argument* is captured by **(RA')** and not **(RA)** has in fact been challenged.⁷⁰ In the next section, I look at this challenge and respond by showing that the *Content Sceptic's Argument* can be refuted even if the appeal to relevant alternatives which it makes is correctly captured by **(RA)**. The argument I present for the rejection of **(RA)** turns on the fact that introspective knowledge, if it is to be compatible with semantic externalism, cannot be based on phenomenological evidence.

2.6 **(RA)** or **(RA')**?

The *Content Sceptic's Argument* relies on the notion of a relevant alternative in order to refute a subject's putative claims to introspective knowledge. As discussed in section **2.5**

⁶⁹ Falvey and Owens (1994) p. 117.

⁷⁰ Tye and McLaughlin (1997).

above, Falvey and Owens argue against the *Content Sceptic's Argument* by maintaining that its appeal to relevant alternatives is correctly captured by **(RA')**, and that **(RA')** does not refute such claims to introspective knowledge. It is crucial to their rejection of the *Content Sceptic's Argument* that the appeal to relevant alternatives is correctly captured by **(RA')**, and not by **(RA)**.

Michael Tye and Brian McLaughlin argue that the *Content Sceptic's Argument* for the incompatibility between semantic externalism and privileged access need make no appeal to **(RA')**.⁷¹ Tye and McLaughlin maintain rather that the argument relies directly on **(RA)**. Tye and McLaughlin's attack on the *Content Sceptic's Argument* therefore comes via a direct attack on **(RA)**. The problem they identify with **(RA)** is that in cases of self-ascriptive beliefs the antecedent is never satisfied. It is never satisfied, they argue, because self-ascriptive beliefs are not based on evidence of any sort, and *a fortiori* not based on evidence that fails to rule out relevant alternatives. This analysis agrees with my formulation of the privileged access claim given in section 2.2, and to this extent I am in accord.

Here is the argument provided by Tye and McLaughlin for the claim that self-ascriptive beliefs are not based on evidence; and *a fortiori* not based on evidence that fails to rule out relevant alternatives.

The *Content Sceptic's Argument*, according to Tye and McLaughlin, appeals to the following three theses.

1 *The Introspective Evidence Thesis* Whenever one is occurrently thinking that *p*, the fact that one is occurrently thinking that *p* fails weakly to supervene on the introspective evidence available to one. (Tye and McLaughlin, 1997 p. 9)

2 *The Privileged Access Thesis* It is metaphysically necessary that if we are able to exercise our normal capacity for introspection to form beliefs about our

⁷¹ Tye and McLaughlin are concerned with travelling arguments in general, and their reasoning therefore applies to the *Content Sceptic's Argument* in particular.

occurrent thoughts, then if we are able occurrently to think that p , then we are able to know that we are thinking that p without our knowledge being justificatorily based on empirical investigation of our environment. (Tye and McLaughlin, 1997 p. 9)

3 *The Introspective Evidential Knowledge Thesis* When we know what we are currently thinking without our knowledge being justificatorily based on empirical investigation of our environment, our knowledge is based upon introspective evidence available to us. (Tye and McLaughlin, 1997 p. 9)

These three theses are incompatible, so the argument goes, and hence at least one of them must be given up. This is how the theses give rise to the alleged incompatibility. A subject cannot know that she believes that p if her putative knowledge is based on introspective evidence, since the introspective evidence available to her underdetermines the fact she claims to know. One of the theses must be rejected. Tye and McLaughlin's suggestion is to abandon the third thesis, the thesis that a subject's self knowledge is gained by introspective evidence. For reasons I will explain at the end of this section, I agree that this is the thesis which the semantic externalist should reject. Privileged access is not based on evidence.

First, I would like to examine a further claim made by Tye and McLaughlin. They claim that the challenge posed by travelling cases is not a problem solely for the semantic externalist. Having made a positive case for privileged access not being evidence-based, they try to turn the argument around against the semantic *internalist*, by accusing the semantic internalist of *endorsing* all three theses, and therefore being similarly forced to reject at least one of them. They write, "unfortunately, the advocate of travelling arguments who embraces the privileged access thesis also seems committed to the other two theses. Insofar as there is any incompatibility, it lies here and *not* between [semantic] externalism about thought-content and privileged access."⁷²

⁷² Tye and McLaughlin (1997) p. 11.

This claim, central to their line of argument, is at best distracting, and at worst false. They write,

Philosophers who attack [semantic] externalism via travelling cases assume that a traveller's introspective evidence can be exactly the same when the traveller is thinking that water is a liquid, while on earth, and when thinking that twater is a liquid, while on Twin Earth. They are thus committed to denying a certain *weak* supervenience thesis: for any world *w*, and any individuals *x* and *y*, then *x* is occurrently thinking that *P* if and only if *y* is. Hence they must accept: [1] The Introspective Evidence Thesis. (Tye and McLaughlin, 1997 p. 9)

But no semantic internalist should accept such a thesis, since no semantic internalist is committed to accepting that two physical duplicates could have different thoughts simply in virtue of inhabiting different physical environments.⁷³ As set out in chapter one, semantic internalism is a thesis fundamentally opposed to the Introspective Evidence Thesis. The force of the *Content Sceptic's Argument* lies in pointing out that the semantic externalist owes us a positive account of privileged access. Tye and McLaughlin are wrong to maintain that the content sceptic *appeals* to the three theses; rather she *demonstrates* that they are incompatible, thereby forcing the semantic externalist to give a positive account of privileged access. Tye and McLaughlin's statement of the alleged incompatibility between the theses amounts to no more than a reformulation of the *Content Sceptic's Argument*. To point out the alleged incompatibility is not to solve the problem. Providing a positive account of privileged access is my concern in sections **2.7** and **2.8** below.

⁷³ It is open to the internalist to maintain that the beliefs *are* different *when widely construed*. However, she would have to maintain that the beliefs were the same *when narrowly construed*, and this is the construal which is of significance for the purposes of psychology. There is a plausible internalist position which would not accept the wide / narrow distinction as alluded to here. See for example Segal (1989a) According to this position, the beliefs are the same *simpliciter*. See also Crane (1991).

This is how I interpret the argument. **1**, *The Introspective Evidence Thesis*, is a statement of semantic externalism.⁷⁴ **3**, *The Introspective Evidential Knowledge Thesis*, is a partial account of privileged access consistent with semantic internalism. The argument purports to show that if you accept semantic externalism, this account of privileged access must be rejected. In rejecting **3**, the semantic externalist must provide an alternative account of privileged access. Given that **1** and **3** are themselves incompatible, one of them must be rejected. However, the issue of which to reject is precisely not, as Tye and McLaughlin contend, “orthogonal to the issue of whether content-externalism is correct”⁷⁵. Rather, this lies at the heart of the debate.

To summarise, whether the *Content Sceptic’s Argument* appeals to the notion of a relevant alternative which is correctly captured by **(RA)** or to the notion of a relevant alternative which is correctly captured by **(RA’)**, the argument does not rule out putative claims to introspective knowledge. Hence the argument fails to establish that there is an incompatibility between semantic externalism and privileged access.

What the argument does, however, is show that there are certain constraints on the kind of knowledge delivered by introspection.⁷⁶ This brings us back to the claim I made earlier, that privileged access is not based on evidence. My aim here is not to provide a full account of the nature of introspection. My aim is, rather, the modest one of pointing out what introspection cannot be if semantic externalism is true.

⁷⁴ In fact, **3** is a statement of one form of semantic externalism. It is consistent with semantic externalism that **3** be false if introspective evidence is itself to be widely construed. This point is discussed further below.

⁷⁵ Tye and McLaughlin (1997) p. 11.

⁷⁶ I will throughout be working with an intuitive characterisation of introspection according to which it is the means by which a subject has privileged access to her thoughts. It follows from this characterisation that if a subject has privileged access to a thought, she knows that thought by introspection. Her knowledge of that thought is introspective knowledge.

Both in the passage quoted from Falvey and Owens, and in the three theses presented by Tye and McLaughlin, crucial reference is made to the notion of “introspectively available evidence”. It is the absence of this “introspectively available evidence” which is taken to support the inference to the conclusion that introspection alone is insufficient to ground self knowledge. In short, so the argument goes, because there is no introspectively available evidence which could distinguish twin thoughts, say a water thought from a twater thought, introspection does not yield self knowledge.⁷⁷

How, then, is this notion of “introspectively available evidence” to be understood? I see only one interpretation which is feasible in this context, an interpretation according to which the evidence for a subject’s introspective knowledge is taken to be *phenomenological* evidence. Phenomenology can be interpreted in one of two ways. Either phenomenology supervenes locally on the subject or it does not. To assume that it does is to assume that phenomenology and thought-content can be prised apart; to assume that it does not is to assume that phenomenology and thought-content can not be prised apart.⁷⁸

Given the former interpretation, phenomenology does indeed underdetermine thought-content: things would feel the same to Susan in the actual situation, were she entertaining the thought that water is a liquid, as they would do in the counterfactual situation, were she entertaining the thought that twater is a liquid.⁷⁹ But what this shows is that if

⁷⁷ I do not mean to imply that Falvey and Owens are committed to this inference. On the contrary, they set up their version of the *Content Sceptic’s Argument* in order to show that it does not work.

⁷⁸ For an account according to which all content is conceptual and hence to be individuated externally see for example McDowell (1994).

⁷⁹ To make the example more vivid, consider the case in which Susan switches between two hypothetical environments which differ in semantically relevant ways. It must of course be assumed that Susan is switched at a slow enough rate to acquire the concept appropriate to each, and thereby to have thoughts with differing contents when situated in the different environments. Semantic externalism admits that although

thought-content is wide, and phenomenology is narrow, introspective knowledge is not, as proponents of travelling case arguments suppose, based on how things feel to the subject. It certainly does not show that there can be no introspective knowledge.

If phenomenology is narrow, then, whatever account of privileged access is to be given by the semantic externalist, it must not be one which relies upon phenomenological differences to individuate thought-contents. Phenomenological differences would individuate propositional mental events too coarsely, and would therefore be inadequate for the purposes of individuation. The correct conclusion to be drawn from this is that introspection is not to be understood as based on phenomenological evidence. Recognition of this fact should lead one to the following conclusion. That a subject entertaining the thought that water is a liquid is in a state phenomenologically indistinguishable from the state she would be in were she to entertain the thought that twater is a liquid has no bearing on the question of whether that subject can know the content of her thought by introspection alone.⁸⁰

Given the latter interpretation, phenomenology can no longer be said to underdetermine thought-content. However,

Susan would have different thoughts in each of the different environments, she would be unable to tell when and where the changes occurred. Phenomenology, if it is narrow, would be of no use as a means to distinguish between such differing intentional states. Note that a case of quick-switching would not be a case in which Susan's thoughts altered, and *a fortiori* would not be a case in which Susan's thoughts altered while the phenomenology remained the same.

⁸⁰ If one were sympathetic to the view that phenomenological facts determine thought-content, the view could be modified as follows. Phenomenological facts will not, if semantic externalism is true, uniquely determine the thought a subject is entertaining *across* possible worlds. Which thought the subject is in fact entertaining will depend upon phenomenological facts plus facts about her environment. However, in any given world, phenomenological facts will uniquely determine the thought a subject is entertaining in that world.

even if phenomenology and thought-content are so closely linked that there would be no phenomenological difference without a difference in thought-content, the phenomenology can not be appealed to as the evidence on which a subject's introspective knowledge is based. This is because if phenomenology and thought-content go hand in hand, there is no possibility that the evidence exist while what it is evidence for not exist: and this is contrary to the notion of evidence.

Either way, the conclusion to be drawn from the *Content Sceptic's Argument* is that introspective knowledge is not based on phenomenological evidence.⁸¹ The upshot is that semantic externalism is in need of a positive account of privileged access. It is to this task which I now turn.⁸²

⁸¹ It is in any case curious that proponents of travelling case arguments have appealed to the fact that a subject would be unable to tell the difference between her thought-contents in the actual and the counterfactual situations, since it is by no means obvious that there is *any* particular way it feels to have a certain thought-content.

⁸² Davidson has also provided an account of how semantic externalism can accommodate the claim of privileged access. See Davidson (1987). I will not discuss his account here in any detail, since I think his response to the problem posed for an account of privileged access by semantic externalism belies a failure to understand the nature of the problem. Davidson maintains that the appearance of an incompatibility arises from the following mistaken inference. The fact that a propositional mental event is *described* by relating it to something outside the head cannot be used as the basis from which to infer that the thought must *itself* be outside the head, and hence unavailable to privileged access. There are two points of contention I have with this response. First, it is far from clear who, if anyone, makes this mistake. Second, it is implausible to suppose that a subject can be attributed with privileged access to a propositional mental event of hers, even while she is unable to provide any relevant description of that propositional mental event.

2.7 The privileged access claim

In section 2.5 it was argued that **(RA')** did not refute putative claims to introspective knowledge. The reason given was that it is impossible to set up a relevant alternative in which the subject retains her self-ascriptive belief, yet in which she is deceived about the object of that belief. That is, Susan could not be on Twin Earth, entertaining the thought that **twater** is a liquid, and yet still have the belief that she is entertaining the thought that **water** is a liquid. For every difference in thought-content, due purely to a semantically relevant difference in environmental factors, there would be a corresponding difference in the content of the associated second-order self-ascriptive thought. This provides the basis for an account of privileged access consistent with semantic externalism. The account is widely accepted amongst semantic externalists. Thus, Crispin Wright maintains that “the content of my second-order beliefs will ... be externally determined ... [and] will, as it were, co-vary with externally determined variation in the content of my first-order attitudes.”⁸³ Sydney Shoemaker claims that “the contents of mental states are fixed holistically ... whatever fixes the content of the first-order belief ... also fixes in the same way the embedded content in the second-order belief.”⁸⁴ Similarly, John Heil claims that the contents of second-order thoughts are determined “just as are the contents of first-order thoughts, by the obtaining of appropriate conditions.”⁸⁵

Burge has offered an account which appeals to similar facts. It will prove instructive to examine Burge’s account further, in particular its appeal to two notions, those of containment and of self-reference. Burge writes:

The content of the first-order (contained) thought [that water is a liquid] is fixed by non-individualistic background conditions. And by its reflexive, self-referential character, the content of

⁸³ Wright (1991) p. 76.

⁸⁴ Shoemaker (1994) p. 260.

⁸⁵ Heil (1988) p. 251.

the second-order judgement is logically locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter. (Burge, 1988 pp. 659-60)

That the first-order thought is taken to be contained in the second-order thought could be regarded as controversial. For instance, such an analysis would go against one interpretation of Frege, according to which instances of a term contained in a first-order thought and in a second-order thought would differ in their sense and in their reference.⁸⁶ The term “water” in “Susan believes that water is a liquid” would be taken to refer to the sense of the term “water” in “water is a liquid”. According to Frege, terms in opaque contexts cannot have their normal reference, as this would violate his principle that when the reference of the parts remains the same, so does the reference of the whole. A sentence is here understood as referring to its truth-value. A further, more powerful concern about the containment principle is this. Maintaining that a thought can literally *contain* another thought commits one to certain metaphysical views about propositional mental events which one might otherwise be unwilling to accept. My formalisation of the privileged access claim will therefore not be couched in terms of containment. I will talk of an *embedded content*, but not of an embedded propositional mental event.⁸⁷

The part of Burge’s account which I do wish to endorse is the notion of self-reference. According to Burge, introspective knowledge is self-referential in the sense that the object of reference just is the thought being thought.

When one knows that one is thinking that p , one is not taking one’s thought ... that p merely as an object. One is thinking that p in the very event of thinking knowledgeably that one is thinking it. It is thought and thought about in the same mental act. (Burge, 1988 pp. 659-60)

⁸⁶ For an interpretation of Frege contrary to this see Dummett (1973).

⁸⁷ The thought that grass is green or grass is blue contains the content that grass is blue, but not the thought that grass is blue.

This is a stronger claim than is made, for example, by Falvey and Owens, or by Wright. Burge's claim is not merely that second-order beliefs will co-vary with the environmental conditions which are taken to explain the covariation in first-order thoughts. Rather, the second-order thought takes itself as its object. Thus, when a subject ascribes a thought to herself, she does not have a second-order thought which is directed towards a distinct first-order thought. Neither is there a second-order "empty shell" which could take any one of a number of thoughts as its object.

The attribution of a propositional mental event to oneself is taken to be a matter of thinking a particular thought, an ability grounded in the environment, self-ascriptively. Burge maintains that "one knows one's thought to be what it is simply by thinking it while exercising second-order, self-ascriptive powers"⁸⁸. This is what provides Burge with the means to refute the *Content Sceptic's Argument*.

If background conditions are different enough so that there is another object of reference in one's self-referential thinking, they are also different enough so that there is another thought. (Burge, 1988 p. 659)

It is the self-referential nature of putative claims to introspective knowledge, and not the containment principle *per se* which characterises introspective knowledge.

2.8 A formalisation of the privileged access claim

I propose the following formalisation of the claim of privileged access.

(PA) For all x , if x believes that she thinks that p , then x thinks that p .

Let "thinks" satisfy the condition that any attitude of the form "S ϕ 's that p " (e.g. "S desires that p ", "S fears that p ") entails

⁸⁸ Burge (1988) p. 656.

“S thinks that p ”. This allows for the possibility that a subject could be mistaken about the relation in which she stands to the proposition concerned. For example, it allows that a subject can believe she *desires* that p when in fact she *fears* that p . The privileged access claim captured by **(PA)** is the modest claim that one cannot be mistaken about the *content* of the proposition.⁸⁹

The second point to notice about **(PA)** is that the truth of the first-order thought is irrelevant to the truth of the second-order belief. A subject who believes that she thinks that water is a liquid, according to **(PA)**, thereby thinks that water is a liquid, independently of whether or not water is in fact a liquid. Thus, a subject who believes she thinks that water is a metal does in fact think, on this account, that water is a metal. “S thinks that p ” does not even entail “S believes that p ”, let alone p . The truth-value of the embedded content is irrelevant to the truth-value of the embedding belief. The self-ascription is correct in both instances.

(PA) is a principle about the self-verifying nature of the *content* given in putative claims to introspective knowledge, and not about either the attitude taken towards that content, or the truth of that content. As expressed by **(PA)**, necessarily, a judgement expressible by a sentence of the form, “I think that p ”, will be true.

We are now in a position to provide an account of first person authority. The account of first person authority is based on the claim of privileged access. However, the authority a subject has over the content of her current thought derives not from the lack of an evidential base for her judgement, but from the fact that such a judgement cannot but be true.

The claim that such a judgement cannot but be true can be brought out by stating **(PA)** contra-positively.

⁸⁹ Note that it follows from this that no contradiction is involved in S’s thinking that p but believing that $\neg p$. For an argument to the effect that privileged access should be construed as privileged access to the relation in which one stands to the proposition expressed by one’s thought see McDonald (1995).

(PA) For all x , if x believes she thinks that p , then x thinks that p ; i.e. for all x , if $\neg(x$ thinks that $p)$, then $\neg(x$ believes that she thinks that $p)$.

(PA) accords with the response to the travelling case arguments as presented by Falvey and Owens: that is, it brings out the reason why **(RA')** does not refute putative claims to introspective knowledge. According to **(PA)**, it would be impossible for a subject to believe that she thinks that p , in circumstances in which she was unable to think that p . For example, it would be impossible for Susan to believe that she is entertaining the thought that water is a liquid without thereby doing so. The existence of a given first-order content is conditionally guaranteed by the existence of, and therefore guarantees the truth of, the corresponding second-order belief. In short, putative claims to introspective knowledge are *self-verifying*.

2.9 Unconscious mental states and past mental events

There are three basic types of self-ascriptive belief: those which ascribe occurrent propositional mental events (I am currently entertaining the thought that p); those which ascribe unconscious propositional mental states (I have an unconscious mental state that p); and those which ascribe past propositional mental events (I once thought that p).⁹⁰ In section 2.8 above, I argued that self-ascriptive beliefs of the first sort are self-verifying. As yet there has been no mention of putative claims to knowledge of unconscious propositional mental states or past propositional mental events. In this section I will argue that a principle of privileged access akin to **(PA)** can be formalised for both types of mental act. These formalisations show the very limited sense in which a subject can be said to have privileged access to such mental acts, and will serve to clarify the force of **(PA)**.

⁹⁰ I will talk of unconscious mental *states* rather than unconscious propositional mental events, or unconscious thoughts, since according to my definition of a thought, there could be no unconscious thoughts.

It might be thought that the natural way to formalise a principle of privileged access with regard to one's unconscious mental states would be as follows.

(PAU) For all x , if x believes that she unconsciously thinks that p , then x unconsciously thinks that p .

But **(PAU)** is clearly false. It is undoubtedly possible for a subject to be mistaken in her beliefs about her unconscious mental states. The access a subject has to an unconscious mental state will generally be based on evidence equally available to others as to the subject. In this respect, a subject's knowledge of that state can not be regarded as privileged. The correlate of **(PA)** for unconscious mental states cannot be **(PAU)**.

I remarked in section 2.8 above that **(PA)** is a principle about the self-verifying nature of the *content* given in putative claims to introspective knowledge, and not about the *attitude* taken towards that content. This suggests an enlightening parallel. Suppose Susan were to believe that she had an unconscious hatred of dogs. Susan may be mistaken in the belief that she has an *unconscious* hatred of dogs, just as she may be mistaken in her belief that she *desires* to hate dogs when she in fact fears that she hates dogs. What Susan cannot be mistaken about is that she is entertaining a thought with the *content* that she hates dogs. The very act of believing that she has an unconscious mental state that she hates dogs determines that she has a conscious propositional mental event with that content.

This suggests the following principle.

(PAU') For all x , if x believes that she unconsciously thinks that p , then x thinks that p .

