The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

University of Cape Town

INTERPRETING THE SELF: AN ANALYSIS OF THE FIRST-PERSON'S PERSPECTIVE OF BELIEFS IN DONALD DAVIDSON'S RADICAL INTERPRETATIONISM

ANNEMIE GILDENHUYS

Thesis submitted in fulfilment for the Master of Social Science in Philosophy

University of Cape Town

2006

The financial assistance of the National Research Foundation of South Africa towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the author, and are not to be attributed to the National Research Foundation.

University of Cape Town

ABSTRACT

In this thesis I examine the ability of Donald Davidson's influential theory of radical interpretationism to accommodate the differences between a first-person's and a third-person's perspectives of beliefs. I first describe the differences between these two perspectives, which collectively I call the asymmetry thesis. I then investigate in detail two interpretations of Davidson's strategy for accommodating the asymmetry thesis in his system of radical interpretationism. I call the standard interpretation the meaning asymmetry, and my own interpretation the sentence held-true asymmetry. I argue that, notwithstanding the moderate success of the meaning asymmetry in accounting for the asymmetry thesis in theory, it cannot be a plausible explanation of it since it cannot be translated into what we do in our linguistic communities with our shared linguistic conventions. I conclude by describing my sentence held-true asymmetry and arguing that it allows Davidson's radical interpretationism to accommodate the asymmetry thesis, both in theory and in practice. This dissertation, thus, opposes the popular belief that Davidson's system of radical interpretationism cannot accommodate the asymmetry thesis, precisely because it rejects the standard interpretation of Davidson's strategy for achieving this, in favour of one that utilises its strengths without falling victim to its weaknesses.

University of Cape Town

ACKNOWLEDGEMENTS

I would like to thank the following people for their contributions to this thesis: my long-suffering supervisor, Dr. Jeremy Wanderer, for his patience and comments towards the end; Dr. Bernhard Weiss, for his help when I really didn't have a clue and had no one else to turn to; the National Research Foundation, for the financial assistance; Reinette Popplestone and Denise Oldham at the Disability Unit, for both the assistance with annoying paper work and the ears when I needed to complain; Ingrid Thompson, Celia Walter and the rest of the staff at the UCT library's Humanities Reference Desk, without whom I would not have been able to do any research; and everyone who was brave enough to employ me, for the confidence and, of course, the sources of income. Thanks to friends like Nafisa Baboo, Dean Chapman, Karien de Wet, Karen du Toit, Elisa Galgut, Denise Oldham, Carmen Perez, Reinette Popplestone, Yolande Ruiters, Sameer Vasta, Ferdi Venter, Bernhard Weiss, Monique Whitaker and Keith (who very much would not want his name mentioned) for the social life and support without which I would not have gotten far. I also want to thank Jeremy Allcock, Lara Davison and anonymous Josh, who were always there both to laugh with and to remove so many obstacles that seemed insurmountable along the way; Ben Gildenhuys, who uncomplainingly supported me financially when all sources of money dried up; Helena Gildenhuys, who demonstrated all the character traits that I want to strive towards; and Liesel, without whom my life would have been so much harder and less interesting. Thanks to Lee, to whom the thesis is dedicated, for teaching me most of what I know about everything worth knowing. Be glad that I can never force you to read it!

University of Cape Town

TABLE OF CONTENTS

University of Cape Town

University of Cape Town

CHAPTER 1
INTRODUCTION

In this thesis I intend to evaluate the ability of Donald Davidson's theory of interpretationism to account for the way in which the first-person's stance towards her own beliefs differs from that of her interlocutors. The project will focus on the first-person's perspective on beliefs specifically, and not on sensations like pain or thirst. I also intend the conclusion not to be generalised to all propositional attitudes, since the problems that the different propositional attitudes pose for the first-person perspective may differ. Before spelling out the problem at the heart of this project, a brief, and regrettably over simplistic, history will reveal the origin of the question. Since this project focuses on beliefs, as opposed to sensations, the historical overview will likewise focus on beliefs.