(PAU') is the correct correlate of **(PA)** for unconscious mental states. Putative claims to knowledge of such states are self-verifying in the limited sense shown.

What about thoughts about past propositional mental events? It might be thought that the natural way to formalise

a principle of privileged access with regard to one's past propositional mental events would be as follows.

(PAP) For all x , if x believes that she once thought that p , then x once thought that p .

This is in line with **(PAU)** above, and is similarly false. This is because semantic externalism opens up the possibility of a change in semantic reference over time which the subject cannot detect. Suppose that at time t_1 Susan has the thought that her cat drowned in water. Suppose that Susan is subsequently transported to Twin Earth without her knowledge. At some much later time, t_2 , by which time Susan is semantically efficient in her "new" environment, Susan thinks to herself, wishing to refer to the thought she had at t_1 , "I once thought that my cat drowned in water". This second order belief is false, according to semantic externalism, since the concept Susan expresses at t_2 by the term "water" is not the same concept as the one she expressed at t_1 by the term "water". The second-order belief therefore fails to capture the content of Susan's original thought.

Once again, I remarked in section **2.8** above that **(PA)** is a principle about the self-verifying nature of the *content* given in putative claims to introspective knowledge, and not about the *truth* of that content. And once again, as in the case of unconscious mental states, this suggests an enlightening parallel. A subject may be mistaken about whether or not her judgement correctly captures the content of one of her past thoughts, but she cannot be mistaken about whether or not she is entertaining a thought with that content.

This suggests the following principle.

(PAP') For all x , if x believes that she once thought that p , then x thinks that p .

(PAP') is the correct correlate of **(PA)** for past propositional mental events. Putative claims to knowledge of such thoughts are self-verifying in the limited sense shown.

In conclusion, that **(PA)** is a principle about the self-verifying nature of the *content* given in putative claims to introspective knowledge, and not about the attitude taken towards that content, nor about the truth of that content, allows there to be a limited sense in which claims to knowledge of unconscious propositional mental states and past propositional mental events are self-verifying. This is reflected in **(PAU')** and **(PAP')**. These principles serve to highlight the limited claim of **(PA)**. It is a formal principle, and does not provide an account of the sort of interesting self knowledge that makes a subject a special object of study for herself.

2.10 The Argument from Memory

It is perhaps the realisation that the contents of past propositional mental events may not be captured by thoughts of the form “I once thought that *p*” which lies behind the following line of argument. Paul Boghossian⁹¹ has argued that semantic externalism has the implausible consequence that a subject could have introspective knowledge of a propositional mental event at time *t*, forget nothing, and yet be unable to say at some later time what the content of her thought at *t* was. He writes,

The only explanation, I venture to suggest, for why [Susan] will not know tomorrow what [she] is said to know today, is not that [she] has forgotten, but that [she] never knew. ... What other reason is there for why our slowly transported thinker will not know tomorrow what [she] is said to know directly and authoritatively today? (Boghossian, 1989 p. 23)

Certainly, when Susan, having been transported to Twin Earth, sincerely utters the sentence “I once thought that my cat drowned in water”, she thereby expresses the belief that she once thought her cat drowned in *twater*, since the content of a memory is fixed by environmental conditions *at*

⁹¹ Boghossian (1989).

the time of recollection.⁹² Boghossian is correct, then, in maintaining that Susan does not know at time t_2 what she knew at time t_1 . The switching example has the consequence that the content of a subject's memories can change without the subject forgetting anything or learning anything, and without her being aware of any such change.⁹³ But in the light of the principle of privileged access, an alternative conclusion to that drawn by Boghossian *must* be drawn, since according to **(PA)**, Susan has privileged access to her thought at t_1 , and according to **(PAP')**, she has privileged access to her thought at t_2 . That is, Susan does know the content of her thought both at t_1 and at t_2 .

So what is wrong with Boghossian's line of reasoning? Boghossian assumes that there are two possible explanations for the fact that Susan will not know tomorrow what she is said to know today: either that Susan will forget the content of her original thought, or that her putative claim to knowledge is illegitimate. Neither possibility is feasible. The correct conclusion to be drawn from the *Argument from Memory*, then, is the following. Semantic externalism opens up a third possible explanation for why a subject may not know tomorrow what she is said to know today. The third possible explanation is quite simply this: that she has undergone a semantically relevant environmental switching.

However unsavoury this consequence may be, I think it must simply be accepted as a natural consequence of semantic externalism.⁹⁴

⁹² For an argument along these lines see Ludlow (1995b).

⁹³ Note that even though a subject may be unable to report the content of her past thought directly, she can nevertheless do so *indirectly*. Consider Susan's two beliefs about her cat. Although Susan's belief that she once thought that her cat drowned in water fails to capture the content of her original belief, she can capture the content of her original belief by referring to it as "that thought which I would then have expressed by the words 'my cat drowned in water'". However, the original scepticism remains.

⁹⁴ Wright has maintained that the mere possibility of a change in semantic reference over time should not undermine our general practice

2.11 The temptation of the arguments

So what is the appeal of the *Content Sceptic's Argument* and of the *Argument from Memory*? At first sight, individualism seems to follow from the possibility of privileged access to the contents of one's thoughts.

A person need not investigate the environment to know what his thoughts are. A person does have to investigate the environment to know what the environment is like. Does this not indicate that the mental events are what they are independently of the environment? (Burge, 1988 p. 650)

The question, then, is “why ... having non-empirical knowledge of our thoughts [is] not impugned by the fact that such thoughts are individuated through relations to an environment that we know only empirically”⁹⁵.

Burge provides the following explanation. That a subject can be immediately aware of the contents of her propositional mental events is logically independent of the individuation conditions of those mental events. Privileged access to one's thoughts, and the individuation of those thoughts, are therefore to be treated as distinct matters.

To know that water exists, or that what one is touching is water, one cannot circumvent empirical procedures. But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought. Such conditions need only be presupposed. (Burge, 1988 p. 654)

of assuming first person authority. On this view, our general practice takes for granted that semantically relevant factors of our environment are not prone to change in reference-altering ways. See Wright (1991) pp. 79-80. However, if semantic reference is socially determined, such switching is presumably commonplace. For an argument to this effect see Ludlow (1995a).

⁹⁵ Burge (1988) pp. 651-2.

The ability to think particular thoughts presupposes certain environmental conditions, and this applies *equally* to second-order thoughts as to first-order thoughts. Introspective knowledge requires *no more than* the conditions presupposed in first-order thought, together with an ability of the subject to self-ascribe propositional mental events. **(PA)** respects this distinction between the ability of a subject to think that *p* and knowledge of the conditions required for a subject to have that ability.

To illustrate the point, Burge draws an analogy between introspective knowledge and perceptual knowledge.⁹⁶ Just as it is a mistake to regard perceptual knowledge of physical entities as resting on some prior justified belief that certain enabling conditions are satisfied, so it is taken to be a mistake to regard introspective knowledge as requiring knowledge of the conditions which make such a judgement possible.

Knowing one's thoughts no more requires separate investigation of the conditions that make the judgment possible than knowing what one perceives. (Burge, 1988 p. 656)

Consequently, in the case of introspective knowledge it is not necessary to rule out counterfeit, or counterfactual,

⁹⁶ While the parallel between perceptual knowledge and introspective knowledge is helpful, fundamental differences are evident. Burge acknowledges such distinctions, and himself repudiates any interpretation of introspective knowledge according to the observational model. According to Burge, perceptual knowledge has the following two characteristics. First, it is possible that any physical object perceived could have been different even while the perceptual states, and other mental states of the subject remained the same. There is, then, no necessary relation between the subject and the physical object. Second, perceptual knowledge, being a form of empirical knowledge, is impersonal. In other words, a different observer could have been equally well-placed to make the same observation. Neither of these characteristics are true of introspective knowledge as explicated by Burge. Introspective knowledge according to Burge is essentially personal and non-contingently related to the object of that knowledge. See Burge (1996).

situations. No comparison need be made between the thought one in fact has, and various other possible thoughts one might have had were one related in the appropriate way to an environment which differed in semantically relevant ways. Given the self-verifying nature of putative claims to introspective knowledge, the possibility of a counterfeit situation does not even arise. It makes no sense to claim that external conditions for *this* thought could be different.

The belief that there is an incompatibility arises, according to Burge, from the assumption of two distinct perspectives. From a first-person perspective a subject thinks that she is thinking that water is a liquid; but from a third-person perspective we know that her thinking that thought depends upon a contingent fact about her environment of which she is ignorant. This gives rise to the worry that the original first-person judgement is unjustified unless either it can “encompass the third-person perspective, or unless the third-person perspective on empirical matters is irrelevant to the character of the first-person judgement”⁹⁷. For the worry to be accommodated in the first way, the first-person judgement must be regarded as containing the knowledge available to the third person. That is, the first-person perspective must somehow rule out alternatives. It is this line of thought which gives rise to arguments such as the *Content Sceptic’s Argument*. To accommodate the worry in the second way, however, would be to give up on semantic externalism.

2.12 Conclusion

I have argued that neither the *Content Sceptic’s Argument* nor the *Argument from Memory* show that semantic externalism is incompatible with privileged access. **(PA)** captures the respect in which a subject has privileged access to her thoughts, since it illustrates how introspective knowledge remains constant under possible unnoticeable variations in environmental circumstances. The self-verifying

⁹⁷ Burge (1988) p. 661.

nature of claims to introspective knowledge in turn accounts for the claim of first person authority.

The Argument from Privileged Access

3.1 Introduction

In chapter two I examined two arguments designed to show that semantic externalism could not accommodate the claim of privileged access. A positive account of privileged access was proposed, formalised as **(PA)**, which demonstrates that the claim of privileged access is indeed compatible with semantic externalism. **(PA)** shows why both the *Content Sceptic's Argument* and the *Argument from Memory* ultimately fail.

Having answered the challenge posed by so-called travelling cases, however, the semantic externalist is now faced with a second, derivative challenge. It has been claimed that the conjunction of privileged access and semantic externalism entails that a subject could have non-empirical⁹⁸ knowledge of contingent empirical facts about her environment; the empirical facts upon which her thoughts essentially depend. This is assumed to be absurd, and hence taken to constitute a *reductio ad absurdum* of that very conjunction. I will refer to this argument against the compatibility of semantic externalism and privileged access as the *Argument from Privileged Access*.⁹⁹

Giving a precise formulation of the *Argument from Privileged Access* has proved to be no easy task. The purpose of the present chapter is to arrive at a plausible formulation, and hence to show that the semantic externalist is faced with a genuine challenge.

⁹⁸ Once again, "non-empirical" is here taken to mean without recourse to observation, or external perception.

⁹⁹ The *Argument from Privileged Access* is generally taken as an argument against semantic externalism. See in particular McKinsey (1991) and (1994), and Brown (1995). One alternative would be to deny the claim of privileged access. However, it is hard to see this as a possibility in the light of **(PA)**.

3.2 The McKinsey recipe

The problem in its current form was first raised by McKinsey¹⁰⁰, who expresses it as an *ad hominem* charge of inconsistency against Burge. However, I will treat McKinsey's argument as a general charge against the semantic externalist who wishes to endorse the claim of privileged access. According to McKinsey, then, the semantic externalist who wishes to endorse the claim of privileged access is committed to the following three propositions, propositions which McKinsey maintains are inconsistent.

- (1) Oscar knows [non-empirically]¹⁰¹ that he is thinking that water is wet.
- (2) The proposition that Oscar is thinking that water is wet necessarily depends upon E.
- (3) The proposition E cannot be known [non-empirically], but only by empirical investigation.
(McKinsey, 1991 p. 12)

E is taken to be a proposition expressing the contingent external facts which determine the content of the subject's thought; that is, facts about the subject's physical

¹⁰⁰ McKinsey (1991).

¹⁰¹ McKinsey uses the phrase "*a priori*" where I have used "non-empirical". Since McKinsey defines *a priori* knowledge simply as knowledge obtained independently of empirical information (McKinsey, 1991 p. 9) nothing is lost by the use of my terminology. I have deliberately avoided talk of privileged access as being a form of *a priori* knowledge, since the issue of how to define *a priori* knowledge is contentious and need not concern us in this context. For one, Burge distinguishes knowledge of one's propositional mental events from a *priori* knowledge, on the grounds that the former is essentially tied to a perspective where the latter is not. For more on this distinction see Burge (1996). For recent interesting accounts of the *a priori* see Peacocke (1993), and Boghossian (1996).

environment or facts about the linguistic practices of the subject's community.

The question, then, is twofold. First, is the semantic externalist who wishes to endorse the claim of privileged access committed to the three stated propositions? Second, are the propositions inconsistent? At first gloss, (1) is a statement of the claim of privileged access, (2) is a statement of the thesis of semantic externalism, and (3) is a plausible claim about our knowledge of the external world. For instance, according to Brueckner, the negation of (3) "embodies a claim which is obviously false on anyone's view"¹⁰². The important point here is that semantic externalists do accept (3). So it would seem that the semantic externalist who wishes to endorse the claim of privileged access is indeed committed to the three propositions.

With regard to the alleged inconsistency, McKinsey's idea is this. If a subject had privileged access to a given thought, where her having that thought necessarily implied the existence of certain external objects, then she could know non-empirically that those objects exist. Since a subject cannot know non-empirically that the external world exists, she cannot have privileged access to the thought in question.¹⁰³

However, a satisfactory answer to the twofold question depends upon the resolution of two crucial issues. An appropriate interpretation needs to be determined for the phrase "necessarily depends upon" in (2), and an appropriate content needs to be determined for the proposition E in (2) and (3). These issues are intimately linked, and will be

¹⁰² Brueckner (1992) p. 111.

¹⁰³ It is important to distinguish the *Argument from Privileged Access* from a line of reasoning which argues that one needs to know the empirical conditions which are conceptually entailed by one's self knowledge before one can have that self knowledge. This is clearly fallacious. Just because p conceptually entails q , it does not follow that a subject could not know p without first knowing q . Think about the (perhaps infinite) conceptual entailments of any given proposition.

addressed respectively in the next two sections of this chapter.

It is evident that the alleged incompatibility between semantic externalism and the claim of privileged access can be brought out only by appeal to a third claim about the nature of our knowledge of the external world. I will refer to this third claim as **(EC)**, where (3) above is an instance of the more general claim captured by **(EC)**.

(EC) x could not have non-empirical knowledge of contingent facts about her environment.¹⁰⁴

The terms “empirical” and “non-empirical” are epistemic terms; they distinguish not between types of fact, but between ways of knowing facts. Thus a fact is neither empirical nor non-empirical *per se*, but can be classified as such, if at all, only by bringing in reference to the means by which it came to be known. Consequently, the categories of empirical fact and non-empirical fact are not mutually exclusive. A good illustration of this point is provided by considering Susan’s current belief that she is jealous of her sister. Assuming that this belief falls within the purview of her introspective knowledge, it would be possible for her to know its content both non-empirically, by introspection, and empirically, by noticing how she behaves towards her sister.¹⁰⁵ Others could have only empirical knowledge of the content of

¹⁰⁴ The force of the phrase “could not” in this context is not one of logical impossibility. The possibility that an Omniscient being could have non-empirical knowledge of contingent facts about the environment need not be excluded. **(EC)** is restricted to creatures who are relevantly similar to us: that is, to beings whose knowledge of the world generally comes via external perception. “Could not” should therefore be interpreted as embodying a weaker form of impossibility, perhaps that of metaphysical impossibility.

¹⁰⁵ It should be clear from the discussion of privileged access in chapter two that the principle of privileged access, **(PA)**, is consistent with Susan gaining a large part of her knowledge of her own psychology via empirical investigation.

her belief.¹⁰⁶ The claim embodied in **(EC)** is that knowledge of contingent facts about the environment will *always* be empirical knowledge; there is no way to come to know such facts other than by empirical investigation.

The *Argument from Privileged Access*, then, takes the following three theses.

(PA) For all **x**, if **x** believes that she thinks that *p*, then **x** thinks that *p*;

(SE) **x**'s thinking that *p* necessarily depends upon contingent facts about her environment;

(EC) **x** could not have non-empirical knowledge of contingent facts about her environment;

and runs as follows.

(pi) **(PA) & (SE)**

(pii) **(EC)**

(piii) **{(PA) & (SE)} \emptyset \neg (EC)**

therefore (c) **\neg {(PA) & (SE)}**

It is now time to turn to the first of the two crucial issues, the nature of the necessary dependence of a thought on contingent environmental facts.

3.3 The first crucial issue: "necessarily depends upon"

How are we to interpret the necessary dependence of thought on environment? Two options present themselves. Either semantic externalism should be interpreted as a

¹⁰⁶ One problem raised for Davidson's account of radical interpretation is how to ensure that the propositional mental event which the radical interpreter attributes to the subject, and the propositional mental event which the subject attributes to herself are one and the same propositional mental event; and hence that the interpreter and the subject have knowledge, when they do, of the very same fact. For Davidson on radical interpretation see his (1973). For a defence of Davidson against this objection see McDowell (1994).

conceptual thesis, according to which interpretation **(SE)** would express the following.

(SEC) *x*'s thinking that *p* conceptually implies contingent facts about her environment;

or semantic externalism should be interpreted as a metaphysical thesis, according to which interpretation **(SE)** would express the following.

(SEM) *x*'s thinking that *p* metaphysically entails contingent facts about her environment.

Conceptual implication and metaphysical entailment can be distinguished in the following way: if *p* conceptually implies *q*, *q* will be deducible non-empirically from *p*; if *p* metaphysically entails *q*, however, *q* may not be deducible non-empirically from *p*, since some metaphysical necessities are knowable only empirically. Each interpretation will thus have different repercussions for the *Argument from Privileged Access*, and hence for the semantic externalist who wishes to endorse the claim of privileged access.

Adoption of the conceptual interpretation brings the alleged incompatibility to the fore. If semantic externalism is correctly captured by **(SEC)**, a subject could come to know non-empirically, for any given thought, which contingent empirical conditions that thought depended upon. Such knowledge, together with introspective knowledge that she was currently entertaining a given thought, would provide a non-empirical route to knowledge that those specific contingent empirical conditions in fact obtained.¹⁰⁷ Hence, the conjunction of privileged access and semantic externalism, granted as a conceptual thesis, does indeed contradict the plausible third claim, **(EC)**.

¹⁰⁷ Exactly which such contingent facts could be thus known is still in question. This matter will be dealt with in section 3.4 where I discuss possibilities for the content of the proposition E, using Susan's thought that water is a liquid as an example.

Adoption of the metaphysical interpretation, on the other hand, may have no such unsavoury consequence. If the thesis of semantic externalism is correctly captured by **(SEM)**, there is no immediate reason to suppose that a subject could have privileged access to the contingent facts upon which her thoughts depend, even if she does have privileged access to the contents of those thoughts.

Since metaphysical dependencies are often only knowable [empirically], propositions that are knowable [non-empirically] might metaphysically depend upon other propositions that are only knowable [empirically]. (McKinsey, 1991 p. 13)

The apparent effect of exchanging conceptual implication for metaphysical dependence, then, is to loosen the epistemic relation between knowledge of a thought and knowledge of the contingent facts upon which that thought necessarily depends. Hence, the conjunction of privileged access and semantic externalism, interpreted as a metaphysical thesis, does not lead immediately to a contradiction of the plausible third claim, **(EC)**.

The *Argument from Privileged Access* has no obvious force if semantic externalism is interpreted as a metaphysical thesis. It might seem, then, that the semantic externalist should endorse the metaphysical interpretation, since this would allow her consistently to maintain both the claim of privileged access and the plausible third claim about the nature of our knowledge of the external world. But is the metaphysical interpretation feasible?

McKinsey thinks not. McKinsey maintains that such an interpretation fails to capture the significance of semantic externalism. He writes,

... to say that a state is wide ... cannot mean *merely* that the state metaphysically entails the existence of external objects. For if it did, then given certain materialistic assumptions ... , it would follow that probably *all* psychological states ... would be wide ... and [semantic externalism] would be merely a trivial consequence of (token)

materialism. (McKinsey, 1991 pp. 13-14, original emphasis)

McKinsey's thought is that since one can endorse physicalism independently of the semantic internalist / semantic externalist debate, semantic externalism cannot amount to the mere claim that having a thought entails the existence of external objects. McKinsey provides the following example.

For instance, it is plausible to suppose that no human could (metaphysically) have existed without biological parents, and that no human could (metaphysically) have had biological parents other than the ones she in fact had. If this is so, then Oscar's thinking that water is wet metaphysically entails that Oscar's mother exists. In fact, Oscar's having *any* psychological property ... would metaphysically entail the existence of Oscar's mother. Thus if metaphysical entailment of external objects were what made a psychological state wide, then probably *all* of Oscar's - and everyone else's - psychological states would be wide. (McKinsey, 1991 p. 14, original emphasis)¹⁰⁸

I agree that the metaphysical interpretation as characterised by McKinsey fails to capture the significance of semantic externalism. Semantic externalism is a thesis about what determines the *content* of certain propositional mental events, and not a thesis solely about what physical conditions must be presupposed in order for a subject to be able to think *per se*. The metaphysical dependence of a propositional mental event upon the (prior) existence of the subject's mother is one which holds independently of the particular content of the given propositional mental event, and will therefore be insensitive to the content of that propositional mental event. The passages from McKinsey thus portray the metaphysical interpretation as failing to capture the crucial distinction between what conditions must hold for a subject to be able to think, and what conditions must hold for a subject to be able to think the thought that *p*. In this way,

¹⁰⁸ See Kripke (1980) pp. 312-4.

McKinsey portrays the metaphysical interpretation as failing to capture the significance of semantic externalism.

But need the nature of the metaphysical dependencies be so general? Could there not be some other objects the existence of which were metaphysically necessary for the thinking of a given thought, objects which *were* sensitive to the content of the thought in question? Perhaps McKinsey simply fails to alight upon the more relevant external objects which are metaphysically entailed by a given propositional mental event, and which are sensitive to its content. If there were such objects, specifying them would presumably give more credence to the thought that semantic externalism could be interpreted as a metaphysical thesis. Consider Susan's thought that water is a liquid. Susan's having this thought may well metaphysically entail the (prior) existence of Susan's mother. It may also, more interestingly, entail the existence of H₂O.¹⁰⁹ This latter possibility gestures towards a metaphysical dependence which is sensitive to the content of the relevant thought.¹¹⁰

Suppose, then, that Susan's thinking that water is a liquid metaphysically entails the existence of H₂O. That is, that the following metaphysical entailment holds.

(SEM1) Susan thinks that water is a liquid $\emptyset (\exists) \text{H}_2\text{O}$.

¹⁰⁹ While it may be feasible to alight upon such dependencies for natural kind concepts, it is by no means obvious that the same can be said for non-natural kind concepts. Since I argued in chapter one that natural kind concepts are in fact mediated through the practices of the linguistic community, this leaves the prospect that the prima facie plausibility of the metaphysical interpretation will not hold for natural kind concepts either.

¹¹⁰ What emerges here is an indication of the integral relation between the nature of the dependence of a thought on contingent facts about the environment, and the nature of the contingent facts upon which that thought depends. I deal with the specific entailments in section 3.4 below.

It is important not to get hung up on the thought that Susan could not have non-empirical knowledge of such a metaphysical entailment *just because* it involves knowledge of chemical theory, knowledge which can be gained only via empirical investigation. It may be true that Susan does not understand the term “H₂O”, and that this fact impedes her knowing **(SEM1)**. But this cannot be *all* that the metaphysical account amounts to for the following reason. While the left hand side of the statement of the entailment must specify the concepts Susan employs to think the thought she does, the right hand side will happily tolerate substitution of co-extensional terms. That is, **(SEM1)** is equivalent to **(SEM1')**.

(SEM1') Susan thinks that water is a liquid \emptyset (\exists) water.

The metaphysical interpretation of semantic externalism, if it is to side-step the *Argument from Privileged Access*, has to maintain that Susan could no more have non-empirical knowledge of **(SEM1')** than she could have non-empirical knowledge of **(SEM1)**. And her lack of knowledge of **(SEM1')** certainly cannot be due to her ignorance of chemical theory, since knowledge of **(SEM')** need involve no knowledge of chemical theory. The crucial reason that Susan should not be taken to have non-empirical knowledge of **(SEM')**, on the metaphysical interpretation, is that the entailment itself is knowable only empirically.

I must confess that I find it hard to believe that semantic externalism is an *empirical* theory. For one thing, the fact that semantic externalism is established by thought experiments might be thought to tell against its status as an empirical theory. However, this is not conclusive. Scientists have variously used thought experiments as a means to establishing empirical theories.¹¹¹ Still, such thought experiments as have been used in science are subject to empirical testing, and will be rejected if the empirical evidence is found consistently to tell against them. In the

¹¹¹ For an interesting account of the use of thought experiments in science see the fascinating Brown (1991).

case of semantic externalism, it is hard to see what kind of empirical evidence would refute it.¹¹²

Although endorsing some form of metaphysical interpretation may have the consequence that the *Argument from Privileged Access* poses no genuine challenge to semantic externalism, in the end I agree with McKinsey that the relevant entailments between thought and environment we ought to be considering are conceptual entailments.

..., to say that the [propositional mental event] is wide is not to say something that is true by virtue of [the subject's] *nature* or the *nature* of the particular event Rather it is to say something about the *concept* ... that is expressed by the English predicate 'x is thinking that [p]'; it is to say something about what it *means* to say that a given person is thinking that [p]. (McKinsey, 1991 p. 14 original emphasis)

So from now on I will treat semantic externalism as a conceptual thesis. The *Argument from Privileged Access* therefore poses a *prima facie* worry for the semantic externalist who wishes to endorse the claim of privileged access. The next important step is to resolve the nature of the specific environmental conditions upon which a given thought depends, ones to which the content of the thought in question are sensitive.

3.4 The second crucial issue: the proposition E

I have argued that semantic externalism should be interpreted as a conceptual thesis. What remains to be determined is which environmental conditions are

¹¹² Noonan has argued that in any particular case, the status of a psychological state as object-dependent is "not ... that of an *a priori* knowable conceptual necessity, but rather that of a merely *a posteriori* knowable Kripkean *metaphysical* necessity". (Noonan, 1993 p. 284). However, the kind of object-dependence he has in mind here is not the kind of object-dependence referred to, if any, in the thesis of semantic externalism.

conceptually implied by any given thought content. Let us take as an example Susan's thought that water is a liquid. Thinking back to McKinsey's argument, the question now is whether there is some E which will be both deducible non-empirically from the fact that Susan thinks that water is a liquid, and such as to ensure that non-empirical knowledge of E is implausible. The proposition E is subject to two constraints. First, the environmental conditions which it states must be necessary to determining the *content* of Susan's thought that water is a liquid. Second, its truth must be conceptually entailed by Susan's thinking that water is a liquid.

Anthony Brueckner¹¹³ has maintained that there is no such E. He thus defends semantic externalism against the *Argument from Privileged Access* by denying that the McKinsey recipe can be satisfied. Brueckner considers various possibilities for the content of the proposition E, and suggests the following.

(E1) [Susan] inhabits an environment containing H₂O and not XYZ

If (E1) gave the content of the proposition E, then the semantic externalist would be in trouble, in line with the general McKinsey argument, since it is indeed implausible to maintain that Susan could have non-empirical knowledge of the conditions specified in (E1). However, as Brueckner points out, (E1) cannot be the correct interpretation of the content of the proposition E. (E1) satisfies neither of the two requisite constraints. The environmental conditions which it states are not specific to determining the content of Susan's thought, and neither is it conceptually entailed by her thought. Brueckner's objection to (E1) is related to its failing to satisfy the first of these constraints.

The environmental conditions which it states are not those upon which Susan's thought depends, since it is false that every environment in which Susan could think that water is a liquid is a world containing H₂O. Burge, for example,

¹¹³ Brueckner (1992).

maintains that “it is logically possible for an individual to have beliefs involving the concept of water ... even though there is no water ... of which the individual holds those beliefs”¹¹⁴. According to Burge, then, an individual can come to have water thoughts in ways other than being in causal contact with H₂O. The worry would be that in such a waterless world, nothing in the individual’s environment would license the ascription of water thoughts as opposed to twater thoughts, or any other counterfactually possible thoughts. Burge claims, however, that if the individual were part of a linguistic community “there might ... be enough in the community’s talk to distinguish the notion of water from that of twater and from other candidate notions”¹¹⁵. This would depend both on the community having sufficiently sophisticated chemical theory, and on the existence in their world of enough physical entities to guarantee reference for some of their theoretical terms. According to Burge, the correct attribution of psychological states has to have *some* grounding in physical objects in the individual’s world, so that reference can be secured. In this particular case, for instance, hydrogen and oxygen may have to exist, even though they do not combine in the subject’s world to form water. Thus the existence of water-thoughts in a world with no water, but in which scientists theorised about H₂O, would be contingent upon the assumption that “not *all* of the community’s beliefs involve similar illusions”¹¹⁶. Burge’s idea seems to be that in such a waterless world, there must exist enough physical entities to fix an appropriate content for the community’s theoretical sentences and psychological states.

Burge does not explicitly consider the possibility of a solitary chemical theoretician. There is no reason to believe, however, that Burge would deny such a solitary figure the ability to entertain thoughts about water, given the same

¹¹⁴ Burge (1982) p. 114.

¹¹⁵ Burge (1982) p. 116.

¹¹⁶ Burge (1982) p. 116, original emphasis. The term “illusion” here is misleading. Strictly speaking there need be no illusion involved. Rather, the terms do not refer to items in the relevant world under consideration.

proviso as was made for the waterless community, that not all of his thoughts involve similar illusions. What Burge denies is that an 'ignorant' and 'indifferent' individual could think water thoughts if neither water nor other people exist.

What seems incredible is to suppose that [Oscar], *in his relative ignorance and indifference about the nature of water*, holds beliefs whose contents involve the notion, even though neither water nor communal cohorts exist.¹¹⁷ (Burge, 1982, p. 116, emphasis added)

These considerations lead Brueckner to modify his statement of the environmental conditions which are entailed by Susan's thought. For the moment I will postpone discussion of these modifications, set out below as **(N)**, since it still remains to explain why (E1) fails to satisfy the second constraint, that it be conceptually entailed by Susan's thought.

With regard to the second constraint, then, there are two reasons why (E1) is not conceptually entailed by Susan's thought. First, it is simply not of the right form. Any proposition which is to follow conceptually from the fact that Susan has the thought that water is a liquid, must have the same structure as the proposition that Susan thinks that water is a liquid. That is, the right hand side of the entailment must match up with the left hand side of the entailment.¹¹⁸ This constraint is respected by both of the following.

(SEC1) Susan thinks that water is a liquid as opposed to that twater is a liquid $\emptyset (\exists) \text{H}_2\text{O}$ and $\neg (\exists) \text{XYZ}$.

¹¹⁷ As Burge sets up the example, it is stipulated that the subject knows no chemical theory. Hence, Burge has no immediate interest in the case of the solitary chemical theoretician.

¹¹⁸ This is not a genuine feature of entailment. Clearly, the following are all conceptual entailments.

$$p \emptyset p$$

$$p \emptyset \neg (\neg p)$$

$$p \emptyset (p \vee q)$$

The point is rather to do with the notion of a relevant alternative.

(SEC1') Susan thinks that water is a liquid $\emptyset (\exists) \text{H}_2\text{O}$.

So which one of these should we choose? Brueckner chooses the former. He focuses on the notion of a relevant alternative, and hence suggests that the proposition entailed by Susan's thought must state the environmental conditions which determine that thought to be a water thought *as opposed to a twater thought*. However, according to **(PA)**, privileged access to a given thought does not require the ruling out of various possible alternative thoughts. Not realising this was the mistake of the *Content Sceptic's Argument*. Hence, **(SEC1')** must be chosen above **(SEC1)**.

The second reason for rejecting (E1) is as follows. Knowledge of (E1), stated in those terms, depends upon knowledge of chemical theory, which is knowable only via empirical investigation. Since we are dealing with *conceptual* entailment, both sides of the statement of the entailment must specify the concepts Susan employs to think the thought she does. The statement of the entailment must be sensitive to the way in which the entailment is specified.

Brueckner's reasons for denying that (E1) correctly captures the conditions which are entailed by Susan's thought - that a subject can think about water without living in a world with H_2O - lead him to the following formula.

(N) It is necessary that if S is thinking that water is wet, then either

- (i) water exists, or
- (ii) S theorises¹¹⁹ that H_2O exists, or
- (iii) S is part of a community of speakers some of whom theorise that H_2O exists. (Brueckner, 1992 p. 116)¹²⁰

¹¹⁹ Burge emphasises that for one to theorise about water, for instance, one must be in appropriate causal relations to other particular substances that enable one accurately to theorise about water. Burge (1988) p. 653.

¹²⁰ While it seems correct to maintain that not every environment in which a subject can think that water is a liquid is a world containing H_2O , it is by no means obvious. Burge states that "thinking that water is not fit to drink is different from thinking that H_2O is not fit to drink"

This, however, is of no more use in the McKinsey recipe than (E1), for the same reasons as those given against (E1). Brueckner maintains that **(N)** is knowable only empirically.

Burge's anti-individualism does not commit him to the view that Oscar can know [non-empirically] that *either (i), or (ii), or (iii) is true*, even if Burge's theory does have the consequence that the disjunction in question is metaphysically necessitated by the proposition (knowable [non-empirically] by Oscar) that Oscar is thinking that water is wet. (Brueckner, 1992 p. 116, original emphasis)

Brueckner goes on to determine whether there *is* some interesting proposition, less specific than **(N)**, of which a subject could have non-empirical knowledge on the assumption that he has privileged access to his thoughts. He

(Burge, 1979 p. 76). Two concepts can be distinguished; a water concept, and an H₂O concept. It is of course possible, although perhaps not necessary, that a subject's water concept and a subject's H₂O concept get fixed by the very same stuff. Now, consider Susan, who lives in the actual world prior to any chemical theorising. For any thought Susan were to have about water, it would involve a water concept, and not an H₂O concept, which is, in her state of ignorance, unavailable to her. What about Susan*, who, in her waterless lonely world, theorises about H₂O? Would the situation not be precisely reversed? It could be argued, in broad agreement with Church, that only one concept is available to Susan* when she entertains thoughts about water, that concept being an H₂O concept, and precisely not a water concept. Now consider Susan**, who, while living in a waterless world and being ignorant of chemical theory, is a member of a community some of whom theorise about H₂O. Presumably she is in the same position as Susan*, in that the concept which features in her water thoughts is dependent on the concept had by the chemists in her community. Her concept is determined by deference. These considerations seem to suggest that Susan, Susan*, and Susan** could *not* be attributed the very same thought content. While this line of reasoning is not conclusive, it does urge caution about whether or not a subject in a waterless world could really have a water concept.

reasons as follows. Semantic externalism tells us that in order for a subject to be attributed a water thought, there must be enough in that subject's world to rule out the attribution to him of various possible twin thoughts. It further tells us that the candidates for such content-determining states of affairs are physical entities distinct from the subject in question. As a result, Brueckner comes up with the following proposition as one which may be knowable a priori, if semantic externalism is true.

(E2) It is necessary that if Oscar is thinking that water is wet, then there exist some physical entities distinct from Oscar.

Brueckner maintains that even if (E2) can be known non-empirically, a subject can still at best know that his environment contains physical entities sufficient to fix the contents of his thoughts; which sorts of entities are required remaining to be settled empirically. He seems to disagree with McKinsey in regarding this is an acceptable consequence.

McKinsey says that 'you obviously *can't* [have non-empirical knowledge] that the external world exists' [(1991 p. 16)]. This does seem obvious if the alleged ... knowledge is said to contain much detail concerning the character of the external world distinct from oneself. But if the alleged ... knowledge is simply knowledge that something or other physical exists distinct from oneself, it is not obvious that such knowledge is impossible.
(Brueckner, 1992 p. 118, original emphasis)

It is by no means obvious, however, that the "weak" conclusion which Brueckner would be willing to accept ought to be regarded as acceptable. It is at least a point of contention.¹²¹ Further, I do not accept that the weak conclusion is the only non-empirical knowledge a subject

¹²¹ One possibility might be to deny that "the world exists" is really an empirical statement. I deal with this issue more fully in my discussion of Putnam and Wright in chapter six.

could gain if semantic externalism and privileged access were both accepted. Brueckner has made the mistake, identified in the discussion above concerning McKinsey, of failing to acknowledge all the entailments between a thought and the environmental conditions necessary for that thought. In the next section, it will become apparent that knowledge of semantic externalism, together with privileged access to a thought, can yield non-empirical knowledge of specific facts about the external world; knowledge which is *prima facie* implausible.

3.5 Brown's elaboration of the McKinsey recipe

In this section I will be concerned with an elaboration of McKinsey's argument, presented by Jessica Brown¹²². Brown claims, contra Brueckner, that there is indeed some proposition E which will be both deducible non-empirically from the fact that Susan thinks that water is a liquid, and such as to ensure that non-empirical knowledge of E is implausible. In other words, Brown argues that the McKinsey recipe *can* be satisfied.¹²³

Brown rightly stresses that any entailment between mind and world of which a subject can be supposed to have non-empirical knowledge must not draw on application conditions of concepts of which that subject is ignorant. For example, if a subject does not know that something is water if and only if it is H₂O, she could not know the following entailment.

(P) Necessarily, if x has a thought involving the concept of water, and x is agnostic about the application conditions of water, then either x is in an environment which contains H₂O, or x is part of a community which has a term 'water' which applies to something if and only if it is H₂O. (Brown, 1995 p. 152)

¹²² Brown (1995).

¹²³ Brown follows McKinsey in taking what I have called the *Argument from Privileged Access* as an *ad hominem* charge of inconsistency against Burge.

However, Brown claims that an entailment between mind and world can be formulated which does not require a subject to know the chemical composition of water; that is, a formulation which does not appeal to facts which the subject could know only via empirical investigation. Brown gives the following general formula.

(Q) Necessarily, if x has a thought involving the concept of a natural kind k, and x is agnostic about the application conditions of the concept of k, then either x is in an environment which contains k, or x is part of a community with the concept of k. (Brown, 1995 p. 152)

In the water example, the specific instance of this formula would be,

(P') Necessarily, if x has a thought involving the concept of water, and x is agnostic about the application conditions of water, then either x is in an environment which contains water, or x is part of a community with the concept of water.

(Q) deals with mind-world entailments for natural kind concepts only. The question is whether it allows a subject to gain non-empirical knowledge of her environment. In order for a subject to be able to use (Q) to gain non-empirical knowledge about her environment, she would have to know non-empirically that the antecedent of (Q) was satisfied. That is, she would have to be able to replace "x" by "I", and "k" by a term referring to a natural kind. However, this would require non-empirical knowledge of the status of one's concept as a natural kind concept. Is such knowledge available non-empirically?

It is plausible to maintain that the status of a concept as a natural kind concept is determined by some combination of the physical nature of the world and the intentions of the linguistic community. It is consequently plausible to maintain that the status of a concept as a natural kind concept is not something of which a subject could have non-empirical knowledge. If this is correct, (Q) cannot be used by a subject to gain non-empirical knowledge of contingent external facts,

since the antecedent of (Q) can itself be known only empirically.¹²⁴

However, Brown maintains that a similar formula which deals with mind-world entailments for non-natural kind concepts can also be formulated. For this she draws on the idea that the attribution to a subject of a specific concept requires that there be some fact which determines that *that* concept is the one that should be attributed, as opposed to some other similar concept. This idea is evident in Burge's statement, given above, that for an ignorant subject without water or communal cohorts, there is nothing licensing the attribution of water thoughts as opposed to thoughts about any other possible substance indistinguishable for the subject.¹²⁵ Brown reasons as follows.

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

¹²⁴ For more on the difference between natural kind concepts and non-natural kind concepts see McGinn (1989) especially pp. 30-6.

¹²⁵ See also Burge (1979) pp. 77-83.

Thus Brown formulates the following for non-natural kind concepts.

(R) Necessarily, if x has a thought involving a non-natural kind concept n , and x is agnostic about the application conditions of n , then x is part of a community which has the concept n . (Brown, 1995 p.154)

However, (R) can no more be used to gain non-empirical knowledge of the environment than (Q). The categories of “natural kind concept” and “non-natural kind concept” are jointly exhaustive. If a subject’s knowledge that a given concept is a natural kind concept must be empirical knowledge, so must her knowledge that a concept is a non-natural kind concept. Hence it would seem that the antecedent of (R), just like the antecedent of (Q), can be known only empirically.

However, the *conjunction* of (Q) and (R) together yield the following further principle (S).

(S) Necessarily, if x has a thought involving a concept c , and x is agnostic about the application conditions of c , then either x is in an environment which contains instances of c and c is a natural kind concept, or x is part of a community which has the concept c , whether or not c is a natural kind concept. (Brown, 1995 pp. 154-5)

On the assumption that a subject can have non-empirical knowledge of (Q) and (R), she can likewise have non-empirical knowledge of (S); and, unlike the antecedents of (Q) and (R), a subject *can* have non-empirical knowledge of the antecedent of (S). Hence (S) can be used by a subject who has privileged access to the content of a propositional mental event to gain non-empirical knowledge of her environment.

The consequent of (S), then, is a schema for the proposition E in the McKinsey recipe. Taking the example of a subject’s thought that water is wet, the schema can be filled in to yield the following specific proposition E of which a subject could have non-empirical knowledge.

(T) Either my environment contains water and the concept of water is a natural kind term, or I am part of a community which has the concept of water.

That such information is available to a subject without recourse to empirical investigation contradicts the claim embodied in **(EC)**, that one can have only empirical knowledge of the external facts which individuate one's thoughts.

3.6 Tye and McLaughlin

Tye and McLaughlin have tried to block Brown's argument by denying that semantic externalism is committed to (R). They argue that one cannot have non-empirical knowledge of (R), since (R) is false. If this is correct, neither can one have non-empirical knowledge of (S), since the truth of (S) depends on the truth both of (Q) and of (R). Hence if (R) is false there would be no non-empirical route to contingent facts about the environment, and the *Argument from Privileged Access* would be fallacious. What does their claim that (R) is false amount to?

Tye and McLaughlin make two claims in support of their view that (R) is false. First, they claim that a subject *could* have a non-natural kind concept without being a member of a linguistic community. I agree that this is plausible. However, what would need to be the case for (R) to be false is not simply that a subject could have a non-natural kind concept without being a member of a linguistic community, but rather that a subject could be attributed a non-natural kind concept without being a member of a linguistic community, *even though that subject is unable to apply her concept correctly*.

It should be stressed that the kind of ignorance which is at issue here is ignorance of a kind which could not be resolved by the subject gaining more information about the object concerned.¹²⁶ For example, suppose Susan was wondering

¹²⁶ The caveat "... other than information which exploits the concept under discussion ..." is necessary, since Oscar could resolve every issue of whether a given object α fell under a concept ϕ simply by being told, by

whether or not Oscar had a wooden leg, but was too shy to ask. Here we have a case in which Susan is unable to apply her concept correctly. However, the issue could be resolved simply by Susan gaining more information about the object concerned - while Oscar was sleeping, for instance. Crucially, Susan's ignorance stems from her lack of perceptual knowledge, and is not purely a product of semantic ignorance. The kind of ignorance which Tye and McLaughlin need to allow for is, however, purely semantic ignorance. According to their position, it must be possible for a subject to possess a given concept which she is unable to apply correctly to a given object, even though she is in an ideal perceptual position with regard to that object. This is what I claim is implausible.