Cartesian dualism owes its past popularity to the way in which it fits with our intuitions regarding our access to our own mental states, of which I will use beliefs as example. It is reasonable to suspect that this was, in fact, Descartes' original motivation for it. According to these intuitions, a first-person has authoritative and immediate access to her own beliefs. She is in a position of authority over her own beliefs because she is usually right about what she believes. She enjoys immediate access to her beliefs since she knows, without having to make use of her own behaviour, speech or environment as evidence, what she believes. These intuitions found their ideal home in dualism, according to which the first-person's authoritative and immediate knowledge of her beliefs was secured by the fact that they were states that were necessarily first-person accessible and not necessarily third-person accessible. This led Descartes to propose that our mental states were located in an inner space accessible only to the person whose space it was. The first-person could then access her own beliefs by scanning this private space, via a mechanism like introspection, for example. She possessed a collection of beliefs which she discovered directly through introspection and which others could discover only by drawing inferences from her behaviour and speech, thereby securing her immediate access. And, because she had direct and unmediated access through introspection, while others had to access her beliefs indirectly through her behaviour, she was in a better

University of Cape Town

position to know what her beliefs were than anyone else, thereby securing her position of authority. The only type of space that Descartes could think of that would not be necessarily third-person accessible, was an immaterial mind, somehow attached to a material body. Cartesian dualism, thus, became a theory that held that minds were immaterial entities that housed immaterial mental states that the person to whom they belonged could access immediately and authoritatively.

The 20th century was a period in which Cartesian dualism was severely crippled by numerous objections that philosophers made to it. It was mostly criticised for making our minds essentially inaccessible to others, which was precisely the strategy whereby it secured first-person authority and immediacy. Firstly, philosophers realised, partly due to Sigmund Freud's theory of the unconscious, that all our mental states were certainly not accessible to us and that third-persons were often in a good position (sometimes even a better position than us) to know what mental state we were in. The introspective scanning of a necessarily first-person accessible inner space made it almost impossible for the first-person to be wrong about her own beliefs, and the fact that her beliefs themselves were necessarily inaccessible to observers made it almost impossible for a third-person to correct her regarding them.

Secondly, many questions arose as to the plausibility of a theory that claimed that internal states could allow us to be in touch with the external world. Beliefs, for example, were meant to be about events in an external world, but if they were wholly private, it was difficult to see how we could ever know that they represented something in an external world. Cartesian dualism, since it insisted on internal states as the vehicles of knowledge about external things, isolated us from the world that we were attempting to access. It, as a result, opened up the possibility that our beliefs were the way they were, independent of what was going on in the world around us; an idea that Descartes himself conveyed through his evil demon thought experiment.

These objections, together with numerous others, gave rise to a shift away from dualism towards materialist theories of mind, according to which there was certainly no talk of immaterial minds and mental states. Most of these materialists, however, still held onto one of the main commitments of Cartesian dualism, namely, the conception of beliefs (and mental states more generally) as internal states. The

University of Cape Town

difference was that they acknowledged that, even though a first-person could access her own beliefs directly and authoritatively, the material nature of the mental allowed a third-person to access them as well. Philosophers provided numerous different explanations for authority and immediacy, but they all essentially relied on some form of internal scanning mechanism to scan the person's internal physical or functional states. The two most prominent theories were the identity theory, according to which beliefs just were brain states, and functionalism, according to which beliefs were functional states that involved their causes, their relations to other functional states and their outputs.

These theories, then, became the target of interpretationists who did not believe that they improved much upon Cartesian dualism. Firstly, interpretationists point out that beliefs (and more generally all propositional attitudes like desires, intentions, hopes, etc. are constitutively governed by rationality, unlike any physical, mechanical or computational device known to us. One of the main advantages of physicalist theories over Cartesian dualism is that they allow empirical, scientifically verifiable theories of the mental in which beliefs exhibit the same sort of law-like relationships (to the world, to each other and to behaviour) that are found in empirical sciences like geography and biology. Interpretationists, however, think that beliefs have to be defined, not in terms of their causal roles in producing behaviour (like functionalism maintains), but rather in terms of their roles in rationalizing other propositional attitudes and behaviour. Rationality, and the rationalisation of other propositional attitudes and behaviour, according to them, cannot involve the same law-like relationships found in empirical sciences, since it prescribes, given our beliefs and desires, firstly, what we ought to do, secondly, how we ought to reason and thirdly, which other beliefs and desires we ought to hold. It involves what we ought to do, not what we, as a matter of empirical fact, do. It involves prescriptions, not descriptions. It involves values, not facts. If we want to respect the constitutive role that rationality plays in beliefs together with the identity of beliefs as rationalisations of other propositional attitudes and of behaviour, then we cannot expect theories that involve them to take the same form as the empirical theories that hold over functional and physical brain states.

University of Cape Town

Secondly, one feature, specifically of beliefs, that was first pointed out by Ludwig Wittgenstein, is the transparency that they seem to have to the world. When the first-person decides that she has a belief, she is not just describing some physical or functional brain state, but she is actually committing herself to the truth of something in the world. If beliefs were just physical brain states, it is difficult to know why it would not be possible for her to describe those states without committing herself to the truth of something in the world. It is certainly a feature that no other bodily state has. Functionalism does not do much better, since it cannot explain how an identification and a description of a functional state can commit us to the truth of something in the world, while an identification and description of exactly that same functional state in another is no more than a mere identification or description of a functional state.