The following scenario might be thought to favour Tye and McLaughlin's position, and hence be raised as an objection to my position. Suppose there is a subject who lives, and has always lived, alone on an island. Call him Solo. Suppose further that from time to time there is a food shortage, and Solo has to go for a number of days with no food. To refer to those times of hardship which last for more than four days, Solo introduces the term "longfast". Now suppose that, at some time later, and having gone without food for four days, Solo wonders whether or not he is in a period of longfast. Solo's deliberation in this instance comes not from any lack of knowledge of the length of time for which he has not eaten, since he always records such things faithfully. Solo's deliberation comes rather from the fact that he can no longer remember whether he introduced the term "longfast" to apply to periods of more than four days or to periods of more than three days. Solo is aware of the fact that if the latter is correct he is in a period of longfast, but if the former is correct he is not yet in a period of longfast.

It looks, then, as if we have exactly the kind of scenario which I maintain is implausible: that is, a scenario in which a solitary subject has a non-natural kind concept even though that subject is unable to apply his concept correctly, and that

a sincere and knowledgeable subject, that α was ϕ , or that α was not ϕ .

not through any perceptual inadequacies. However, on closer examination it will be seen that this is not so. The example is one in which the crucial notion of deference is used. While it is true that Solo does not defer to a linguistic community consisting of current subjects other than himself, he nevertheless defers, on this view, to a subject other than his present self, namely his past self, the introducer of the term “longfast”. This response depends upon the notion of deference being the crucial notion employed by semantic externalism, rather than the notion of a current linguistic community.

I maintain that the attribution of any non-natural kind concept to an independent subject will depend upon the subject’s knowing the application conditions of that concept. If this is correct, (R) will stand up against the Tye and McLaughlin’s first claim, so long as a subject can be assumed to have non-empirical knowledge, at least some of the time, of whether or not she can apply her concept correctly, and this I see no reason to dispute.

The second claim that Tye and McLaughlin make in support of their view that (R) is false is that, in establishing (R), Brown commits herself to the claim that Oscar actually *has* a non-natural kind concept. But this cannot be right. At most, the wording of the example is misleading. The fact that Oscar in his actual linguistic community has the concept *sofa*, and that counterfactual Oscar in the counterfactual linguistic community would have the concept *chofa*, does not obviously falsify the claim that Oscar in a counterfactual solitary world could be attributed no relevantly similar concept. Clearly, it is not Oscar-as-he-is who is under consideration, but a counterpart of Oscar who has always been in the solitary situation.¹²⁷ The question, then, is what concept Tye and McLaughlin believe can be attributed to *that* counterpart of Oscar.

¹²⁷ It is easier to take the extreme example of a subject who has always been alone in his world, since this avoids issues such as whether a subject could keep a non-natural kind concept after a long period of time after travelling to an otherwise uninhabited world.

Twin Earth thought experiments involving natural kind concepts differ from those involving non-natural kind concepts in one important respect. In the former type of example, we are presented with subjects who, because of the physical difference in their environment, have distinct, mutually exclusive concepts, referring to distinct sets of instances. Susan can have water thoughts but not twater thoughts; and the instances her concept ranges over are necessarily distinct from the instances over which her twin's concept ranges. In the latter type of case, on the other hand, the referents of the actual concepts, say *sofa*, or *arthritis*, constitute a *subset* of the referents of the twin concept, *chofa*, or *tharthritis* respectively. Oscar and twin Oscar can refer to the very same object by means of their distinct concepts. The difference is that twin Oscar can refer to more with his concept than Oscar can with his. It is perhaps this idea that a non-natural kind concept and its twin concepts form a hierarchical set which leads Tye and McLaughlin to attribute a concept to solitary Oscar: he at least has *one* of the possible concepts in the hierarchy. However, it seems that for any concept we could attribute to solitary Oscar, there is no reason to attribute him *that* concept as opposed to any of the other possible concepts constituting the relevant hierarchy. If that is the case, it is implausible to attribute him with *any one* of the concepts in the relevant range, since any specific attribution would be completely arbitrary. The only alternative would be to stipulate that solitary Oscar be attributed with either the *least* inclusive concept, or the *most* inclusive concept. But this is implausible. First, the attribution would be just as arbitrary; and second, it is not clear that there *are* concepts at either end of the range.

Semantic externalism, at least as endorsed by Burge, is committed to (R). Hence, semantic externalism is *prima facie* committed to the claim that a subject could come to have non-empirical knowledge of (S), knowledge which, together with privileged access to her thought-contents, can be used to gain knowledge of the external world via conceptual

analysis and introspection alone. This is inconsistent with **(EC)**.

3.7 Conclusion

The *Argument from Privileged Access* demonstrates that from the conjunction of semantic externalism and privileged access, one can generate specific arguments such as the one below. I will refer to such arguments as *instance arguments*, since they take one from an instance of a thought to a contingent fact about the external world.

(1) I am thinking a water-thought
 (2) If I'm thinking a water-thought, then I'm in a water-world
 therefore (3) I'm in a water-world.

It is a necessary condition for a subject's being in a water-world that the following disjunction be true. Either the subject's environment contains water and water is a natural kind, or the subject is part of a community which has the concept of water, whether or not water is a natural kind.¹²⁸

Given that introspection can yield knowledge of premise (1), and conceptual analysis can yield knowledge of premise (2), it would seem that if semantic externalism is true, the conclusion (3) can be known on the basis of introspection and conceptual analysis alone. This contradicts the plausible claim embodied in **(EC)**.

Hence the *Argument from Privileged Access* urges the rejection of one of the following theses.

(PA) For all x , if x believes that she thinks that p , then x thinks that p .

¹²⁸ It may be objected that there are other disjuncts which have not been taken into account. However, if semantic externalism is a conceptual thesis, non-empirical knowledge of all the disjuncts will be available. That a complete set has not yet been put forward would tell at most against the current state of our knowledge, and not at all against the *Argument from Privileged Access*.

(SE) x 's thinking that p necessarily depends upon contingent facts about her environment.

(EC) x could not have non-empirical knowledge of contingent facts about her environment.

Privileged Access to the World

I cannot resist a methodological reflection. It may happen that one is committed to delivering a paper and that one discovers, at the last minute, and to one's horror, that one's theory has an absurd consequence, a consequence so absurd that if it is pointed out by a critic it will, without further ado, be taken as a refutation of one's position. Now the best thing to do in this deplorable situation is to point out the disastrous consequence oneself, before anyone else can notice it, and to embrace it. (Schiffer, 1987 p. 80)

4.1 Introduction

In chapter three, I set up what I have termed the *Argument from Privileged Access*. The Argument demonstrates that from the conjunction of semantic externalism and privileged access to a thought, one can generate an *instance argument* such as the one below.

(1) I am thinking a water-thought
 (2) If I'm thinking a water-thought, then I'm in a water-world
 therefore (3) I'm in a water-world.¹²⁹

As we have seen, knowledge of the premises is non-empirical knowledge. Privileged access yields knowledge of premise (1), and conceptual analysis yields knowledge of premise (2). Consequently, it would seem that a subject could gain non-empirical knowledge of contingent facts about her environment.

¹²⁹ It is important to remember that it is a necessary condition for a subject's being in a water-world that the following disjunction be true. Either the subject's environment contains water and water is a natural kind, or the subject is part of a community which has the concept of water, whether or not water is a natural kind.

The *Argument from Privileged Access*, then, puts pressure on anyone who wishes to maintain the following three theses.

(PA) For all x , if x believes that she thinks that p , then x thinks that p .

(SE) x 's thinking that p necessarily depends upon contingent facts about her environment.

(EC) x could not have non-empirical knowledge of contingent facts about her environment.

Responses to the Argument have thus far taken one of two lines. Proponents of semantic internalism have claimed that the derivation of empirical knowledge from introspection and conceptual analysis constitutes a *reductio ad absurdum* of semantic externalism.¹³⁰ Proponents of semantic externalism, accepting that such a derivation *would* constitute a *reductio* of their position, have expended their time and energy trying to show that no such derivation is possible.¹³¹ I call those who take the first line *incompatibilists*, and those who take the second line *compatibilists*. It is their dislike of instance arguments which they have in common. In this chapter I will examine a number of compatibilist responses, explaining why each ultimately fails. I will then offer a solution which disagrees with the standard divide of the debate; one which challenges the common assumption behind the debate. My position is that the conjunction of semantic externalism and privileged access does indeed have the consequence that one can come to know truths about one's environment via introspection and conceptual analysis. In this respect I am in accord with the incompatibilists. However, I maintain that this consequence should in fact be *embraced*, and in this respect my position falls squarely within the compatibilist camp. I argue that inferences from introspective knowledge to empirical knowledge are not to be seen as *intrinsically*

¹³⁰ See McKinsey (1991) and (1994), Brown (1995), and Boghossian (1997).

¹³¹ See Brueckner (1992) and (1994), Warfield (1992), Millar (1997), and Tye and McLaughlin (1997).

unacceptable; on the contrary, there is a certain class of such inferences which are legitimate, and the *Argument from Privileged Access* deals only with those inferences which fall into this class.

My claim is that we should reject neither **(SE)**, semantic externalism, nor **(PA)**, the claim of privileged access. Rather, taking semantic externalism seriously, we should reject the third claim, **(EC)**, that we could not have non-empirical knowledge of contingent facts about our environment. Instance arguments are to be accepted as a valid means of arriving at knowledge of contingent facts about the environment.

4.2 Compatibilist responses

The purpose of this section is to review a number of ways in which it might be thought that compatibilists could argue against instance arguments. As will become clear, no response along these lines is satisfactory.

A good argument can be thought of as a way of transferring knowledge: deductive inference can yield knowledge of a proposition if it is validly inferred from premises which are themselves known. One compatibilist response, then, would be to deny that a subject has *knowledge* of her thoughts. This would ensure that the conclusion of instance arguments could not be *known*. Is it feasible to maintain that a subject's judgements concerning her thoughts somehow failed to be knowledgeable judgements, even in the face of **(PA)**. Certainly, more than truth is needed if a judgement is to count as knowledgeable. But **(PA)** yields not just truth, but infallibility. Is it feasible to maintain that a subject's judgements concerning her thoughts somehow failed to be knowledgeable even though those judgements were infallible?

Consider the following Cartesian scenario. Suppose that in the middle of his meditations, Descartes decides that the project of enquiry upon which he has embarked is proving far too difficult, and that rather than continue on this tortuous

path, he will toss a coin. He resolves, in the event of the coin landing heads up, to believe that the proposition, *I exist*, is true, and in the event of its landing tails up, to believe that it is false. Suppose further that the coin, having duly been thrown towards the heavens, lands heads up. Descartes, according to his resolution, duly believes a proposition which could not but be true; but we would not, I think, want to accord his belief the status of *knowledge*, since he could just as easily have come to believe the opposite.¹³²

It would seem, then, that infallibility is no more sufficient for knowledge than truth. What is missing in the Cartesian scenario, I take it, is *reliability*.¹³³ Although the proposition, whilst thought, could not but be true, the method leading to its adoption could just as easily have secured the alternative, necessarily false belief. **(PA)**, however, differs from the Cartesian scenario in this very important respect. Reflection on **(PA)** gives a subject a *reliable method* by which to arrive at truths about the thought she is currently entertaining.¹³⁴

The only remaining possibility of denying that a subject could have knowledge of her thoughts appears to be to maintain, along with Wittgenstein, that it *makes no sense* to say of a proposition that it is known, if that proposition is guaranteed to be true¹³⁵. On this account, infallibility is not

¹³² For this example I am grateful to John Watling.

¹³³ I am here adopting a reliabilist epistemology. For reliabilist accounts of knowledge and justification see Armstrong (1973), and Goldman (1986).

¹³⁴ This is not to say that when a subject entertains the proposition *I exist* she can not know that it is true. Reflection on the nature of the proposition will yield such knowledge. The point is rather that *in the situation described*, Descartes' belief would not have counted as knowledge. Similarly, if a subject came to believe that she was thinking of water on the basis of the toss of a coin, her belief would be true, infallible, and yet not known. It is not mere conformity with **(PA)** which yields knowledge of any given thought, but rational reflection on the truth of **(PA)**.

¹³⁵ Similar reasoning is widespread with respect to the status of a representation: nothing can be a representation unless it can

insufficient for knowledge, rather it is *incompatible* with knowledge. According to Wittgenstein, then, the possibility of error is a necessary condition for a belief to count as knowledge. Thus the account of privileged access given to accommodate semantic externalism is not, on this view, an account of self *knowledge*, since the possibility of error is ruled out.

For the moment, let us set aside particular responses, such as Wittgenstein's, and turn to a positive proposal to the effect that self-ascriptions *must* typically be knowledgeable. Ironically, perhaps, Burge has been concerned to defend such a claim.¹³⁶ Burge does not himself address the problem raised by instance arguments. However, his claim that self-ascriptions are generally knowledgeable rules out one line of response against them. His defence of the claim that a subject's judgements about her thoughts must count as knowledgeable comes as the second stage in a three-part argument. First, he argues that an epistemic entitlement to one's second-order judgements is required for critical reasoning. Second, he argues for the stronger thesis that critical reasoning further requires that one *know* one's thoughts. Third, he argues that this knowledge must take a distinctive non-observational form. This third stage will be of no concern to us here; but an examination of the first two stages of his argument will prove instructive.

The notion of epistemic entitlement is assumed to be broader than the ordinary notion of justification.

An individual's epistemic warrant may consist in a justification that the individual has for a belief But it may also be an *entitlement* that consists in a status of operating in an appropriate way in accord with norms of reason, even when these norms cannot be articulated by the individual that has that status. We have an entitlement to certain perceptual beliefs or to certain logical inferences even though we may lack reasons or justification for them. (Burge, 1996, p. 93, original emphasis)

misrepresent.

¹³⁶ Burge (1996). See also Peacocke (1996).

The epistemic warrant to much of our self-knowledge is taken to be of this sort. Judgements concerning our thoughts often lack justifying argument or evidence. They are immediate and non-inferential. The characteristic feature of the epistemic entitlement to self-ascriptions on the semantic externalist's account is that it presupposes understanding, which according to semantic externalism is local to a given environment, and yet the entitlement itself is capable of surviving environmental switches.

Burge's argument for our epistemic entitlement to our self-ascriptions runs as follows,

... if one lacked entitlement to judgements about one's attitudes, there could be no norms of reason governing how one ought check, weigh, overturn, confirm reasons or reasoning. For if one lacked entitlement to judgements about one's attitudes, one could not be subject to rational norms governing how one ought to alter those attitudes given that one had reflected on them. If reflection provided no reason-endorsed judgements about the attitudes, the rational connection between the attitudes reflected upon and the reflection would be broken. So reasons could not apply to how the attitudes should be changed, suspended, or confirmed *on the basis of* reasoning depending on such reflection. But critical reasoning just is reasoning in which norms of reason apply to how attitudes should be affected partly on the basis of reasoning that derives from judgments about one's attitudes. So one must have an epistemic entitlement to one's attitudes. (Burge, 1996 pp. 101-2)

Burge argues that the self-ascriptions to which we are epistemically entitled should be regarded as knowledgeable since systematic error and Gettier-type counterexamples are impossible. Systematic error is ruled out because reflection could not add a rational element to the complex process of critical evaluation and subsequent action unless reflective judgements were normally true. Similarly, if reflection were connected to the truth of such judgements in an accidental

way, the entitlement itself would be undermined. Critical reasoning would not be possible if reflective judgements were either systematically false, or true in an accidental, non-knowledge-yielding way. Critical reasoning is not only possible, but is a commonplace occurrence. Critical reasoning requires knowledge of our thoughts. Therefore, we must have knowledge of our thoughts.

There is a question as to whether the account of privileged access embodied in **(PA)** provides a sufficiently rich form of knowledge to bear any connection with the sort of complex rational reflection on one's beliefs that concerns Burge. However, denying the compatibilist this line of response does not depend upon accepting Burge's reasoning here. **(PA)** by itself provides sufficient reason to attribute a subject with knowledge of her thoughts.

One further response runs as follows. While it is generally held that arguments provide a means by which to gain knowledge, certain exceptions have been acknowledged. For instance, Robert Nozick's counterfactual theory of knowledge has the consequence that knowledge is not closed under known entailment.¹³⁷ This is supposedly shown by arguments such as the following.

(1') I am working at my desk
 (2') If I am working at my desk then I am not a
 brain-in-a-vat
 therefore (3') I am not a brain-in-a-vat

According to Nozick's counterfactual theory, (1') and (2') can be known, but (3') cannot be known as a result of knowing (1') and (2'). The principle of closure under known entailment will fail only in cases where the relevant propositions are assessed relative to different sets of possible worlds. However, it is implausible to maintain that the principle of closure under known entailment fails for instance arguments for *this* reason, since there is little prospect of maintaining

¹³⁷ See Nozick (1981).

that the premises and the conclusion should be assessed relative to different sets of possible worlds.

Claiming that the conclusion of instance arguments are not *known* does not, I think, allow one to side-step the real problem posed by the *Argument from Privileged Access*. Those worried by the Argument will surely take little comfort in the response that one cannot *know* the conclusion of an instance argument, since it is worrying enough that by using such instance arguments we always come to have true beliefs about our environment. Moreover, *we can know that this is so*. The potency of the *Argument from Privileged Access* does not turn solely on whether or not we can *know* the conclusions of instance arguments. Having the means to arrive systematically at true beliefs about our environment is surely bad enough.

Personally, I find it highly implausible to maintain that the conclusions of instance arguments are not known, given that they provide one with a reliable method of obtaining true beliefs. I see no reason to deny a subject epistemic warrant for believing the conclusion of an instance argument. If a subject knows that she can systematically come to believe true propositions, what more epistemic warrant could be needed? The epistemic warrant for holding the belief that I am in a water-world comes from my knowledge that this belief was reasoned to by way of a reliable method. Hence the *Argument from Privileged Access* stands.

4.3 Instance arguments reconsidered

Compatibilists have made no headway in responding to the *Argument from Privileged Access* by criticising instance arguments. The Argument establishes that if semantic externalism is true, and privileged access possible, substantial knowledge of empirical facts can be inferred from introspective knowledge and conceptual analysis. I want to argue that this should be embraced as a natural and acceptable extension of semantic externalism. The claim which ought to be rejected is the claim embodied in **(EC)**.

My vindication of instance arguments will take the following line. First, I will explain why it is that instance arguments deliver truths about one's environment. Second, I will identify a causal feature which is generally lacking in inferences from introspective knowledge to empirical knowledge, the lack of which renders such inferences unacceptable. That is, inferences from introspective knowledge to empirical knowledge are not to be seen as *intrinsically* unacceptable, rather they should be seen as unacceptable only in so far as they lack the necessary causal feature which would, were it present, justify any such inference. Finally, I will show that semantic externalism ensures the presence of this necessary causal feature in instance arguments, and therefore that the inference from introspective knowledge to empirical knowledge is in these cases acceptable. In this way I will disarm the *Argument from Privileged Access*.

How is it, then, that instance arguments can without fail deliver truths about the subject's environment? **(PA)** is a principle of privileged access to one's thoughts. Since it would be impossible to know the content of a thought without knowing which concepts occurred in that thought, the claim that a subject has privileged access to her concepts is entailed by the claim that a subject has privileged access to her thoughts. This entailment is important for the argument which follows.¹³⁸

Concept-acquisition, on the semantic externalist view of things, is rather like photography. Photography is a method by which information about the external world can be recorded for future reference. Various complex, physical, causal processes are in play, whereby certain amounts of light reflecting off objects in a given situation interact for a precise amount of time with a piece of photographic paper, thus producing an image of the original scene.¹³⁹ The resulting photographs, once developed, can be brought out at any later

¹³⁸ Given the way I have formalised the claim of privileged access, there is in fact little more to privileged access to a thought than access to the concepts involved in that thought. The entailment, then, is trivial.

time, and used as the evidence from which to infer to the existence of a past event or state of affairs in the world. The *Argument from Privileged Access* relies upon making explicit the similarity between photography and concept-acquisition. According to semantic externalism, one of a unique set of possible causal processes is necessary for the acquisition of any given concept. Causal interaction with the environment imprints concepts, mental photographs, in our minds. These concepts can be thought of as items which are essential to the storing of information about the external world, information which can be used at a future time for the purposes of thought and communication. As a consequence, any mental concept, once acquired, can, just like a photograph, be used as the evidence from which to infer to the past existence of the state of affairs which led to the existence of the requisite concept.

It could be objected that, while it may not be physically possible, it is certainly logically possible that “phoney photographs” be produced, which depict scenes that do not in fact exist. The crucial claim I am endorsing on behalf of semantic externalism is that there could *never* be an analogous situation in the case of concepts.¹⁴⁰ This is precisely where the analogy breaks down. Instance arguments can be used to yield beliefs about one’s environment which are *guaranteed* to be true, whereas there is at least a logical possibility that inferences from photographs to the environment yield false beliefs. Clearly the possibility of error is not by itself sufficient to undermine knowledge. Photographs provide a perfectly good route to knowledge about the past. However, that there is a possibility of error leaves one open to the sceptic. The inference from a thought to a fact about the world is, on the other hand, demon-proof.

¹³⁹ There are various ways in which photographs can be produced, but all of them result in an image of the original scene.

¹⁴⁰ It should be remembered that I am throughout this discussion assuming a version of semantic externalism as espoused by Burge. I accept that there may be forms of semantic externalism which are not committed to this specific kind of causal theory of reference.

Why is the analogy certain to break down? The phoney photograph case depends on there being a possible way of getting just the right amount of light in just the right places on a piece of photographic film, so that the resulting photograph looks exactly the same as it would have done if the state of affairs depicted had in fact caused the image, but where the appropriate causal connection, between possible state of affairs and film, has been lost. This is a possibility. But according to semantic externalism, there is precisely no way to get just the right concept in the mind without the actual causal connections being in place; and herein lies the crux: *causal contact (either to a natural kind or to a linguistic community) is a necessary condition for the acquisition of a concept*. To maintain that one could break the causal link, and yet produce the same effect, is to deny semantic externalism, which is not to win the game, but rather to give it up.

Could one be mistaken in one's thought that one had a certain concept? Given the entailment between privileged access to one's thoughts and privileged access to one's concepts, one could no more be mistaken in one's thought that one had a given concept than in one's thought that one was entertaining a given propositional mental event. Putting **(PA)** contrapositively illustrates this.

(PA) For all x , if $\neg(x$ thinks that p), then $\neg(x$ believes that she thinks that p).¹⁴¹

There are two further important disanalogies between a photograph and a concept. In the former case the relevant inference takes one to a specific fact about the world: that Ralph once went digging for gold, say. In the latter case, however, the relevant inference takes one not to a specific

¹⁴¹ **(PA)** entails that a subject could not think she had a concept which she did not in fact have. I take this claim to be backed up by the fact that semantic externalism is at least partly motivated by the thought that a subject can have a concept while having only partial understanding of the reference of that concept. How much understanding can be tolerated in any given case is presumably a vague issue.

fact, but to a *general* fact about the world: that either there's gold, or there's a community that thinks about it. This difference can be invoked as an explanation for the crucial difference mentioned above; that photographs concern specific incidents allows for the possibility that there could be phoney photographs. It is precisely because instance arguments take one to a *general* fact that rules out the possibility of error. The second further difference is this. The main purpose of photographs is arguably to store information about the past.¹⁴² The main purpose of concepts, on the other hand, is presumably not to store information about the past, but to be used for the purposes of thought and communication. However, that concepts *can* be used for the purposes of thought and communication is precisely because they encode information about the world.

Time to make a qualification explicit. Instance arguments were initially presented as arguments by means of which one could come to know facts about one's current environment. However, it is not facts about one's *current* environment, but rather facts about one's *recently past* environment which can be known. Consider Susan. Suppose she is abducted from Earth, and placed on twin-Earth. According to semantic externalism, a change in one's concepts is not immediate, but requires the passing of a sufficient length of time, so that appropriate causal interaction with the new environment can be established. For an indefinite period of time, then, Susan can think, for example, about water, and run through various instance arguments. However, Susan would conclude something false if she concluded that she was in a water-world. She would be correct, on the other hand, to conclude from her introspective knowledge and conceptual analysis that *in the near past* she *had been* in a water-world.¹⁴³ That is, a subject can be sure, by running through instance

¹⁴² I would not of course want to rule out photographs as pieces of art, but even in these cases, original scenes are depicted.

¹⁴³ This qualification results from the discussion in chapter two concerning the possibility of privileged access to past propositional mental events, and the *Argument from Memory*.

arguments, that she used to lie in certain specific causal relations in her near past.¹⁴⁴

4.4 A necessary causal connection

I claimed that inferences from introspective knowledge to empirical knowledge were not intrinsically unacceptable, but rather that they were so only in so far as a certain causal feature was lacking. The reason that we are resistant to embracing instance arguments, I maintain, is because we are used to such arguments lacking this causal feature, so that the situation which evolves from the *Argument from Privileged Access* bears a superficial similarity to a situation which we would be justified in rejecting as impossible.

Consider individualism. According to individualism, no causal contact is needed for the acquisition of a concept; rather, subsequent causal contact is needed to be able to know whether the concept one possesses has reference in the world.¹⁴⁵ On the picture individualism presents, one can imagine a subject equipped with certain concepts prior to exposure to the world. It would certainly make sense to be worried if it turned out that such a subject could correctly, and infallibly, infer from her thoughts involving experience-independent concepts to the nature of her world; it is always possible on the individualist picture that the subject have just those concepts she does have, and yet those concepts pick nothing out in her world. Individualism presents us with a picture of an isolated subject trying to determine what relation she bears to an independent, external world. The main obstacle to accepting the soundness of instance

¹⁴⁴ This means that the sceptical hypothesis that one is currently a brain-in-a-vat remains a logical possibility. The sceptical hypothesis that in the near past one was a brain-in-a-vat, however, is ruled out. Since this is the concern of chapter six, I will postpone further discussion until then.

¹⁴⁵ It is of course consistent with individualism that the way we in fact learn concepts is via causal contact with the world. What individualism must maintain, however, is that it is logically possible that a subject have thoughts, and therefore concepts, independently of the way the world is.

arguments is the fear that they present as a genuine possibility a subject who can know about the environment independently of any causal contact with that environment.¹⁴⁶

But now consider semantic externalism. Semantic externalism denies the very coherence of the picture described above. There is simply no way a subject could ever have the concepts she does have without either the referents of those concepts, or other people existing.¹⁴⁷ Concepts cannot, as it were, be programmed in by anything other than the actual referents of those concepts or the practices of the linguistic community. Semantic externalism requires that there be causal contact right at the stage of concept-acquisition; and this means that there is already no room for the possibility that one's concepts do not refer.¹⁴⁸

The widespread resistance to instance arguments assumes a view of the self, and in particular of introspection, which is largely a hangover from Cartesianism. Semantic externalism is precisely the view that an individual cannot be regarded as complete with thoughts independently of any prior causal contact with a specific given environment. The self can no longer be regarded as an entity completely separate from her environment. As a result, the apparent clear divide between the mind and the world is eroded. The world we inhabit determines our mental capacities, our ability to think certain thoughts. To suppose that we can "look into our minds" and see things which are themselves independent of the

¹⁴⁶ The difference between semantic internalism and semantic externalism is similarly illustrated by the way in which each theory treats *accidental replicas*. Proponents of semantic externalism typically deny that an accidental replica should be attributed any intentional states. For further discussion of accidental replicas see Millikan (1984) p. 94, and Papineau (1993) pp. 91-4.

¹⁴⁷ Again, this is for any concept about the application conditions of which the subject is agnostic.

¹⁴⁸ I am using the term "reference" here loosely. As is evident from the preceding chapter, a subject cannot infer, for instance, that the substance referred to by her concept exists in the world, since this would be to ignore the role of the linguistic community to which she belonged.

environment is the mistake of the individualist.¹⁴⁹ According to semantic externalism, the concepts to which we have privileged access themselves bear the trace of the specific empirical conditions which led to their acquisition.

Of course it would be unacceptable if a subject could come to know about the external world *just* by looking inside her mind; that is, despite the lack of prior causal contact between that subject and the world. But instance arguments do not allow a subject to argue from world-independent facts to facts about the world, but rather to argue from the way the world is, via the mark the world leaves on her, back out to the way the world must have been to leave such a mark. Without prior causal contact, there is no concept available to introspection.

One entailment of **(PA)** is that we have infallible access to the concepts which feature in our propositional mental events; to the concepts we possess at a given time. The concepts themselves are quite clearly dependent on contingent empirical facts, and semantic externalism tells us about the acquisition-conditions of those concepts. So the conjunction of the claim of privileged access with semantic externalism amounts to the claim that we can have direct knowledge both of the concepts we possess at a given time, and of the acquisition-conditions of those concepts. Putting things this way highlights the fact that for a subject to gain empirical knowledge of the world via instance arguments, it is not enough that semantic externalism be true, the subject must have knowledge of semantic externalism. For any truth, however, it is at least possible that it be known; and that a subject *could* come to know such empirical truths must be accounted for on this basis.

This does not, I think, mean that we have a new crisis in epistemology, or in the philosophy of mind. It does not even mean that empirical science becomes a purely non-empirical activity. Certainly, introspection becomes a viable method of acquiring knowledge of our environment; but it must be

¹⁴⁹ The metaphor of looking into one's mind is particularly suited to the individualist picture, and should probably be rejected along with the rejection of individualism.

recognised that introspection will yield knowledge *only* of those empirical facts that the subject could already have come to know via empirical means. Here it is worth reflecting on the function of memory. As I am, at the present moment in time, I can “look inside my mind” and produce various pieces of empirical knowledge: for instance, that the battle of Hastings was fought in 1066. Why are people willing to accept memory as a route to empirical knowledge? Presumably because memory is *recognised* as a way of *retrieving* information which was acquired via empirical means at an earlier time, even if the means by which it was acquired can themselves no longer be remembered.¹⁵⁰ Semantic memory is of this type. I may be unable to recollect when and how I learnt certain of the concepts I possess, but this does nothing to impugn my knowledge of those concepts. Instance arguments similarly yield knowledge which is based on beliefs that were acquired empirically at an earlier time. This is not, however, to say that instance arguments do not yield new knowledge. They do. They yield new knowledge in just the same way that deductive arguments generally yield new knowledge: they clarify the consequences of the knowledge we already have.

4.5 What should we say about Vatbrain?¹⁵¹

There is a further worry with instance arguments which takes the form of a question: namely, what should we say about Vatbrain? There are two available options. But it is important to notice that no matter which option is preferred, instance arguments can still be admitted as a genuine route to non-empirical knowledge of contingent facts about the

¹⁵⁰ The memory case bears a close resemblance to the photograph case. First, it is possible to have false memories. Second, the inference from a memory is an inference to a specific as opposed to a general state of affairs. Once again, that memories concern specific incidents allows for the possibility that a subject could have false memories. Third, the main purpose of memories is arguably to store information about the past.

¹⁵¹ Vatbrain is, as the name would suggest, a brain-in-a-vat.

environment. That is, on no account does Vatbrain constitute a counterexample to the validity of instance arguments.

The two options, then, are the following. Either Vatbrain can think, or he can't. To assume the first disjunct is to maintain that it is legitimate to attribute propositional mental events to Vatbrain. On this assumption, there is no reason not to suppose that Vatbrain could reason through instance arguments and thereby come to know, without recourse to empirical investigation, certain facts about his environment. Although the instance arguments available to Vatbrain would be syntactically identical to those available to Susan, for example, the propositions expressed by the respective thoughts of the subjects would, of course, be different. Vatbrain can no more reason to the fact that he is in a water-world than Twin Susan can. With their respective, syntactically identical instance arguments, Susan would conclude that she was in a water-world, Twin Susan would conclude that she was in a twater-world, and Vatbrain would conclude something like, perhaps, that he was in a water-in-the-image-world.¹⁵² Each of their respective conclusions would be true.

What about the second possibility, according to which Vatbrain cannot think? For this scenario to present a counterexample to the validity of instance arguments, it would have to be assumed that it could *seem* the same to Vatbrain, even though he somehow failed to have propositional mental events. The reason this would provide a counterexample is because it would raise the sceptical possibility that *we could be in that position*. We could think we were gaining knowledge about our environment, but would be unable to rule out the possibility that we were just like Vatbrain; that there was nothing but phenomenology¹⁵³.

¹⁵² This possibility is owed to Putnam (1981), where it is used as the basis for a transcendental argument against scepticism. Once again, since this is the concern of chapter six, I will postpone further discussion until then.

¹⁵³ This possibility would require that phenomenology be locally supervenient. On the supposition that Vatbrain has no propositional mental events, I am strongly inclined to deny Vatbrain phenomenological

However, on closer examination it becomes evident that this line of reasoning is misguided. The supposition that a subject could *think* she was gaining knowledge and yet it be illegitimate to attribute propositional mental events to her is clearly incoherent. The only scepticism that could be raised here is a full-blooded scepticism about whether we think at all; and I take it that this very possibility is incoherent. A minimal requirement on a theory of thought is that it be descriptive of whatever it is we do. It is not as if we could settle on a theory, and then discover that whatever it is we do, it's certainly not *thinking*.

Thus, whatever we say about Vatbrain, he provides no counterexample to the validity of instance arguments.

4.6 Conclusion

Introspection and conceptual analysis can together yield knowledge of contingent facts about the external world; but only in so far as those contingent facts are themselves a necessary condition for the very existence of the objects introspected. Of the three following claims,

(PA) For all x , if x believes that she thinks that p , then x thinks that p ;

(SE) x 's thinking that p necessarily depends upon contingent facts about her environment;

(EC) x could not have non-empirical knowledge of contingent facts about her environment;

the one which should be rejected is **(EC)**. If one accepts a new theory about what a concept is, namely semantic externalism, one should expect a new theory about what knowledge of a concept entails. The *Argument from Privileged Access* fails to establish its intended conclusion.

states also. That is, I am sympathetic to the view that phenomenology is broad.

Semantic externalism and privileged access are compatible.
We can have privileged access to the world.¹⁵⁴

¹⁵⁴ The claim that, on the assumption of a certain theory of the mind, if knowledge of mental items is demon-proof, knowledge of the items to which they correspond is demon-proof, has roots at least as far back as John Locke. Thus Locke not only acknowledges that his theory of simple ideas has this consequence, though he would no doubt have expressed it differently, he actually takes it as a point in favour of his theory. He invokes the allegedly unsavoury consequence as providing the solution to a problem about knowledge of the external world. Locke writes,

§ 3. ... How shall the Mind, when it perceives nothing but its own *Ideas*, know that they agree with Things themselves? This, though it seems not want to difficulty, yet, I think there be two sorts of *Ideas*, that, we may be assured, agree with Things.

§ 4. *First*, The first are simple *Ideas*, which since the Mind, as has been shewed, can by no means make to it self, must necessarily be the product of Things operating on the Mind in a natural way, Thus the *Idea* of Whiteness, or Bitterness, ... has all the real conformity it can, or ought to have, with Things without us. And this conformity between our simple *Ideas*, and the existence of Things, is sufficient for real Knowledge. (Locke, 1689 Book IV, Chapter IV, original emphasis)

Singular Thought

5.1 Introduction

In chapter four I argued that instance arguments are a valid means by which a subject can come to have non-empirical knowledge of her environment. It was stressed that instance arguments take one from an instance of a general propositional mental event to knowledge of a general fact about the world.

It is tempting to think that semantic externalism characterises thought as essentially *object-dependent*, and that it is this fact which explains the validity of instance arguments.¹⁵⁵ If this reasoning were correct, it ought to be possible to use analogous instance arguments to infer from any object-dependent thought to the existence of the object upon which that thought depended. A *prima facie* example of an object-dependent thought is a singular thought, or *de re* thought.¹⁵⁶ Such thoughts, broadly speaking, concern individuals, or particular objects.

In this chapter I discuss two opposed theories of singular thought, each of which is defined by its answer to the following question: is the content of a *de re* thought to be understood as object-dependent, or object-independent? To maintain that the content of a *de re* thought is object-dependent is to maintain that the content would not be available to be thought in the absence of the object which the thought concerns. If this view is correct, it will be possible to gain non-empirical knowledge of the existence of the particular object or objects upon which a given *de re* thought depends. If the alternative view is correct, no such non-

¹⁵⁵ For example see Brown (1995), who has claimed that what I have called the *Argument from Privileged Access* works in effect because it can be generalised to apply to all cases of object-dependent thought. See also Noonan (1993).

¹⁵⁶ I will use the terms “singular thought” and “*de re* thought” interchangeably.

empirical knowledge is possible, since the content of a *de re* thought would exist even in the absence of the object or objects the thought in fact concerns.¹⁵⁷

Given the complex nature of the theories I will be discussing, it will be worth while to examine them in some detail before moving on to establish my intended conclusion.

5.2 *De dicto* and *de re* attributions

To give an account of singular thought has been, in the first instance, to provide a criterion of differentiation between the attribution of a *de dicto* thought and the attribution of a *de re* thought.¹⁵⁸ For ease of exposition, I will throughout the following discussion use belief as the exemplar type of thought. At first sight, then, it might be thought that the distinction between the attribution of a *de dicto* belief, and the attribution of a *de re* belief could be drawn by appeal to a surface-level, grammatical distinction; a distinction between belief in a proposition and belief *of* something that it has a given property. To illustrate this, consider the following.

(1) Ortcutt believes the proposition that someone is a spy.

(2) Someone in particular is believed by Ortcutt to be a spy.¹⁵⁹

¹⁵⁷ As will become clear, the debate centres on the notion of “content”. *De re* thoughts are object-dependent according to *both* theories. The crucial issue is whether to admit some form of content even in the absence of a truth-evaluable thought.

¹⁵⁸ On the face of it, there is no reason to suppose that an examination of the nature of the attributions of beliefs should provide any insight into the nature of the beliefs thereby attributed. Consider an analogous approach towards the nature of baldness. There is no obvious reason why a subject would learn anything about the nature of baldness by investigating the nature of baldness-attributions. However, I take it that this was the traditional approach, and that the surface-level grammatical distinction is the traditional one.

¹⁵⁹ The examples are taken from Quine (1956).

According to the grammatical distinction, (1) is taken to relate Ortcutt to a proposition, and is thus the ascription of a *de dicto* belief; (2) is taken to relate Ortcutt to an individual - a *res* - and is thus the ascription of a *de re* belief.

Burge¹⁶⁰, however, has argued that the grammatical distinction does not provide a sufficient condition for drawing the relevant distinction between the attribution of a *de dicto* belief and the attribution of a *de re* belief. This is shown by a proposal about the structure of propositions put forward by Bertrand Russell. Russell held that sentences containing logically proper names express propositions whose components include the individuals thereby named.¹⁶¹ Burge claims that since this notion of proposition was introduced specifically to account for the notion of *de re* knowledge, a statement that says that this sort of proposition is believed should not be regarded as *de dicto* but as *de re*. An example of such a proposition is given in (3).

(3) Ortcutt believes the proposition that this is red.

According to the grammatical distinction, the ascription expressed by (3) is that of a *de dicto* belief, since it relates Ortcutt to a proposition; yet the proposition is such that believing it would be to have a singular belief. Hence, according to Burge, the grammatical distinction can not underpin the relevant distinction between *de dicto* belief and *de re* belief.

Quine's discussion of the topic has led to the relevant distinction being drawn instead in terms of a substitutivity criterion. On this account, the distinctive mark of the attribution of a *de re* belief about a given object is that it is legitimate to substitute *salva veritate* any correct description

¹⁶⁰ Burge (1977).

¹⁶¹ According to Russell, a proposition is an abstract entity with objects and properties in it. See Russell (1912) and (1956). A logically proper name is a genuinely referring expression, one which presupposes the existence of the object named.

of that object. For example, suppose Ortcutt believes *de re* of the woman in the alleyway that she is a spy. In characterising Ortcutt's belief, it would be legitimate to use any correct description of that woman, regardless of whether Ortcutt could describe the woman in that way. Thus, on the supposition that the woman in the alleyway is the prime minister, Ortcutt's belief would be correctly characterised both as the *de re* belief of the woman in the alleyway that she is a spy, and as the *de re* belief of the prime minister that she is a spy. Theoretically, the ascription relates Ortcutt directly to the woman, without attributing to Ortcutt any particular description or conception that he could use to represent her.¹⁶²

If, on the other hand, Ortcutt believes *de dicto* that the woman in the alleyway is a spy, it would be incorrect, according to Quine, to infer that he has the *de dicto* belief that the prime minister is a spy. Even were Ortcutt to be aware of the relevant identity, and hence to believe both propositions, the beliefs are to be regarded as distinct. Essentially, it is possible for a rational subject to take opposing attitudes to the propositions. On this view it is a mark of belief *de dicto* that an ascription of it will not tolerate co-extensional substitution.

Once again, however, Burge maintains that the distinction cannot be adequately drawn in these terms either, since there are ascriptions in which substitutivity fails, yet which are nevertheless *de re*. Consider the following.

(4) Ortcutt believes that the woman in the alleyway is a spy.

The problem Burge sees here is that the expression "the woman in the alleyway" may be "doing double duty at the surface level - both characterising Ortcutt's conception and

¹⁶² It is a notable point from Quine that he assumes that when a subject is related directly to an object, the subject's conceptions of the object are not taken into account at all. As will become clear, this mutual exclusion should be abandoned.

picking out the relevant *res*¹⁶³. Burge concludes that the distinction between *de re* and *de dicto* attributions can not be adequately drawn by differences at the surface level. Essentially, the difficulty is that attributions of a given thought need not reflect the type of thought which is attributed. Hence, any account which tries to draw the distinction between *de dicto* and *de re* beliefs by appeal to a distinction between *de dicto* and *de re* attributions of beliefs will ultimately fail.

5.3 Burge's distinction

The realisation that differences at the surface level are inadequate for the purposes of distinguishing *de re* from *de dicto* attributions leads Burge to provide an account of two distinctions: a distinction between the attribution of *de dicto* thought and the attribution of *de re* thought, as well as a distinction between *de dicto* thought and *de re* thought, where the latter, epistemic distinction grounds the former, semantic distinction. In this way, Burge focuses not only on the attributions of thought, but on the logical status of the thoughts entertained.

Let us look at the semantic distinction between the attribution of a *de dicto* and a *de re* belief first. Burge expresses this distinction in terms of logical form, and represents (1) and (2) above as,

$$(1') B_d(\text{Ortcutt}, \phi(\exists x) \text{Spy}(x) \ddot{U})$$

$$(2') (\exists x)(B_r(\text{Ortcutt}, \langle x \rangle, \phi \text{Spy}(y) \ddot{U}))$$

'B_d' represents that the subject has a *de dicto* belief, and 'B_r' that the subject has a *de re* belief. The pointed brackets contain a bound variable representing the object presented, and indicate that it is the object itself which is of importance, rather than any description of that object; substitution of co-

¹⁶³ Burge (1977) p. 341. For essentially the same point see Castañeda (1967), and Loar (1972).

extensional descriptions within slanted brackets is therefore legitimate. The corner quotes, according to Burge, “may be regarded as a convenience for denoting the proposition, or component of proposition, expressed by the symbols they enclose”¹⁶⁴. No singular term will appear inside the corner quotes, and the predicates which do appear will not be substitutable by co-extensive predicates.