These problems brought about theories, of which interpretationism was one, that essentially abandoned the conception of mental states as internal states. According to interpretationists, beliefs (and more generally all propositional attitudes) are not inner states. That is, claims about what a subject believes, for example, are not claims about her internal physical (or any other) states. The interpretationist trend is to define such attitudes as properties or statuses that individuals have in virtue of being interpreted as such by others. A belief that P, for example, is a property that we attribute to a person, like a length of thirty centimetres or twelve inches are properties that we attribute to a ruler. We obviously have physical brain states, and we may have functional states, in the same way as a ruler has length. But in order to call something a belief that P, it has to be attributable to us by a third-person's judgement of its rationalising role in our behaviour. As in the second objection to the identity theory and functionalism above, interpretationists think that beliefs just are properties that rationalise other propositional attitudes, behaviour and speech. There is no element of beliefs that falls outside a project of interpretation that rationalises our attitudes, behaviour and speech. For something to qualify as mental, it, thus, has to be attributable by a third-person in a context of this type of rationalisation or interpretation. That is, we cannot settle the fact of what we believe on our own or in conjunction with a neurologist that is examining our brains or our nervous systems. Conversely, something cannot be such an attitude if it is not possible for a third-person to attribute it to us. Beliefs do not exist independent of being attributed to us because third-person attributions fix the

University of Cape Town

beliefs, desires, intentions, etc. that we have. Third-person attributions are often wrong, as we all know, but this is because they do not use proper methods to interpret us, or because they know too little about us to fix our propositional attitudes accurately. They can be fixed accurately only by a fully informed interpreter that is employing the right sort of system of interpretation. What a fully informed interpreter, who employs a proper system of interpretation, cannot find out about the propositional attitudes, is, thus, not a propositional attitude at all. If a proper method of interpretation is used, it will yield everything that can be called propositional attitudes. So, if we know what a proper system of interpretation is, and we know everything such a system of interpretation can interpret, then we know everything there is to know about the propositional attitudes. More precisely, if we want to know what a belief is, we can study the best method on the basis of which it is attributed. So, if we want to know which beliefs we have, we can study the best method whereby they are attributed to us.

The particulars of the interpretationist proposal have been elaborated in a variety of ways, but they all share the claim that the process of interpretation, whereby others attempt to make sense of us, fixes the propositional attitudes that we possess. By this they do not mean that the epistemological question of what a person, for example, believes can be answered by conducting a project of interpretation. They literally mean that a belief is constituted by, or just is, whatever the best scheme for making sense of us will attribute to us. No neurophysiological examination of a brain or a nervous system can contribute information about the mental that an appropriate interpretive scheme cannot. This claim has, as a result, been termed interpretationism and its original and leading advocate is Donald Davidson. Interpretationists would probably want to extend their theory to the sensations as well, but, since this is still in its infancy, I have not, and will not, include them in the discussion.

The interpretationist trend has given rise to its own difficulties, of which the question of the first-person perspective is the most significant and persistent. If Cartesian dualists, identity theorists and functionalists all deliberately constructed theories with beliefs as inner states in order to accommodate authority and immediacy, interpretationists' commitment to mental states as socially ascribed properties or social statuses (rather than inner states), poses an immediate problem for them. If our

University of Cape Town

CHAPTER 2
THE ASYMMETRY THESIS

Philosophers often suggest that there are differences between the way in which we know, and relate to, our own beliefs and the way in which we know, and relate to, those of others. They generally agree that a theory of beliefs and belief attributions is more desirable if it is able to leave room for the differences between these two perspectives. In this chapter I intend to examine what these differences are and why it is desirable that theories of beliefs and belief attributions are able to leave room for them. In section 1, I will consider various methodologies for highlighting these differences, with the aim of arguing that we should derive them from a pre-theoretical description of our everyday practices rather than from theories of self-knowledge or from personal reports of believing. In section 2, I will show that this pre-theoretical description of our everyday practices reveals three such differences between the first and third-person perspectives on beliefs, that individually I call the evidence, the transparency and the authority asymmetries, and collectively I call the asymmetry thesis. In section 3, I will show why it is desirable that any theory of beliefs is able to leave room for the asymmetries. In section 4, I will suggest four desired constraints on the explanations of the asymmetry thesis.

University of Cape Town

1. METHODOLOGY

If theories of beliefs are more desirable if they can leave room for the asymmetry thesis, the need to spell out the asymmetries becomes evident. The difficulty with doing so is that there exists no fixed collection of such asymmetries that such theories can be assessed on. Philosophers disagree about the exact ways in which the two perspectives differ, and their disagreement can often be traced to the method by which they arrive at the asymmetries.

Before proceeding to a description of the asymmetries, I wish to justify my view that, in a project that proposes to assess a theory of beliefs, the ways in which the first- and third-person perspectives differ should be derived from a description of our common everyday epistemic practices involved in justifying beliefs and our knowledge about them. I propose to do this by describing three possible strategies for determining the

asymmetries together with an explanation of why, in this type of project, the third is preferred to the first two.

The ways in which the two perspectives on beliefs differ can be established, firstly, by appealing to the best theories or constructing a new theory, secondly, by consulting our personal reports about believing and, thirdly, by describing those of our practices that exhibit the ways in which we routinely treat our own beliefs differently from those that we attribute to others.

1.1. THE ASYMMETRIES AT A THEORETICAL LEVEL

The nature of the asymmetries often depend, either upon the theories that attempt to explain them, or upon more comprehensive theories of mind. For example, most logical behaviourists argue that we are more likely than others to be correct about which beliefs we hold because we spend more time observing our own behaviour than others do.[1] Several contemporary Theories that employ concepts like rationality and avowal in their accounts of self-knowledge, on the other hand, claim that the greater likelihood of being correct about our beliefs is, in some sense, a constitutive part of self-knowledge.[2] Further, we are all familiar with the way in which traditional Cartesianism brought about the first-personal notions of infallibility and omniscience; that is, the idea that we cannot be mistaken about what beliefs we hold and the view that we cannot be unaware of a belief that we do hold, respectively.[3] These are all versions of an authority asymmetry, but the theories within which they appear advance very different construals of such authority. If theorists propose different dissimilarities between the first- and third-person stances towards beliefs, then they will inevitably expect theories of beliefs to accommodate different asymmetries. The consequent disagreement will then concern the problem of which asymmetries need to be accommodated, as opposed to the question of the plausibility of such theories. A project that is aimed at evaluating a theory, like interpretationism, in terms of whether it provides a plausible explanation of the multi-perspectival character of beliefs will, as a result, have to be designed to persuade the proponents of such theories of three conclusions: firstly, that the specific asymmetries defended in the project are correct (which may require a lot of theorising), secondly, that any theory that fails to account for them is an implausible theory of beliefs and, thirdly, that interpretationism cannot

University of Cape Town

account for them. This threatens to be a lengthy and laborious task and, if attempted in one project, it runs the risk of neglecting to deal with all three these questions in sufficient depth.

A second disadvantage of evaluating a theory according to whether it can accommodate asymmetries obtained from a specific theory, is that the conclusion of such a project will be of little importance to many. There is not much to gain from a project that attempts to establish whether one theory (such as interpretationism) can accommodate conditions spelled out in another theory (such as the rationality or avowal models of self-knowledge). This is true precisely because everyone does not hold these theories. An interpretationist is unlikely to be convinced that her theory is defective just because it cannot accommodate a theory that she may not even endorse. She is more likely to work on a reply that rejects the other theory. She may be more easily convinced of the need to revise her theory if she is shown that it cannot accommodate some of the strongest intuitions about beliefs which she most probably shares.

1.2. THE ASYMMETRIES AT A PRE-THEORETICAL LEVEL

Instead of relying on a theory to provide us with the asymmetries, we can work at a pre-theoretical, intuitive level to determine what the asymmetries are that theories of beliefs are required to account for. This resembles the procedure most commonly employed by moral and political philosophy according to which we construct theories that are consistent with our intuitions and according to which our theories are permitted to disregard such common intuitions only in cases where they can be demonstrated to be in error. In this case we expect all theories of beliefs to be able to allow for an explanation of the way in which we commonly think of our own beliefs as being different in some respects from those we attribute to others.

Such pre-theoretical data can be derived from two sources, either from people's personal reports about beliefs and believing, or from some of our most common everyday practices.

University of Cape Town

1.2.1. THE ASYMMETRIES DERIVED FROM PERSONAL BELIEF REPORTS

There are good reasons to suspect that the latter option is more feasible than the former. If the pre-theoretical level is preferred to the theoretical level because it promises more agreement over what exactly the asymmetries are, then the strategy of deriving the asymmetries from individual reports is a poor direction to pursue. One example of an asymmetry obtained from an individual's statement about beliefs is the following: I am sure that I am right about what my beliefs are, but I experience some uncertainty about the accuracy of the belief-attributions I make to someone else.