Burge maintains these formulations make the relevant distinction explicit: ‘Bd’ relates the subject to a complete proposition, expressed by a closed sentence; ‘Br’ relates the subject in part to an incomplete proposition, expressed by an open sentence, and in part to a *res*. He writes,

Purely de dicto attributions make reference to complete propositions - entities whose truth or falsity is determined without being relative to an application or interpretation in a particular context. De re locutions are about predication broadly conceived. They describe a relation between open sentences (or what they express) and objects. (Burge, 1977 p. 343, original emphasis)

According to Burge, this way of drawing the distinction captures the intuition behind both the substitutivity criterion and the surface-level grammatical distinction.

Examining the function and placement of the corner quotes makes it clear why Burge claims that Quine’s substitutivity criterion will be unable to distinguish the *de re* from the *de dicto*, for it is no longer the individuating criterion of the *de dicto* that it will not tolerate substitution of co-extensional terms. No expression inside corner quotes will tolerate substitution of co-extensional terms, so according to Burge’s distinction attributions of belief *de re* will not tolerate substitution of co-extensional terms *simpliciter* either. It is clear that substitution is legitimate only in the first and second argument places. It is not legitimate anywhere in the third argument place, which specifies the open sentence,

¹⁶⁴ Burge (1977) p. 341.

since this is characterised in terms specific to the way the subject is thinking about the supposed referent.

This can be illustrated by examining the following examples of the logical form of beliefs which could make (4) true.

(4') Br (Ortcutt, <the woman in the alleyway>, ϕ Spy (y)Ü)

(4'') Br (Ortcutt, <the woman in the alleyway>, ϕ Spy (y) & Woman (y) & In Alleyway (y)Ü)

The second, and not the first, attributes the notion of a woman in the alleyway to Ortcutt.¹⁶⁵

To reiterate, then, the semantic distinction is this. A *de dicto* attribution of a belief relates the subject to a complete, or completely expressed, proposition; a *de re* attribution of a belief relates the subject to a *res* and an incomplete proposition. Underlying this semantic distinction is an epistemic distinction. The epistemic distinction is between *de dicto* beliefs that are fully conceptualised, and *de re* beliefs, "whose correct ascription places the believer in an appropriate nonconceptual, contextual relation to objects the belief is about"¹⁶⁶. This is not to say that the non-conceptual relation precludes any conceptual relation between the subject and the object, as is explicit in (4''). While there is no constraint on the involvement of concepts available to the subject entertaining a *de re* belief, there is a requirement that the relation between the subject and the object not be *merely* conceptual. A further relation must hold, a description of which is external to the content of the belief, and which need not correctly be described in terms of concepts available to the subject. The connection between the subject and the object is a brute causal connection. The subject need not, however, be denied the concepts required to express the relevant connection.¹⁶⁷

¹⁶⁵ Burge assumes (4'') to be a fairly ordinary reading for cases in which Ortcutt sees the woman.

¹⁶⁶ Burge (1977) p. 346.

The semantic and the epistemic distinction are of course interrelated. If a subject satisfies the epistemic criterion for having a *de re* belief, any correct ascription of that belief will satisfy the semantic criterion for the attribution of a *de re* belief. *Mutatis mutandis* for a *de dicto* belief and its correct ascription. Thus, while all beliefs will be partly characterisable in terms of the subject's concepts, it is, according to Burge, the distinctive mark of a *de dicto* belief that it be characterisable *purely* in such terms. He writes,

A correct ascription of a *de dicto* belief identifies it purely by reference to a "content" all of whose semantically relevant components characterise elements in the believer's conceptual repertoire. (Burge, 1977 p. 346)

Similarly, according to Burge, if a belief is essentially *de re*, that is in virtue of the fact that a context involving the *res* itself enters into determining how the belief can be correctly ascribed.

At first sight, the following could be regarded as a threat to Burge's characterisation. Suppose Ralph, in conversation with Ernie, sincerely asserts, "The richest man in Britain is mean", not knowing who the richest man in Britain is. Suppose later that Ernie, an acquaintance of the richest man in Britain, reports Ralph's belief to him by saying, "Ralph believes you are mean". *Prima facie* what we have here is a case in which the belief ascription relates Ralph both to an open sentence and to the richest man, thus fulfilling the semantical characterisation of *de re* belief, even though Ralph's epistemic state depends completely on concepts in his repertoire, thus fulfilling the epistemic characterisation of *de dicto* belief. The semantic and epistemic characterisations

¹⁶⁷ Burge (1977) p. 361. Although Burge admits that there is no adequate general explication of this contextual, not purely conceptual relation, he takes perception as the paradigm, with further examples of such possible relations being those involved in memory, introspective beliefs, certain historical beliefs, beliefs about the future, and perhaps beliefs in pure mathematics.

provided by Burge thus appear to diverge. However, the characterisations can be seen to diverge only in so far as the belief-ascription is true. The question, then, is whether or not the belief ascription *is* true: and I take it that strictly speaking it is not. What we have here is not a correct ascription of a *de re* belief, but rather a *de re* ascription of a *de dicto* belief. All the example can be taken to show is that the attribution of a belief may come apart from the type of belief which is taken to make the attribution true; and there is no reason why Burge should not accept this.¹⁶⁸

Burge, then, holds a dual-component theory of *de re* thought. A *de re* thought is composed of a reference-independent content and a context.¹⁶⁹ The content is the “mentalist” part of the thought, expressed by an open sentence. Which object the content applies to may vary with context. Orcutt could have the same thought-content were he to think of Ralph as a spy, or of Ralph’s twin as a spy. The thoughts would nevertheless have different truth-conditions. Without the contextual component, then, there is no thought. The first would be true if and only if Ralph were a spy, the second if and only if Ralph’s twin were a spy.

Burge maintains that this *de re* content cannot be reduced to a complete *de dicto* thought. Any incomplete *de dicto*

¹⁶⁸ I take it that the distinction I invoke between an ascription being literally true and an ascription being understood and accepted even though literally false lies behind Burge’s proposal that, in the example given, the term “you” would be anaphoric, acting as a “pronoun of laziness” for the description “the richest man in Britain”. If this were correct, the ascription would then relate Ralph to a closed sentence, thus fulfilling the semantic characterisation of *de dicto* belief. Consequently, the semantic and epistemic characterisations would remain in step. (Burge, 1977 p. 346). The treatment of “you” as anaphoric could, however, be criticised, since if it were anaphoric, it ought to be a condition on understanding the attribution that the richest man in Britain know the relevant description under which he was referred to by Ralph, which does not seem right.

¹⁶⁹ For alternative ideas about reference-independent content, see Fodor (1987), Searle (1983) chapter 8, and Loar (1988).

content will by itself be insufficient to pick out the actual referent of the *de re* belief. Since Burge places no restriction on the number or type of concepts which can enter into the content of a *de re* thought, he must presumably allow that it be possible for the content of a *de re* thought in fact to pick out the referent of that *de re* thought. However, in any given case in which the concepts featuring in the content of a *de re* thought do in fact uniquely identify the object of the *de re* thought, it will not be in virtue of the fact that the referent satisfies the relevant description in the thought-content that the thought concerns the object it does, but in virtue of the non-conceptual relation which holds between the subject and the object. Thus Burge expresses his commitment to a non-reductivism with respect to *de re* thought. Further, of the two types of thought, it is the *de re* which Burge maintains is the more fundamental. It is *de re* thoughts which he maintains are essential for the use and understanding of language.

It might be thought that the content invoked in a dual-component theory is naturally individualistic. However, Burge's non-reductive fundamentalism about *de re* thought ties in neatly with his semantic externalism. A dual-component theory will be distinguished as individualistic or anti-individualistic depending on the status of the concepts involved in the content of the thought, in the propositional fragment. Since the concepts which constitute the content of the thought will, on Burge's theory, be externally individuated, his dual-component theory of *de re* thought is anti-individualistic. While Ortcutt could have the same type of content were he to think of Ralph as a spy, or of Ralph's twin as a spy, this would depend on his remaining in the same linguistic environment, such that the term "spy" has the same meaning in each case. The content of his thought is not locally supervenient, but externally determined.

5.4 Object-dependent thoughts

There are two defining characteristics of dual-component theories such as Burge's. First is the claim that the elements

of the content of the thought, the incomplete propositional fragment, may be insufficient to pick out the object the thought concerns. Second is the claim that the content of a *de re* thought is available to be thought whether the object which it purportedly concerns exists or not. It is McDowell's contention that this view is adopted as a result of the mistaken belief that,

... a Fregean philosophy of language and thought can represent an utterance, or a propositional attitude, as being about an object only by crediting it with a content that determines the object by specification, or at least in such a way that the content is available to be thought or expressed whether the object exists or not. (McDowell, 1984 p. 98)

The work of both Evans and McDowell is strongly opposed to such a belief.¹⁷⁰ Their concern is to demonstrate that a Fregean philosophy of language and thought can accommodate a notion of *de re* thought, according to which the very *content* of a *de re* thought is object-dependent. On this view, the existence of a *de re* content depends essentially upon the existence of the object the content concerns. If the supposed object does not exist, neither does the supposed content. Crucially, the SSTT admits of no distinction between content and thought; to say that a thought is object-dependent is to say that its *content* is object-dependent. The dual-component theory, on the other hand, divorces *content* from *thought*, and hence allows that the content of a *de re* thought is not object-dependent, even if the thought itself is regarded as object-dependent.

¹⁷⁰ My concern here is not directly with the motivational contention of McDowell's, but rather with the claim that a *de re* thought is not such that it is available to be thought whether the object exists or not. However, both Evans and McDowell are largely motivated by an attempt to account for certain types of thought within a Fregean framework. See McDowell (1977) for a Fregean account of proper names. See Evans (1981) for a Fregean account of demonstratives.

The idea that the content of a thought can be object-dependent is evident in Russell's account of singular propositions. Russell held that sentences containing genuinely referring expressions express propositions whose components include the individuals thereby named. From this it follows that a singular proposition is only available to be thought in so far as the object referred to exists. Singular propositions are object-dependent.

Since Russell maintained that "[e]very proposition which we can understand must be composed wholly of constituents with which we are acquainted"¹⁷¹, where a subject is acquainted with something if it is "immediately known to [her] just as [it is]"¹⁷², he was naturally wedded to the claim that we can understand a proposition only if we cannot be mistaken about which object it is that the proposition concerns. The requirement that one have such demon-proof knowledge of the object concerned led Russell to restrict the possible constituents of singular propositions to universals, sense-data and, tentatively, the self.¹⁷³

Evans and McDowell propose to accept the Russellian notion of an object-dependent thought, and yet reject the restrictive nature of Russell's theory of singular propositions. That is, they allow acquaintance also with objects knowledge of which is fallible; and hence propose that object-dependent thoughts could also concern physical objects in the world.¹⁷⁴ Acquaintance is not undermined by fallibility, as Russell believed. The question is whether object-dependent thoughts can plausibly be recognized outside Russell's restriction.

It might be suggested that the most natural way to adopt a theory of object-dependent thought without also adopting Russell's restriction is simply to maintain that the physical objects which are now supposed to be the proper concern of object-dependent thoughts can themselves be the constituents of the relevant thoughts. But there is a problem

¹⁷¹ Russell (1912) p. 32.

¹⁷² Russell (1912) p. 25.

¹⁷³ See Russell (1912) especially chapter 5.

¹⁷⁴ See Evans (1982), and McDowell (1986).

with this suggestion; how is one to account for the fact that a subject could rationally take different cognitive attitudes towards a thought with the very same constituents?

For if we say that the object itself is actually *in* the thought - like a pea in a pod - then we seem to have made no provision for a distinction between different thoughts in which the same property is ascribed to the same object. (Evans, 1982 p. 82)

If a thought is conceived of as an ordered pair consisting of an object and a property, there can be no way to distinguish thoughts containing the same constituents, and therefore no way in which to account for someone's having a differing cognitive attitude towards such an ordered pair. There are possible situations for which the ordered pair conception will not be adequate. Take the following example, given by Evans.

Suppose a person can see two views of what is in fact a very long ship, through two windows in the room in which he is sitting. He may be prepared to accept 'That ship was built in Japan' (pointing through one window), but not prepared to accept 'That ship was built in Japan' (pointing through the other window). (Evans, 1982 p. 84)

The problem comes when we try to describe the situation on the ordered pair conception of Russellian thought. There is one ordered pair – <the ship, the property of having been built in Japan> – to which the person both does and does not stand in the belief-relation. Not only does this fail to capture the state of mind of the person involved, but it also appears to be contradictory.¹⁷⁵

¹⁷⁵ It might be thought that Burge's account faces the same difficulty in distinguishing pairs of thoughts which ascribe the same property to the same object. Taking Evans's example, Burge's account yields the following.

$$\begin{aligned} & Br (S, \langle \text{that ship} \rangle, \text{Built in Japan} (y)) \& \\ \neg & Br (S, \langle \text{that ship} \rangle, \text{Built in Japan} (y)) \end{aligned}$$

There is perhaps no problem here for Russell. On his theory the objects contained in singular propositions are restricted to things which were conceived to be so fleeting and insubstantial that it seemed unintelligible to suppose a person might identify the same one twice without knowing that it was the same. This way of making the account consistent is unavailable to those who, like Evans and McDowell, wish to apply Russellian object-dependent thoughts to objects in the world.

The account which Evans and McDowell propose, then, is a combination of the following two compelling insights. First,

... the idea that there are things which we say and believe whose content cannot be faithfully represented without the reporter himself making a reference to an object in the world which those utterances and beliefs concern - so that, where there is no such object, there would be no such content available to be faithfully represented ...
(Evans, 1982 p. 82)

and second, the idea that there needs to be a way in which to distinguish pairs of thoughts which ascribe the same property to the same object, and that only Fregean sense will suffice for this purpose. The result is an account of *de re* thoughts as object-dependent, but as containing *senses*, and not objects, as constituents. On this account, unlike Russell's, a thought is not determined by the sheer identity of its referent. There are, in McDowell's terms, *de re* senses.

However, Burge has a way of overcoming the difficulty by an appeal to the necessary contextual relation between the subject and the object for any given *de re* thought. There will be no inconsistency if a different contextual relation relates the subject to the ship in each case. However, this solution leads to a modification of the logical form of *de re* beliefs. Rather than the above, we should have,

$$\text{Br}(S, \mathbf{R}, \langle \text{that ship} \rangle, \text{Built in Japan } (y) \ddot{U}) \ \& \\ \neg \text{Br}(S, \mathbf{R}', \langle \text{that ship} \rangle, \text{Built in Japan } (y) \ddot{U})$$

For a discussion of whether the non-conceptual relation should be incorporated into the content of the thought, and of the difficulties involved, see Bach (1987).

According to the dual-component theory, it will be remembered, the content of a thought is available to the subject whether or not there is an object which it concerns, and, consequently, the object a belief is about will not necessarily be determined by elements of the content of the thought, since features of the proposition or content will in certain cases not be sufficient to pick out the relevant object. In stark contrast, according to the theory of object-dependent thought proposed, not only is the content of a thought available to be thought only on the condition that the object or objects which it concerns exist, but the very constituents of the thought will therefore themselves always be capable of securing reference. This theory of object-dependent thoughts has been termed the “Strong Singular Thought Theory” (hereafter “SSTT”).¹⁷⁶

There is a certain unease felt about *de re* senses, which Segal expresses by means of the following example.

If at some time, *t*, you think *it's bedtime now*, then you have a thought about *t*. You represent *t* as now (from your perspective at the time) and as bedtime. What makes it the case that it is *t*, in particular, that your thought is about? Is it because you had a thought of a certain kind at *t*, so the referent was determined in part by the nature of the thought, in part by the fact that you had the thought at *t*? Or is it because you had a thought of a certain kind, a thought that contained a special purpose sense, a sense that differs in kind from any later or earlier sense, a sense part of whose essence it is to refer to *t* and no other time? Which of these sounds strange? (Segal, 1989, p. 53)

The passage quoted is clearly rhetorical; and the idea that the proposition expressed by an indexical or demonstrative thought is determined by a content, or character, together with a context external to the cognitive world of the subject, is perhaps the entrenched view.¹⁷⁷ But rhetoric is not

¹⁷⁶ This terminology comes from Blackburn (1984) chapter 9.

¹⁷⁷ See Kaplan (1980), and Perry (1977).

argument. The very point at issue here is whether there *is* a conception of sense which determines reference in the very strong way which Evans and McDowell advocate.¹⁷⁸

Certainly, the notion of character can look very appealing. One purpose which it serves is to allow for the intuition that subjects who think, for example, of themselves as being hot by thinking respectively *I am hot*, thereby have some thought content in common. Classifying thoughts by their content is unavailable to the SST theorist. Segal writes,

If thoughts are type-identified by their contents and singular thoughts about different objects have different contents, then such singular thoughts must be type distinct. SSTT, in rejecting reference-independent content, entails that there is no taxonomy of thoughts that identifies them by their contents and groups together singular thoughts about different objects. (Segal, 1988, p. 39)

Certainly, some form of systematic taxonomy is required; but there is no pre-theoretical reason to suppose that content is the best, or the only, way to taxonomise thoughts. McDowell suggests the following.

Particular *de re* senses, each specific to its *res*, can be grouped into sorts. Different *de re* sense (modes of presentation) can present their different *res* in the same sort of way: for instance, by exploiting their perceptual presence. And the univocity of a context-sensitive expression can be registered by associating it with a single sort of *de re* sense. (McDowell, 1984 p. 103)

¹⁷⁸ In his fascinating paper on demonstratives, Evans proposes that a sense be thought of as a *way of thinking* about an object. This then allows for so-called “dynamic thoughts”, where a subject keeps track of a day, for instance, as it recedes into the past, by thinking of it respectively as today, yesterday, two days ago, and so on. See Evans (1981). The paper, which builds on McDowell’s (1977), is intended as a response to Perry (1977), where it is claimed that a Fregean framework cannot accommodate demonstrative thoughts.

Whether McDowell's suggestion is plausible or not, to assume that thoughts should be grouped together by only their *content* is to beg the question against the SSTT.¹⁷⁹

5.5 Against Burge

The contrast between the two opposed accounts of singular thought can be highlighted by examining the major objections raised against each. It will therefore be worth while spending time on these. I will start in this section by discussing two criticisms of the dual-component theory. Criticisms of the SSTT tend to focus on its treatment of a deluded subject, a subject who is mistaken in her belief that she is presented with an object, and correspondingly mistaken in her belief that she is entertaining a singular thought. This will largely be the concern of the three sections which follow.

The first objection to the dual-component theorist, then, is as follows. As mentioned in section **5.4** above, it is McDowell's contention that the dual-component theory is adopted as a result of the mistaken belief that a Fregean philosophy of thought and language could not accommodate singular thought in any other way. However, McDowell objects to the separation of context from content which characterises the dual-component theory. The SSTT demonstrates that there is no need for such a forced separation, since singular thought can be accommodated within a Fregean framework which does not require any such separation. He writes,

Given that conceptual content is made up of *means* of representation in thought, a belief's being fully conceptualised can mean only that it has a fully propositional content exhausted by some collection of thought symbols; and it would follow that there is no room for contextual factors to contribute to determining how such a belief may be correctly ascribed. This makes Burge's picture of the relation between conceptual

¹⁷⁹ For arguments against the idea that a psychological taxonomy should taxonomise by content, see Stich (1983), and McGinn (1982).

content and context obligatory. (McDowell, 1984 p. 101)

McDowell maintains that the separation is plausible only by a conflation between the two readings the term “concept” allows; between what is expressed and what does the expressing. If there is a possible account according to which what is expressed is context-dependent, while the thought content is nevertheless fully conceptual, then this account, according to McDowell, is surely to be preferred.

He argues that given Frege’s account of sense, and the assumed equation between “conceptual content” and what can be “completely expressed”, there is no need to accept that contextual factors are extraneous to the content-determining powers of a conceptual repertoire; and no need to accept that what is expressed by a context-sensitive utterance cannot be partly determined by the context in which it is made. In effect, McDowell maintains that *de re* Fregean senses yield thoughts which are both *de re* and yet part of the subject’s cognitive world.

... for all that Burge shows, a conceptual repertoire can include the ability to think of objects under modes of presentation whose functioning depends essentially (say) on the perceived presence of the objects. Such *de re* modes of presentation would be parts or aspects of content, not vehicles for it; no means of mental representation could determine the content in question itself, without benefit of context, but that does not establish any good sense in which the content is not fully conceptualized. (McDowell, 1984 p. 102)

Both theories accept the importance of the role of context in determining thought. McDowell’s contention is that this importance can be acknowledged without siphoning off the contextual factors from the very content of the thought in question.

The second criticism McDowell raises for the dual-component theory arises as a result of this separation of

thought content and context. This objection concerns the supposed relation between the incomplete propositional content and the *res*. McDowell writes,

How does the relational expression relate the *res* to the propositional fragment? In the state of affairs that the attribution represents, the propositional fragment should figure as somehow tied to the *res* by a predicational tie; can this intuitive requirement be met? If as in the Fregean position, the *de re* attribution is conceived as true in virtue of the truth of a *de dicto* attribution, this question holds no terrors: in the underlying *de dicto* the predicational tie will be explicitly expressed. But if, as in Burge's framework, the *de re* attribution is conceived as "barely true", the belief relation has to secure the presence of the predicational tie all on its own; and it is quite unclear that it can be explained so as to carry the weight. (McDowell, 1984 p. 107)

It will be remembered that Burge holds that *de re* thoughts are the fundamental type of thought: the description of a *de re* thought will not always be reducible to the description of a *de dicto* thought, and neither will there be, for every *de re* thought, an underlying *de dicto* thought. The problem McDowell sees is that in the absence of an underlying *de dicto* thought, there is no underlying explicit predicational tie for the given *de re* thought, and hence nothing to hold the elements of the thought together.