The advantages appear to be that, firstly, this version of the authority asymmetry does not rely on any theory for its accuracy and, secondly, it is likely to be endorsed both by philosophers of different persuasions and by most laypersons. Nevertheless, even though this approach helps us to overcome the theory-specific nature of the asymmetries, it bears the disadvantage of relying on something even more person-specific. Individual reports may give rise to very different construals of the asymmetries. The resulting asymmetries will depend on the amount of self-examination that people have engaged in and judgments that they have made about beliefs. Such asymmetries will once again rely on the unique theories that people hold about beliefs. Individual reports about beliefs and believing, accordingly, are likely to lead to asymmetries whose characters are person-specific in a way that places it beyond any method of independent investigation.[4]

In the context of my specific project, it is also worth observing that one of the chief motivating factors behind interpretationism is the avoidance of this intrinsically subjective, individualistic portrayal of the mental. If the asymmetries are derived from our subjective, individual reports of beliefs, interpretationists may be inclined to dismiss them for this reason alone. Obtaining the asymmetries from personal reports of beliefs and believing is, thus, not a good way of revealing our intuitions regarding the asymmetries.

University of Cape Town

1.2.2. THE ASYMMETRIES DERIVED FROM PRACTICE

The most promising approach to describe the asymmetries, as a result, seems to be to draw attention to some of our everyday epistemic and conversational practices (such as the giving and asking for reasons, the justification of beliefs and belief-claims, etc.) that demonstrate that we do not treat first- and third-person knowledge of, and stances towards, beliefs identically. This approach respects not only our need to keep the asymmetries theory-free, but it also keeps our intuitions independent of individual experiences or reports. Thus, whatever the exact nature of the theoretical asymmetries that philosophers derive from such practices, and whatever the nature of our individual reports of believing, a more feasible question is whether theories can explain the practices themselves.

I alluded to the advantages of the practical approach above. Firstly, critics may deny the everyday practices that I posit below, much as I deny first personal reports of believing. But, given the fact that such practices are publicly observable and participated in by all, they are harder to deny than personal reports of believing are.

Secondly, since the practices are harder to deny than information mined from the other two approaches, they are the least misleading sources of the asymmetries. My critics are not going to be able to reject the conclusion of the thesis purely because of rejecting a theory or a personal report from which the asymmetries are derived. The increased likelihood of the shared point of departure will contribute to the force of the conclusion.

In the following section I, therefore, intend to describe a number of our everyday epistemic and conversational practices that suggest that there are three ways in which we, in our ordinary interactions with each other, treat our own beliefs differently from those that we attribute to others.

University of Cape Town

2. THE ASYMMETRY THESIS

2.1. THE EVIDENCE ASYMMETRY

The first way in which we treat the beliefs of others in a manner different from our own, is the role that we allow verbal and behavioural evidence to play in our knowledge of which beliefs are held. When asked how I know that another person believes something, I usually cite aspects obtained from her environment, some inferences from her other beliefs and her verbal and non-verbal behaviour (hereafter just behaviour) as evidence for my suspicion that the belief is held. When asked how I know that I believe something, on the other hand, I may cite elements from my environment, infrequently but possibly draw inferences from some of my other beliefs, but it is unlikely that I will cite any information about my behaviour. In other words, I treat others, and they treat me, as if my observations of what they say and do are important when acquiring knowledge of their beliefs, while I treat myself, and others treat me, as if I can know what my own beliefs are without observing my own behaviour.

If I am asked what reason I have for claiming that Jane believes that a man is a murderer, I am likely to say that she believes it because she was startled and fled when coming upon him in the dark alley. If I am asked what reason I have for thinking that I believe that the man is a murderer, I am very unlikely to say that I think that I believe that the man is guilty because I observed myself being startled and fleeing from him when coming upon him in the dark alley.

In case 1, where I am asked to provide the evidence on the basis of which I achieve, or wish to acquire, my knowledge of the beliefs of someone else, a response is given effortlessly. It takes the form of citing or seeking the behaviour of others, together with some wider environmental evidence. Conversely, Case 2, where the evidence that gives rise to my knowledge of my own beliefs and desires is sought, I do not cite behaviour, not of my own or of others. In fact, in practice, such a question is very rarely asked, precisely because we do not expect that such behavioural evidence is used.[5] If I claim that I know that Jane believes that the man is guilty, I will be expected to provide at least some behavioural evidence for my claim to knowledge. It

University of Cape Town

is not only an appropriate question to ask, but it is, in reality, almost always asked when a third-person (like Jane) is discussed by two other persons. My interlocutors, on the contrary, are unlikely to ask me for behavioural evidence for my claim that I know that I believe that the man is guilty. Not only is it an inappropriate question for which there does not seem to be an adequate response, but it is, in reality, hardly ever asked.

Further, our interlocutors usually notice the switch from typically first-personal to behaviour-based self-attributions and typically treat them as less confident.

Let us imagine that the murderer in question is a speaker (S)'s husband. Now, the question about S's belief regarding a man's guilt is a question about whether she believes her husband to be guilty of murder. She declares that she does not think that her husband is capable of killing anyone because he is such a sensitive person. But her friends and family point out to her that she acts as if she believes that he is guilty. She still thinks that he couldn't possibly be a murderer, but she can understand that there is good evidence that she, as a matter of fact, does believe that he is guilty. This perplexing situation may compel her to assert that she knows that she believes that he is guilty, because she finds herself acting fearfully in his company.

This is a familiar case of emotional conflict which we are able to accommodate in our relations with others. Those with whom she interacts will understand her conflict and will be able to make sense of her evidence-based self-attribution. But they will notice that she is not relating to herself in the same way that people usually do, evidenced by the fact that they will ask why she is treating it as something that she would have attributed to anyone who exhibited some specific behaviour, instead of attributing it directly to herself in a specific circumstance. They will also assume that she is uncertain about what she believes. She says that she believes X, because behaviour Y (when engaged in both by herself and by others) would make sense only in the presence of a belief in X. She seems confused about what she believes, so she attempts to draw inferences from the same sort of evidence that she would have used when attributing beliefs to someone else. And instead of giving herself and others more confidence that she is sure of what she believes, the inference from her

University of Cape Town

behaviour to a likely belief will be taken to indicate uncertainty, rather than confidence, in the self-attribution.

Evidence-based self-attributions are also treated as inappropriate, or as if there is something wrong with the way the speaker relates to herself. The case above illustrates this clearly. It is improbable that a speaker will say that she concludes that she believes that a man is guilty because she observed herself acting fearfully in his company. And if she were to do so (such as in the case above), her interlocutors will immediately know that there is some sort of psychological conflict or unusually low level of self-awareness involved. They are likely to start questioning why she seems unwilling to attribute the belief to herself via some other method that does not involve behavioural evidence. And if they judge that she is incapable of reaching a conclusion about what she believes without consulting such evidence, her interlocutors will treat her as someone who has, for some reason, temporarily lost touch with herself.

Lastly, these differences in the way we treat first- and third-person beliefs is even more obvious when we move from the treatment of one individual evidence-based self-attribution to the handling of general self-attributions of this sort. As seen above, individual cases of evidence-based self-attributions occur infrequently and they are treated as being fairly peculiar. And if this is true for individual cases of evidence-based self-attribution, we should not be surprised that, in practice, we do not have a way of dealing with someone that generally self-attributes beliefs based on behavioural and verbal evidence.

University of Cape Town

2.1.1. THE EVIDENCE ASYMMETRY AND ORDINARY CASES OF BELIEVING

Some critics may argue that I chose the above case specifically because it leads to the conclusion I am attempting to defend. Such thinkers may try to construct more ordinary cases of evidence-based self-attribution to attempt to prove that we do not usually treat such self-attributions as uncertain or peculiar. What about a case where, for example, a friend asks me which chocolate bar I believe I like the best. I am sure that I do not believe that any one chocolate bar is tastier than the others, which is what I tell him I believe. He then suggests that I must believe that Tex bars are tastier than

other chocolate bars, since the last fifteen chocolate bars I bought included ten Tex bars. I then agree with him that I do believe it, since my behaviour suggests it.

But this is where the example collapses. In most cases the interlocutor will be surprised if I self-attribute the belief based on behavioural evidence only. My friend will probably continue to question whether his observation of my behaviour genuinely indicates that I believe what he thinks I do. He will expect me to think about it some more, arriving at a conclusion about my chocolate bar-related beliefs by some method other than observing my own past and present behaviour. Others can point out some of our behavioural patterns together with their interpretations of them, but whether we agree that the suggestions are accurate is expected to be settled in some other, first-personal, way.

2.1.2. THE EVIDENCE ASYMMETRY AND PAST BELIEFS

Some critics may try to argue that the difference between our treatments of the two perspectives is not significant in the case of the attribution of past beliefs.

But even most past beliefs are dealt with in this way. If a friend asked me what reason I have for thinking that Jane believed that the man was a murderer, I will still say that Jane must have believed it since she was startled and fled when coming upon him in the dark alley. If he asks me what reason I have for thinking that I believed that the man was a murderer, I am likely to respond in one of two ways. I will either tell him that I remember that I believed it and that I did not use my behavioural and verbal evidence to self-attribute it, or I might tell him that I cannot remember what I believed, but that I must have believed that the man was a murderer since I remember being startled and running off when coming upon him in a dark alley.