Segal has elaborated Burge's account, and provided a response to this challenge.¹⁸⁰ According to Segal, the predicational tie is taken to be achieved by a reference relation between the variable in the propositional fragment and the object, together with the logical form of an open sentence. The open sentence will then be true if and only if the referent of the variable satisfies the predicate. On this account, the predicational tie is effected in the same way for *de re* as for *de dicto* beliefs. The important difference, claims Segal, is that for the former the relation must be supplied by

¹⁸⁰ Segal (1989) pp. 50-2.

the context, whereas for the latter it is supplied entirely by elements of the thought itself.

However, this is not so much an answer to the problem, as a restatement of the theory under consideration. It is precisely the nature of the predicational tie between the content of the thought and the object external to that content which is in question. Something has to point the propositional fragment toward the object concerned, and concepts will not always be sufficient to do the job.

5.6 Deluded subjects

As mentioned above, criticisms of the SSTT tend to focus on its treatment of a deluded subject; a subject who is mistaken in her belief that she is presented with an object. It is the purpose of this section to present the treatment of deluded subjects by both the SSTT and the dual-component theory.

Let us start with a look at the dual-component theory. According to this theory, the object of any given singular thought is not part of its content, rather it is an external object which, together with the content constitutes a thought. This appeal to such reference-independent content allows that thoughts be treated as psychologically the same if and only if they have the same content, independently of whether or not there is an object with which the content can combine. The criterion for being in the same psychological state can be given independently of how things are in the environment. On the Burgean conception of *de re* belief, then, if the context fails to supply an appropriate object, the subject is still alleged to have, because of the descriptive content of the open sentence, something in her head sufficient to explain her actions, even if this does not amount to a truth-evaluable thought. Deluded subjects are regarded as being in some kind of intentional state, since the content is available to make up a thought when coupled with an object, even in the immediate absence of an appropriate object.

Evans has argued that any such so-called “content”, since it must be given in schematic terms, ought not to be regarded as representational at all. He writes,

It is of the essence of a representational state that it be capable of assessment as true or as false. If a state is a representational state, it represents something other than itself as being thus and so. But a schema is not assessable as true or false, nor is any state whose ‘content’ can be given only in schematic terms So, since [a *de re* thought] has a ‘content’ which is strictly specifiable only in schematic terms, the [*de re* belief] is not assessable as true or false; hence it is not a representational state. (Evans, 1982 p. 202)

The problem Evans sees is not that a representation must be assessable either as true or as false, since the possibility that certain representations are simply vague should not be ruled out. Rather, the problem is that a propositional fragment and an object do not yet make up anything which is truth-evaluable.¹⁸¹

If *de re* thoughts involve object-dependent senses, on the other hand, it follows that there can be no thought at all in the absence of an appropriate object. McDowell writes,

Given a context, a *de re* sense may determine a *de re* sense ... , or else it may determine nothing. And in the latter sort of case, according to this way of thinking, there can only be a gap - an absence - at ... the relevant place in the mind - the place where, given that the sort of *de re* sense in question appears to be instantiated, there appears to be a specific *de re* sense. (McDowell, 1984 p. 103)

This entails an essentially disjunctive conception of the state a subject is in when she believes herself to be entertaining a singular thought. If she is correct, the subject has direct

¹⁸¹ The objection posed by Evans can be seen as another form of the objection raised by McDowell that there is no predicational tie between the content (of the thought) and the object (of the thought).

access to the world. If she is incorrect, the subject has merely apparent access to the world, and fails to think a singular thought at all. Following Evans, I will refer to such a deluded subject as entertaining a “mock thought”.¹⁸² Crucially, there are no such things as empty singular thoughts.

Segal¹⁸³ has argued that the attribution of empty singular thoughts to deluded subjects allows us to understand how things seem from their point of view, and that this sort of understanding of how things seem to the deluded subject is unavailable to proponents of the SSTT. Consider the following example. Little Johnny is drawing up a list of toys he would like for Christmas. When asked what he is doing, little Johnny says, “I’m writing a letter to Santa Claus”. The dual-component theorist can explain little Johnny’s actions and utterances by appeal to a series of reference-independent psychological states. If, however, the senses of proper names are object-dependent, little Johnny can have no such psychological states, so the argument goes, and *a fortiori* no such psychological states which could explain his actions. According to Segal, the SST theorist cannot therefore accept our description of little Johnny’s state of mind.

McDowell claims that this is no real objection, since in such cases we are simply “playing along with the deluded subject - putting things his way”¹⁸⁴. Similarly, Evans claims that in

¹⁸² See Evans (1982) p. 82. To avoid the “mock thought” analysis, it might be tempting to analyse the singular thoughts of deluded subjects as follows. In the event of their being no object to which the subject intends to refer, a particular location in space would be regarded as the object of the *de re* thought. For instance, Orcutt could be taken to be predicating spyhood of that space which would be occupied if a woman were present. However, Evans has argued against the reduction of “the demonstrative identification of spatially located objects” to “the identification of positions in egocentric space” for the reason that it is possible to identify an object perceptually even when it is moving too rapidly for us to be able to assign it a precise location in space (Evans, 1982).

¹⁸³ Segal (1989).

¹⁸⁴ McDowell (1977) p. 154.

attributing a belief to a deluded subject we are “only sympathetically approximating to his state of mind”¹⁸⁵. Unsatisfied with this response, Segal writes that “understanding someone *is* putting things their way. Put them any other way and you miss the point”¹⁸⁶.

Evans offers the following thoughts to explain, and dismiss, the attractiveness of the view that a deluded subject must have some thought before her mind.

To hallucinate is precisely to be in a condition in which it seems to one as though one is confronting something. So of course it will seem right to the hallucinator to say that he is actually confronting something; the situation is very like one in which he *is* confronting something. ...

If after it has been acknowledged on all sides that it seems to the hallucinator that he is confronting something ... one says that it seems reasonable to the generality of mankind to suppose that the hallucinator is actually confronting something, ... then one is attempting to double-count the fact that has already been acknowledged.

Now it is essentially the same consideration that underpins the view that if it seems to a person that he is thinking something, then there must be some thought before his mind. Perhaps the same explanation of the attraction this consideration has for us will serve to undermine it in this application as well. (Evans, 1982 p. 200)¹⁸⁷

The passage quoted from Evans draws upon an instructive parallel between the view that particular uses of singular terms are credited with senses that determine objects in such a way that the senses are expressible whether the objects exist or not, and representative realism in the theory of perception, where representative realism postulates items

¹⁸⁵ Evans (1982) p. 199.

¹⁸⁶ Segal (1989) p. 57, emphasis added.

¹⁸⁷ It would seem that criticisms such as Segal’s rest on a failure to distinguish between what Evans calls “situation-specifying attitudes” and “content-giving specifications of attitudes”. See Evans (1982) p. 199.

that are before the mind whether objects are perceived or not.¹⁸⁸

Given that it is possible for it to appear to a subject as if she is perceiving, or thinking about, a particular object when the object is presented to her, and to have a phenomenally indistinguishable experience when there is no such object, it is tempting to assume that there is something which the subject has before her mind in both cases. McDowell and Evans both reject this line of inference. According to them, that things are the same for the subject does nothing to establish that worldly circumstances are only externally related to experiences: “to think otherwise is to fall into a fully Cartesian conception of the [subjective]”¹⁸⁹. Both in perception, and in the case of singular thought, the legitimacy of the subjective is consistent with an essentially disjunctive conception of the state a subject is in.

McDowell claims that such dual-component theories of the mind fail to capture the nature of demonstrative thought, because the relevant object is “before the mind only by proxy”¹⁹⁰. The force of *de re* senses is that they allow the subject to be in direct contact with the world, while respecting rational explanation by explaining how the subject can have differing cognitive attitudes to the same object and property.

5.7 The Two List Argument

The treatment of empty cases by the SSTT has, however, led to the following powerful line of argument. Noonan has argued that reference to object-dependent thoughts is never required for the purposes of psychological explanation, and hence that either there are no singular thoughts reference to which is essential to the adequate psychological explanation

¹⁸⁸ The corresponding analogy is between object-dependent singular thought and the disjunctive conception of experience. For an interesting defence of the disjunctive conception of experience see McDowell (1982).

¹⁸⁹ McDowell (1986) p. 157.

¹⁹⁰ McDowell (1986) pp. 292-3.

of action, or those singular thoughts reference to which is essential are not object-dependent.¹⁹¹

Whenever an action is directed towards a concrete, contingently existing object, other than its agent, in the sense that it is intentional under a description in which there occurs a singular term denoting that object, then an adequate psychological explanation of it is available under a (possibly distinct) description in which occurs a term denoting that object; *and in this explanation the only psychological states of the agent referred to are ones which would also be present in a counterfactual situation in which the object did not exist.* (Noonan, 1986 p. 68-9, emphasis added)

Consider the following example.¹⁹² While walking in his garden, Ralph spies the cat he believes to have killed his beloved canary. Angry Ralph lashes out and kicks the cat. If object-dependent thoughts are essential for the purposes of psychological explanation, the explanation of Ralph's action must invoke his object-dependent thoughts about the cat. It is these object-dependent thoughts which, at least in part, explain his action.

Now imagine a counterfactual situation exactly like the actual situation except for the fact that there is no cat: Counterfactual Ralph is subject to a hallucination. Since things seem the same to Counterfactual Ralph as they do to Ralph, we can assume that Counterfactual Ralph will move in the very same way as Ralph in fact moves. That is, Counterfactual Ralph will lash out at what he takes to be the cat in question.

This is where the challenge to the SST theorist comes into play. The challenge is to explain Counterfactual Ralph's behaviour. Noonan writes,

¹⁹¹ Noonan (1986), (1991) and (1993). Essentially the same argument can be found in Segal (1989).

¹⁹² The example is in essence taken from Noonan (1993).

The [proponent of object-dependent thoughts] is thus faced with a dilemma: he must either deny that the behaviour of [Ralph] in the hallucinatory situation is rationally explicable by reference to his contentful psychological states, or he must acknowledge that reference to a proper subset, X, of the thought contents available to [Ralph] suffices to explain [Counterfactual Ralph's] actions. (Noonan, 1993 p. 286)

The first option is unattractive, since, on the face of it at least, Counterfactual Ralph's behaviour does make sense.¹⁹³ The second option, however, is assumed to be equally unpalatable. The argument for this runs as follows. Counterfactual Ralph has a set of beliefs and desires that constitutes a sufficient reason for him to lash out. This is what Noonan refers to as "X" in the passage quoted above. Since Ralph is Counterfactual Ralph's twin, Ralph has this set of beliefs and desires, and possibly some extra ones, namely the object-dependent thoughts he was originally assumed to have. X is a proper subset of Ralph's thoughts. But X is sufficient reason for a subject to lash out - which is just what Ralph did. Hence, X is sufficient to explain Ralph's behaviour as well as Counterfactual Ralph's. To endorse the second option, so the argument goes, is to accept that object-dependent thoughts are *redundant* in the psychological explanation of action.

This line of reasoning is, however, fundamentally flawed. It can be agreed on all sides that Counterfactual Ralph has a set, X, of beliefs and desires that constitute a sufficient reason for him to lash out. Noonan goes on to claim that since Ralph is Counterfactual Ralph's twin, Ralph also has this set of beliefs and desires: X is a proper subset of Ralph's thoughts. But this is where the SST theorist should object: the SST

¹⁹³ I will leave aside the issue of whether counterfactual Ralph does act rationally. Even if one were to adopt an externalist account of rationality, and hence deny that counterfactual Ralph acts rationally, one could still accept that his behaviour has an explanation. For an externalist account of action according to which the actions of agents are rational only within a given context, see Millikan (1993).

theorist should simply reject the claim that Ralph has X. According to the SSTT, Ralph, since he is not hallucinating, *does not have X*. The SSTT is essentially a disjunctivist theory with respect to psychological states. The defining characteristic of such a theory is precisely that there will be no psychological state which subjects have in common simply in virtue of things seeming the same to them. To maintain that because the set X is sufficient to move Counterfactual Ralph to act, X is sufficient to move any physical duplicate of Counterfactual Ralph to act is already to assume the falsity of the SSTT.¹⁹⁴

¹⁹⁴ There is an important description under which the actions of Ralph and Counterfactual Ralph differ; the former kicks a cat, while the latter lashes out into thin air. Noonan acknowledges this difference, and goes on to explain it as follows.

What makes [Ralph's] action a kicking of a cat then, is simply: the presence of a cat. (Noonan, 1995 p. 287)

The explanation offered by Noonan of the difference between Ralph's action and Counterfactual Ralph's action relationally described is far from satisfactory. This can be illustrated by the following example. Consider a third situation in which Ralph is subject not to a standard hallucination, but to a veridical hallucination. That is, suppose that at the very place at which he hallucinates the presence of a cat, there is in fact a cat. Call him Veridical Ralph. When Veridical Ralph lashes out, he kicks a cat. Thus, according to Noonan's account, Veridical Ralph's action, relationally described, is just the same as Ralph's action, relationally described, where both actions differ from the action of Counterfactual Ralph. But surely this taxonomises actions in the wrong way. Veridical hallucinations, on Noonan's account, yield just the same actions, relationally described, as genuine perceptions; whereas ordinary hallucinations and veridical hallucinations yield different actions. But surely ordinary hallucinations and veridical hallucinations should be classified together for the purposes of psychological explanation. I do not wish to argue that a dual-component theory cannot be modified so as to alter the groupings of the actions relationally described. One plausible way to do this would be to add in a causal constraint. What makes Ralph's action a kicking of a cat would then be: *the presence of a cat*

I have argued that the SSTT can withstand the *Two List Argument*, since a crucial assumption should be rejected by the SSTT. However, the SST theorist does owe some explanatory account of the actions of deluded subjects. To this extent the *Two List Argument* does have some force. It might be thought that one could make sense of the behaviour of a deluded subject by attributing her a second-order belief. On this account, a deluded subject believes herself to have an object-dependent singular thought, and it is this second-order belief which explains her behaviour, which makes her action intelligible. Note that the second-order belief cannot itself be *de re*. For a second-order belief to be *de re* it would have to take as its object a first-order belief. But if a subject is entertaining a “mock” thought, there is no appropriate first-order belief for the second order belief to take as its object. There would then be no second-order belief, rather a further mock thought. Such a second-order mock thought can be invoked in the psychological explanation of a subject’s actions no more than the first-order mock thought it was supposed to replace.

The second-order belief, in order to do the explanatory work required, would therefore have to be *de dicto*. However, an appeal to second-order *de dicto* beliefs as a way to rationalise the actions of deluded subjects gives rise to the following worry. It is at least contentious that second-order beliefs provide the appropriate explanans for a subject’s actions. Since second-order beliefs tell us how we conceive of our own minds, and not how we conceive of the external world, why should a belief that a subject has a *de re* belief of a certain type be sufficient to move her to act?¹⁹⁵

which caused the kicking in some non-deviant way. My claim is merely that Noonan’s account, as it stands, will not suffice.

¹⁹⁵ Note that a version of the *Two List Argument* could be run here also. Once it is conceded that the second-order belief will suffice to explain the actions of the deluded subject, it would be claimed, there would be no need to invoke any object-dependent thoughts in order to give a rational explanation of the non-deluded subject. As we have seen, this argument is fallacious.

The following might be a more plausible option. In the absence of a *de re* belief, when a subject is deluded, what explains her action is a general, *de dicto* belief. In Counterfactual Ralph's case, then, what would explain his lashing out would be something like the general belief that there is a cat in front of him: it would even be possible to attribute Counterfactual Ralph with a *de dicto* belief which uniquely identifies the cat he believes to be in front of him, on the assumption that there is such a cat.

However the SSTT is to account for the actions of deluded subjects, it will give a distinct account of the actions of genuine subjects. No unified account will be forthcoming.

5.8 Instance arguments

Having come to an understanding of the two opposed theories of singular thought, we are now in a position to establish the intended conclusion of this chapter. Let us consider the dual-component theory first.

According to the dual-component conception of *de re* thought, the content of a *de re* thought is an incomplete, *de dicto* proposition. The object the thought concerns, and the contextual relation which relates the subject and the object, fall *outside* the content of that thought. **(PA)**, the claim of privileged access, was formalised as follows.

(PA) For all x , if x believes that she thinks that p , then x thinks that p .

(PA) states that for any given thought, a subject can have privileged access to the *content* of that thought only. This will hold for privileged access to *de re* thoughts also: that is, for any given *de re* thought, introspection can yield knowledge of the content of that thought only: it can yield knowledge neither of the object of the thought, nor of the non-conceptual relation which relates the subject to the object. Consequently, for any given *de re* thought, introspection and conceptual analysis cannot yield non-empirical knowledge of the existence of the object of that thought. The non-empirical

knowledge which a subject can gain about her environment will, on this view, be restricted to knowledge of general facts, and will not extend to knowledge of the existence of particulars.

As noted above, however, the content of a *de re* thought is itself an incomplete *de dicto* content. On Burge's view, this *de dicto* content will be individuated externally, and will necessarily depend on contingent facts about the subject's environment. Hence introspective knowledge of the incomplete content of any *de re* thought can be used as the basis from which to infer to non-empirical knowledge of the environment in just the same way as any complete *de dicto* content. Consider once again the following.

(4'') Br (Ortcutt, <the woman in the alleyway>, Φ Spy (y), Woman (y) & In Alleyway (y) \ddot{U})

On the assumption that Ortcutt has privileged access to the content of his thought, Ortcutt can produce valid instance arguments for any of the three concepts which occur within its content. That is, he can validly infer that he is in a spy-world, a woman-world, and an alleyway-world.

If the dual-component theory is correct, then, *de re* thoughts can be used as the basis from which to infer to the nature of the external world only in so far as *de re* thoughts have a *de dicto* content, and not in virtue of the fundamental characteristics which, according to the dual-component theory, determine that they are *de re*.

Now let us turn to the SSTT. According to this theory, singular thoughts involve object-dependent senses. The cost of lifting Russell's restriction, and allowing that the content of a subject's thought could depend upon objects about the existence of which that subject could be mistaken, is that,

... we open the possibility that a subject may be in error about the contents of his own mind: he may think there is a singular thought at, so to speak, a certain position in his internal

organization although there is really nothing precisely there. (McDowell, 1986 p. 145)

The important question is how this affects the privileged access claim, **(PA)**. This question is important since it is privileged access as characterised by **(PA)** which, I have argued, allows a subject to generate instance arguments, and thereby come to have non-empirical knowledge of her environment.

At first sight it might seem that the truth of the SSTT actually falsifies **(PA)**, since the SSTT allows that a subject can think she has a singular thought when in fact she has no such thought. However, this is not so; the SSTT does not falsify **(PA)**. To see this, we need simply reflect on the fact that the *content* of the second-order belief corresponding to the antecedent of **(PA)** would itself be object-dependent. Ralph could not believe he thought *that cat should be kicked* without thereby thinking *that cat should be kicked*. That is, it would be impossible for a subject to entertain the second-order belief if she were unable to entertain the first-order object-dependent content.

A subject has privileged access to the content of her *de re* thoughts in the same way as she has privileged access to the content of her general thoughts. Maintaining that a subject can be mistaken about the *existence* of a thought is consistent with maintaining that she cannot be mistaken about *the content of a thought she has*.¹⁹⁶ **(PA)** is not refuted by the SSTT: a subject has privileged access to the contents of all her thoughts.

So where does this leave us with respect to instance arguments? In chapter four I argued that a subject can use her introspective knowledge of the content of a given propositional mental event to gain non-empirical knowledge of her environment. If there are propositional mental events which involve *de re* senses, a subject can equally use her introspective knowledge of the content of such propositional

¹⁹⁶ The force of a disjunctive theory such as the SSTT is that it advocates direct acquaintance with the world even in the face of possible error.

mental events to gain non-empirical knowledge of her environment. The crucial symmetry is this. From the fact that Susan has a *water* concept, Susan can gain non-empirical knowledge that she is in a water-world, since if she were not in a water-world she would not have a *water* concept: likewise, from the fact that Ralph has a *that-cat* concept, Ralph can gain non-empirical knowledge that that cat exists, since if that cat did not exist Ralph would not have a *that-cat* concept.

5.9 Conclusion

If the content of a *de re* thought is object-dependent, a subject can come to have non-empirical knowledge of the existence of the particular object or objects upon which her *de re* thought depends. If, on the other hand, the content of a *de re* thought is object-independent, such non-empirical knowledge is unavailable to her. In neither case is the validity of instance arguments undermined. Instance arguments allow one to infer from introspective knowledge of the content of a thought to a fact about one's environment.

Scepticism

6.1 Introduction

I have argued that semantic externalism entails the *prima facie* absurd thesis that contingent facts about the environment can be known on the basis of introspection and conceptual analysis alone. Further, I have argued that this consequence is acceptable and unsurprising once one rejects fully the Cartesian view of the mind. The nature of the dependence of the contents of one's propositional mental events on one's environment which defines semantic externalism serves to erode the once-clear divide between subject and world. Consequently, if semantic externalism is true, there is a class of valid inferences which can lead one from knowledge of the semantic contents "within" to knowledge of the world "without".

That such knowledge is available non-empirically might be thought to constitute a refutation of global external-world scepticism, such as that captured by Descartes' dreaming argument, or the supposition that one is being systematically deceived by an evil demon. In short, it would seem that semantic externalism, together with privileged access, can reassure one that such sceptical hypotheses are not true, since such sceptical hypotheses could not in such circumstances be true.