As the last response indicates, we do occasionally cite our behaviour as sources of evidence for what we believed in the past. But then the only difference between self-attribution of present and past beliefs is that occasional evidence-based self-attributions of past beliefs are likely to be treated more favourably than such self-attributions of present beliefs. They are not viewed as indicative of some psychological conflict or an unusually low level of self-awareness. If it is done on a

University of Cape Town

propositional attitudes are fixed by attributions from a third-person stance, it casts doubt on our ability to take up a first-person stance towards such attitudes. That is, it seems to imply that our knowledge of our own propositional attitudes will have to be obtained by adopting the position of a third person towards ourselves; that our self-attribution and revision of such attitudes will have to be achieved in the same way that we, as someone else's third-persons, achieve the attribution and revision of their attitudes; that we will have to relate to our propositional attitudes in the same way as we, as third persons, relate to those held by others. We do, however, by now know that the knowledge that we have, and the relation that we adopt, towards our own propositional attitudes is different from our knowledge of, and our relation towards, those of others.

The objective of this research project is to explore the relation between Donald Davidson's version of interpretationism and the first-person's perspectives of beliefs. Once again, since different propositional attitudes may pose slightly different problems, I will focus on beliefs specifically. The type of questions that will be addressed are the following: In chapter 2, I will investigate the nature of the differences between our knowledge of, and relation to, our own beliefs and our knowledge of, and relation to, those possessed by others. I will argue that three such differences can be derived from our common everyday epistemic and conversational practices, which I collectively call the asymmetry thesis. In chapter 3, I will describe Donald Davidson's influential version of interpretationism, called radical interpretation, and sketch his strategy for accommodating the asymmetry thesis in it. The common view is that Davidson's radical interpretationism cannot account for the asymmetry thesis, but I want to argue that this view is motivated by a misunderstanding of Davidson's strategy to accommodate it. I will, consequently, claim that his reconciliation strategy can be interpreted in two different ways: the standard interpretation of it, which I will call the meaning asymmetry, and my own interpretation, which I will call the sentence held-true asymmetry. In chapters 4 and 5, I will argue that, notwithstanding some moderate success in accounting for the asymmetry thesis, the meaning asymmetry interpretation cannot give a satisfactory account of it. In chapter 6, I want to offer my sentence held-true asymmetry interpretation, with the aim of arguing that it allows Davidson's radical interpretationism to accommodate the asymmetry thesis.

University of Cape Town

regular basis, though, observers are likely to treat them as being just as unusual and peculiar as regular present-tense ones. We can accommodate people who occasionally self-attribute past beliefs based on behavioural evidence, but we just do not easily accommodate a person who generally self-attributes past beliefs based on such evidence.

2.2. THE TRANSPARENCY ASYMMETRY

I treat my own beliefs as consistent with what in the world I hold true, while I treat the beliefs of others as rationalisations of their behaviour. Likewise, they treat my beliefs as states that make sense of my behaviour, while they treat their own as states that express their view of what is true in the world. That is why, from the third-person perspective, beliefs can be divorced from the question of their truth, whereas first-person attributions cannot come apart from an assessment of their truth.[6] If I related to my own beliefs primarily as causes of my behaviour, their truth would not matter to me.[7]

When self-attributing a belief, or when revising a belief attribution to myself, I invariably attribute something that is consistent with my view of some states of affairs in the world. If I claim that I believe that leopards are dangerous pets, I am thereby claiming that, from my point of view, leopards are dangerous pets. I do not attribute a belief to myself without taking a stand on the truth of something external to me. I simply will not claim that I have the belief about leopards without concerning myself with questions about what they eat, whether they behave aggressively, whether they can be taught to be submissive, etc.; that is, with the question of whether they are, in fact, dangerous. I, moreover, certainly will not attribute the belief that they are dangerous pets to myself while deciding that I think that they are quite harmless. My interlocutors will not ask me both what I believe and how I see the world since they know that my responses to the two questions will be the same.

When I attribute a belief, or revise a belief attribution to someone else, I do not concern myself with the truth of any proposition or state of affairs. If I claim that James believes that leopards are dangerous pets, I attribute this belief to him on the basis of his verbal reports of what he says he believes and his non-verbal behaviour

University of Cape Town

when visiting someone that keeps such pets, both interpreted in the context of what else I know that he believes. In such cases, I am not concerning myself with leopards at all. If I want to know what to attribute to him, I concern myself with questions like: "Does he have the tendency to be afraid of animals that bite and scratch?" "Does he own a leopard?" "Does he act responsibly around it?" etc. It is, accordingly, common for my interlocutors to ask me both the questions about what someone else believes and what I hold true, since my responses to the two questions may differ.