In this chapter I will examine an argument the purpose of which is to rule out one such sceptical hypothesis. Putnam has provided a transcendental argument to demonstrate that, on the assumption of a certain weak causal constraint on reference to which semantic externalism is committed, the proposition *I am a brain-in-a-vat* is self-refuting and necessarily false.¹⁹⁷ In line with my defence of instance

¹⁹⁷ Putnam (1981) chapter one. Putnam's argument has provoked a number of criticisms including the following. Smith (1984), McIntyre (1984), Brueckner (1986), Casati and Dokic (1991), and David (1991). For purported defences of Putnam see Dell'Utri (1990), and Warfield (1995).

arguments, I defend Putnam's argument. If semantic externalism is true, I maintain, a subject cannot be radically mistaken in her beliefs about the nature of her world, since interaction with the world provides the only means by which that subject could think about that world.¹⁹⁸

6.2 Putnam's argument

The sceptical hypothesis which Putnam is concerned to refute is a carefully formulated version of the sceptical hypothesis that you are a brain-in-a-vat. According to the standard brain-in-a-vat hypothesis, you are asked to imagine that your brain has been removed from your body by an evil scientist and placed in a vat of nutrients which keep it alive. You are then asked to imagine that the scientist, by stimulating the nerve-endings in your brain, causes you to have the illusion that everything is perfectly normal. It seems to you as if you are interacting with the same world in the same way as you always have done. As with all forms of global external-world scepticism, the sceptical hypothesis that you are a brain-in-a-vat characterises a situation which is compatible with all the evidence available to you, and which is nevertheless consistent with the world being largely other than you perceive it to be. The sceptical hypothesis takes its force from the fact that it does not seem to you as if you are a brain-in-a-vat; rather, it seems to you as if you inhabit a world with forests, lakes, cities and people.¹⁹⁹

A large part of the discussion in this chapter will focus on a formalisation of Putnam's argument presented by Wright (1991), who, as will become clear, effectively holds a middle position.

¹⁹⁸ The weak causal constraint on reference employed in Putnam's argument is by no means definitive of semantic externalism. However, even while Putnam does not relate his argument directly to semantic externalism, it is specifically the consequences of semantic externalism which interest me here; in particular, the relationship between scepticism and instance arguments.

¹⁹⁹ In fact, it is hard to understand what would be involved in having a sense impression as of being a brain-in-a-vat. It is not as if one could

Putnam's hypothesis differs from the standard brain-in-a-vat hypothesis in three broad ways. First, you are asked to suppose not only that you are currently a brain-in-a-vat, but that you have *always* been a brain-in-a-vat. Thus the vat is, as it were, your natural habitat; to whatever extent you can be said to have concepts, those concepts were acquired in your current environment, and are hence appropriate to it.²⁰⁰ Second, rather than supposing that you are alone in your vat, with an external scientist causing your hallucinations, you are asked to suppose that the universe simply consists of a vat which sustains every living subject, and that your hallucinations are caused by automatic machinery programmed for that very purpose. Thus your belief that your world is populated by beings similar to yourself is true, even though the beings which share your world are not as you perceive them to be. Third, you are asked to suppose that the automatic machinery is programmed to generate a coherent, *collective* hallucination. Thus your beliefs about the nature of your environment largely cohere with those of the other subjects who have the misfortune to share your predicament.

The significance and plausibility of each of these assumptions will be discussed in some detail below. For the moment, let us move directly to Putnam's argument. Putnam writes,

when the brain-in-a-vat (in the world where every sentient being is and always was a brain-in-a-vat) thinks 'there is a tree in front of me' his thought does not refer to actual trees. On some theories ... it might refer to trees in the image, or to the electronic impulses that cause tree experiences, or to the features of the program that are responsible for those electronic impulses. ...

"look" around and see that one was ensconced in a vat full of nutrients, or that one was surrounded by other brains, since one would not have the eyes with which to see anything. Perhaps it is possible to imagine being a brain-in-a-vat, but it is certainly not possible to imagine it *seeming* to one as if one were a brain-in-a-vat.

²⁰⁰ Remember that we are assuming a causal constraint on reference such as that embodied in semantic externalism.

These theories are not ruled out ... for there is a close causal connection between the use of the word 'tree' in vat English and the presence of trees in the image By the same argument, 'vat' refers to vats in the image in vat English, or something related ..., but certainly not to real vats, since the use of 'vat' in vat English has no [relevant] causal connection to real vats It follows that if [the brain's] 'possible world' is really the actual one, and we are really the brains-in-a-vat, then what we now mean by 'we are brains-in-a-vat' is that *we are brains-in-a-vat in the image* or something of that kind (if we mean anything at all). But part of the hypothesis that we are brains-in-a-vat is that we aren't brains-in-a-vat in the image (i.e. that what we are 'hallucinating' isn't that we are brains-in-a-vat). So, if we are brains-in-a-vat, then the sentence 'we are brains-in-a-vat' says something false (if it says anything). In short if we are brains-in-a-vat then 'we are brains-in-a-vat' is false. So it is necessarily false. (Putnam, 1986 pp. 14-5)

The causal constraint on reference ensures that the sceptical hypothesis expressed by the proposition that we are brains-in-a-vat cannot be true.²⁰¹

Let us return to the promised discussion of the three assumptions explicit in Putnam's formulation of the sceptical hypothesis which sets it apart from the standard brain-in-a-vat hypothesis. According to the first of these, we have always been brains-in-a-vat: hence our thoughts have the same content now as they always did. Of what significance is this? In chapter four it was argued that a subject could use instance arguments to yield knowledge not of her *current* environment, but strictly speaking only of her *recently past* environment. Since meaning-conferring relations take some time, there is always a possibility that one run through an instance argument having been recently switched from the environment to which one's concepts connect. But Putnam's

²⁰¹ The proposition that I am a brain-in-a-vat is not necessarily false, since there is a possible world in which I am a brain-in-a-vat. But if I were a brain-in-a-vat, my utterance of "I am a brain-in-a-vat" would no longer express the proposition that I am a brain-in-a-vat.

hypothesis rules out this very possibility. It might therefore be thought that the conclusion of Putnam's argument could escape the qualification: that we could genuinely conclude that we *are* not brains-in-a-vat. On reflection, however, this is not so. The sceptical hypothesis against which Putnam's argument is directed is the sceptical hypothesis that *we have always been brains-in-a-vat*. The conclusion of the argument should therefore be that *we have not always been brains-in-a-vat*.²⁰²

The second assumption is that there is no evil scientist, and no "real" world outside the vat; there is nothing other than us as brains-in-a-vat and the automatic tending machinery. In this there is a tangible advantage, since it serves to rule out the possibility that, despite our envatted predicament, our concepts nevertheless refer to real objects; to objects *outside* the vat. Putnam's argument depends on the concepts of a brain-in-a-vat having a reference which differs from the reference of a non-envatted subject. In the evil scientist scenario, it would at least be open to the proponent of a causal theory of reference to maintain that the brain-in-a-vat's token of "tree" referred to real trees; for instance if the evil scientist were consistently to cause the brain-in-a-vat to have images which reflected what he, the evil scientist, was perceiving. Putnam's hypothesis ensures that the conceptual repertoire of the brain-in-a-vat is not derivative; derived that is from sentient beings the direct cause of whose perceptions differs from the direct cause of the perceptions of the brain-in-a-vat.

The third assumption is that the hallucinations to which we are subject largely cohere. Putnam goes on to maintain that in such a situation we would, in effect, be able to communicate. He writes,

Let us suppose that the automatic machinery is programmed to give us all a *collective* hallucination, rather than a number of separate unrelated hallucinations. Thus, when I seem to

²⁰² Having noted this qualification, I will for the moment leave it to one side.

myself to be talking to you, you seem to yourself to be hearing my words. Of course it is not the case that my words actually reach your ears – for you don't have (real) ears, nor do I have a real mouth and tongue. Rather, when I produce my words, what happens is that efferent impulses travel from my brain to the computer, which both causes me to 'hear' my own voice uttering those words and 'feel' my tongue moving, etc., and causes you to 'hear' my words, 'see' me speaking, etc. In this case we are, in a sense, actually in communication. for you do, after all, really hear my words when I speak to you, even if the mechanism isn't what we suppose it to be. (Putnam, 1981 pp. 6-7, original emphasis)

Perhaps Putnam is right about this. After all, it is our mental states which are assumed to cause, albeit indirectly, our collective hallucinations. However, the claim is at least controversial, and it is not clear that it adds anything to the force of the argument. In what follows, I will proceed as if the conclusion were the singular proposition *I am not a brain-in-a-vat*. That this conclusion can be reached non-empirically leaves us with a substantial issue.²⁰³

6.3 Some qualifications and a proof

Putnam's argument could be regarded as a *meta-argument*. It concerns the propositions which would be expressed by subjects whose syntactically identical utterances of "I am a brain-in-a-vat" differ semantically. More importantly, it concerns the *arguments* such subjects could produce to counter whichever sceptical hypothesis would be expressed by their respective utterances of that sentence syntactically individuated. It is important to distinguish clearly between Putnam's meta-argument and the arguments the subjects he considers could run through. Putnam's meta-argument

²⁰³ Putnam's sceptical hypothesis actually appears to diminish the horrific nature of the original "nightmare": you are able to communicate, and you have a certain degree of control over your "actions". Would such a situation be all that bad?

consists of philosophical reflection on the arguments of such subjects. Wright has formulated the common form of such arguments as follows.²⁰⁴

- (i) My language disquotes
- (ii) In BIVese, “brain-in-a-vat” does not refer to brains-in-a-vat
- (iii) In my language “brain-in-a-vat” is a meaningful expression
- (iv) In my language “brain-in-a-vat” refers to brains-in-a-vat (from (i) and (iii))
- (v) My language is not BIVese (from (ii) and (iv))
- (vi) If I am a brain-in-a-vat, my language, if any, is BIVese (definition of BIVese)
- (vii) I am not a brain-in-a-vat. QED. (Wright, 1991 p. 74)

Henceforth, Putnam’s argument will be referred to as the meta-argument; Wright’s argument-form will be referred to as The Proof; the argument which a brain-in-a-vat would express by running through The Proof will be referred to as **(ABiv)**; and the argument which a normal subject would express by running through The Proof will be referred to as **(ANorm)**. As will become clear, perceived problems with Putnam’s meta-argument emerge from a failure to distinguish between the specific arguments **(ABiv)** and **(ANorm)** on the one hand, and the meta-argument which concerns them on the other.

In his meta-argument, Putnam moves from a claim about *thought* to a claim about *language*. He moves from the claim that “when the brain-in-a-vat ... thinks ‘there is a tree in from

²⁰⁴ I have chosen to focus on Wright’s formalisation because it is, as far as I know, the one which is most true to the letter of Putnam’s argument. Other tend to beg central questions, and therefore fail to do justice to the force of Putnam’s argument. For instance, in their response to a formalisation by Dell’Utri (1990), Casati and Dokic (1991, p. 91) remark that “[t]he reconstruction has the advantage of presenting the argument in such a way that it is easy to detect some major flaws in it”. See also Brueckner (1986), who attempts to formulate in English the argument available to a brain-in-a-vat, and Warfield (1995).

of me' his *thought* does not refer to actual trees", to the conclusion that "if we are brains-in-a-vat, then the *sentence* 'we are brains-in-a-vat' says something false"²⁰⁵. While there may be significant differences between psychological content and linguistic content, the move between them is legitimate in this context. This can be seen by reflection on the genesis of semantic externalism. As will be remembered from chapter one, semantic externalism arose from the following realisation: any difference in the extensions of terms which occur in the content-clauses of propositional attitude attributions will reflect a difference in the propositional attitudes thereby attributed. This means that the content of a thought and the content of an utterance will be causally constrained, and will vary accordingly with variation in environmental differences.

Wright's formalisation focuses explicitly on the meaning of linguistic utterances, and makes explicit appeal to the notion of a language. The question as to the nature and individuation conditions of a language is substantial. However, there is an interpretation of the expression "my language" which should be uncontroversial in this context. I will take a subject's language to consist of all the expressions in her language which she understands: all those expressions which correspond to concepts she possesses.²⁰⁶

It is worth noting that in neither Putnam's argument nor Wright's proof require a specific form of causal theory of reference. Rather, Putnam and Wright make appeal to a weak notion of causal constraint. This means they can avoid *prima facie* worries such as the one illustrated by the following example. Doug, whenever he eats cheese, suffers the misfortune of hallucinating that he is surrounded by brains-in-a-vat. Doug, not being particularly bright, eats cheese on a frequent basis without realising the effect it has on him. It might be thought to be compatible with what Putnam writes that Doug's expression "brain-in-a-vat" refers to brains-in-a-

²⁰⁵ Putnam (1981) pp. 14-5, emphasis added.

²⁰⁶ Again, the understanding involved here can be partial, and there may well be vague cases.

vat-in-the-image: “[t]hese theories are not ruled out ... for there is a close causal connection between the use of the word [“brain-in-a-vat”] ... and the presence of [brains-in-a-vat-in-the-image]”²⁰⁷. Yet any theory of reference which has as a consequence that Doug’s expression “brain-in-a-vat” refers to images of brains-in-a-vat surely ought to be rejected. Any adequate theory of reference must explain why his expression “brain-in-a-vat” refers to actual brains-in-a-vat and not to images of brains-in-a-vat.²⁰⁸

As noted above, neither Putnam’s argument nor Wright’s formalisation are committed to the claim that the tokening of a propositional mental event, or linguistic utterance, refers to whatever causes it. All that is required is that reference be *causally constrained*: if there is *no* causal connection between “F” and x’s, “F” cannot refer to x’s. Since there is no causal connection between “brain-in-a-vat” as thought or uttered by brains-in-a-vat, and actual brains-in-a-vat, “brain-in-a-vat” cannot for them refer to brains-in-a-vat.

The premise needed for the argument to go through is the weaker, and more plausible premise, that *whatever* the brain-in-a-vat can do with an utterance of “brain-in-a-vat”, it certainly can’t refer to brains-in-a-vat.

This also means that neither Putnam nor Wright are committed to the claim that a brain-in-a-vat could think or have a language.²⁰⁹

²⁰⁷ Putnam (1981) p. 14.

²⁰⁸ Answers to this problem are particularly forthcoming in naturalistic theories of content. For example Fodor (1987) offers an asymmetric-dependence theory, according to which Doug’s expression “brain-in-a-vat” would refer to brains-in-a-vat and not to images of brains-in-a-vat because of an asymmetric dependence of “brain-in-a-vat” on brains-in-a-vat, and not on images of brains-in-a-vat. See also Millikan (1984) and (1993).

²⁰⁹ This is explicit in premise (ii) of The Proof.

6.4 Brueckner's problem

Brueckner has argued that the assumptions of **(ABiv)** and **(ANorm)** “engender a skepticism about knowledge of meaning, or propositional content, which undercuts [their] anti-skeptical force”²¹⁰. Brueckner remarks that,

... if I do not know whether *S* is speaking English or vat-English, then I cannot apply a disquotational principle ... to *S*'s utterance of '*S* is a BIV' and conclude that those utterances are true iff *S* is a BIV. (Brueckner, 1986 p. 164)

In this he is surely correct. To identify the referent of a subject's expression one first has to know which language the subject is speaking. Such knowledge is plausibly *comparative* knowledge; knowledge that the subject speaks *English as opposed to BIVese*, for instance. But Brueckner continues,

Similarly, if I do not know whether *I* am speaking English or vat-English, then I cannot apply [a disquotational principle] to my own utterances of 'I am a BIV' as a step toward the conclusion that I know that I am not a BIV (Brueckner, 1986 p. 164, original emphasis)

The problem he identifies is that if, prior to running through The Proof, I do not know whether or not I am a brain-in-a-vat, I cannot know the referent of my expression “brain-in-a-vat”, and am therefore in no position to identify it by disquotation as at line (iv). But in this he is surely mistaken. To identify the referent of an expression *in my language* no such comparative knowledge is necessary. Even on the assumption that I do not know that I am speaking English as opposed to BIVese, I do know that whatever “brain-in-a-vat” refers to in my language I may identify its reference by using that very expression.²¹¹ Homophonic disquotation is not rendered illegitimate by semantic externalism.

²¹⁰ Brueckner (1986) p. 148. Brueckner would not express his point like this, since he fails to distinguish clearly between the specific arguments, which I have called **(ABiv)** and **(ANorm)**, and the meta-argument.

Brueckner's worry about knowledge of meaning exactly parallels the worry about knowledge of propositional thought-content expressed by the *Content Sceptic's Argument* considered in chapter two. The *Content Sceptic's Argument*, it will be remembered, seeks to undermine a subject's knowledge of her propositional thought-content as follows. If Susan is on Earth, she is thinking that water is a liquid. If Susan were on Twin Earth, she would be thinking that *twater* is a liquid. The evidence available to Susan is consistent both with her currently thinking that *water* is a liquid and with her currently thinking that *twater* is a liquid. Hence Susan does not know that she is currently thinking that *water* is a liquid. The principle of privileged access, **(PA)**, shows why this line of reasoning is mistaken; it demonstrates that knowledge of propositional thought-content is not based on evidence, and does not require that one rule out various possible twin thoughts. It is consistent with semantic externalism that a subject cannot be mistaken in her self-ascriptive beliefs.

Similarly, Brueckner seeks to undermine knowledge of linguistic meaning as follows. If Susan is a normal subject, her expression "brain-in-a-vat" refers to brains-in-a-vat. If Susan were a brain-in-a-vat, her expression "brain-in-a-vat" would refer to brains-in-a-vat-in-the-image. The evidence available to Susan is consistent both with her being a normal subject and with her being a brain-in-a-vat. Hence Susan does not know that her expression "brain-in-a-vat" refers to brains-in-a-vat. But just as knowledge of propositional thought-content is not undermined by semantic externalism, neither is knowledge of linguistic meaning for expressions in one's language. Knowledge of linguistic meaning is not based on evidence, and does not require that one rule out various possible twin meanings.²¹²

Note how the problem arises from a failure to distinguish clearly between **(ABiv)**, **(ANorm)**, and the meta-argument

²¹¹ I take it that knowledge of which language you are speaking would be equally guaranteed by semantic externalism.

²¹² As noted above, if "X" is an expression in S's language, then S has the concept x.

which reflects upon them. The *Content Sceptic's Argument* trades on the mistaken idea that there could be one second-order thought, and yet two possible first-order thoughts, only one of which could be correctly attributed. Similarly, Brueckner's worry trades on the mistaken idea that there could be one sceptical hypothesis, and yet two possible arguments, one of which refutes it and one of which does not. However, just as a subject could not entertain the very same second-order belief – that she was thinking that water is a liquid – on Earth and on Twin Earth, a subject could not entertain the very same sceptical hypothesis – that she was a brain-in-a-vat – were she a normal subject or a brain-in-a-vat. Whatever a brain-in-a-vat can do, it certainly cannot entertain *that* sceptical hypothesis.

Brueckner's problem is not legitimate, and Putnam's meta-argument works. Both a brain-in-a-vat and a normal subject would establish the falsity of their respective sceptical hypotheses by running through The Proof.²¹³

6.5 The remains of scepticism

Putnam's argument yields a transcendental refutation of global external-world scepticism. On the assumption of semantic externalism it is a short step from understanding the expression "brain-in-a-vat" to knowing that you are not a brain-in-a-vat. To the extent that running through The Proof yields non-empirical knowledge of the external world, it is of a piece with instance arguments, and is to be defended in the same general way.

However, semantic externalism may well leave room for a rather strange form of scepticism. Although Wright has been concerned to defend the claim that Putnam's argument is sound, he nevertheless maintains that "it does not sustain

²¹³ Taking Brueckner's problem seriously in fact engenders a further form of scepticism which undermines the original sceptical hypothesis. If I do not know what my expression "brain-in-a-vat" refers to, I presumably do not even know which sceptical hypothesis it is that I want to refute.

the conclusion that, in the way we would like, the nightmare is refuted"²¹⁴. He writes,

... the real spectre to be exorcised concerns the idea of a thought *standing behind* our thought that we are not brains-in-a-vat, in just the way that our thought that they *are* mere brains-in-a-vat would stand behind the thought ... of actual brains-in-a-vat that 'We are not brains-in-a-vat'. What we should really like would be an assurance that there is no such true thought: an assurance not just that most of what we think is actually true – for semantic externalism might well deliver that result for the brains-in-a-vat (Wright, 1991 p. 93, original emphasis)

I think this is ultimately correct. Instance arguments *can* provide a subject with substantial knowledge of her environment, and such knowledge should not be underestimated. But an uneasy feeling remains. To acknowledge that one is securely related to one's environment in the way semantic externalism proscribes is also to acknowledge that one's cognitive capacities are *restricted* by that environment. This is manifest in the claim that actual subjects and counterfactual subjects *cannot* have the same concepts. Reflection on this becomes worrying, since it opens us to the possibility that the fundamental nature of our relation to our environment may be outside our cognitive grasp. As Wright says,

But of course, if there were such a true thought, standing behind us as it were, it would no more be available to us than the thought that they are brains-in-a-vat would be available to the envatted brains. (Wright, 1991 p. 93)

6.6 Conclusion

In this chapter I have defended Putnam's argument. Semantic externalism does indeed provide a subject with the means to refute such sceptical hypotheses as that presented by

²¹⁴ Wright (1991) p. 73.

Putnam. This is because, if semantic externalism is true, the world provides the very means by which we can think about it. This, however, turns out to be a double-edged sword. It does ensure that we cannot be radically mistaken in our beliefs about the world; but it also ensures that we may not be able, as it were, to get outside our cognitive world to reflect upon the mechanisms which connect us to it. If this latter thought gives rise to a genuine worry, it is not one of which we could easily rid ourselves.