Lastly, if I change my view of an aspect of the world, I do not have to take on a separate task of revising my self-attribution. The self-attribution is revised involuntarily when I change my view of something in the world. If my interlocutors want me to change the belief that I attribute to myself, they simply try to change my view of the state of affairs that my belief is about.

2.3. THE AUTHORITY ASYMMETRY

While communicating with others, everyone assumes that we are usually more likely to be correct about our own beliefs than those with whom we interact.[8] In special cases, it may be possible for others to override our first person claim about what we believe, but, in general, it is accepted that our judgement of what we believe is a better indication of what we really believe than a judgement offered by an observer. Many of our practices can serve as evidence for this.

We expect to be treated as if we are usually correct about our own beliefs. We expect others to treat us in a way that acknowledges that they rely on us for information of what we genuinely believe, even in cases where they happen to know us well. This is demonstrated clearly by the fact that we become angry or argumentative when others do not respect our claim about what we believe. Imagine a situation in which I claim that golf is a boring game, while a friend uses the fact that I play golf every Saturday as evidence that I in fact believe that it is exciting. I argue that I do not play because I enjoy golf, but because I hope to meet new business partners. My friend may then cite my enthusiasm on the golf course as a reason to think that I do in fact believe that golf is exciting. I will respond that I fake the enthusiasm in order to interest enthusiastic businessmen. Let us assume that my friend concludes his side of the discussion by

University of Cape Town

claiming that, no matter what I think I believe, I actually do believe that playing golf is exciting.

In this situation, I am likely to become annoyed and either terminate the discussion and ignore the interlocutor, or become angry and argue that I am more likely to be right about what I believe than my interlocutor is since they are my own beliefs, based on my reasons to believe that I know better than anyone else.[9] If, however, my interlocutor does turn out to be right, everyone is likely to be surprised that I managed to misidentify my own belief, which provides another practice in support of the authority asymmetry.

Such debates occur infrequently, but when they do occur, the third-person usually takes a questioning, rather than a positive, stance. Exchanges are fairly common in which an interlocutor doubts that a speaker has identified her belief correctly and asks her for more information about her belief and her behaviour. The interlocutor's final confident claim that he is right, while the speaker is wrong, is highly improbable, though. And, if an interlocutor were to take this confident attitude that his conclusion about the speaker's beliefs is right, then he would be treated as if what he is doing is peculiar.

University of Cape Town

2.3.1. CHALLENGES TO AUTHORITY

Some critics may want to argue that it is not that uncommon for those with whom we have close long-term relationships to attempt to correct our understanding of what we believe. But the truth is that even then we are treated as authoritative because of the way in which they suggest beliefs that they think we hold, indicate their reasons for thinking that we hold them, request our own reasons for thinking that we do not hold them, ask for our reasons for behaving in ways that suggest that we hold them, seek our approval of their interpretations of our beliefs and other attitudes, attempt to change our view of our own circumstances and attitudes with the aim of bringing about a recognition or avowal of the beliefs they suggest to us as the ones we genuinely hold. I may not have been aware of my belief that Tex Bars taste better than other chocolate bars, but my interlocutor will not simply decide that this is definitely what I believe just because my behaviour indicates it. He will ask me to verify it. Or,

if I appear unable to verify it, such as in cases where I honestly do not know what I believe, then he is still likely to enter into a discussion with me during which he seeks my approval of what he is suggesting. He thinks that he is right about what I believe, but he continues to treat me as if I am the only person who can determine whether he is, in fact, right.

2.3.2. THE AUTHORITY ASYMMETRY AND PAST BELIEFS

My critics may try to object that the authority asymmetry does not apply to past beliefs, but all these practices are present in discussions about past beliefs just as much as presently held ones. A past-tense version of the golf-related discussion would have been very similar to the present-tense one. Once again, I would have become annoyed with my friend, he would not have taken such an aggressive stance and we would both have been surprised if he had been found to be right. If there is a difference between the discussions of past and presently held beliefs, it is the fact that the issue of a past belief might not be very passionately debated. But this does not show that the first- and third-person are treated as if they are equally likely to be right about a person's past beliefs. If it is indeed the case that the question would have roused less emotion, it could be ascribed to the fact that we care more about our presently held beliefs than about our past ones. There seems little point in convincing someone that he is wrong or right about a belief that has been abandoned or is no longer of great concern.

University of Cape Town

2.4. IMPORTANT THEORETICAL ISSUES

The evidence asymmetry is one that is clearly reflected in our practices, but it problematically rests upon some assumptions about the mind that the majority of people unquestioningly accept. When deriving the asymmetries from the ways in which people treat one another, this type of problem is unavoidable. Our practices are inevitably closely linked with the most common, intuitively pleasing, assumptions of the majority of us. And, as we have discovered often enough, our intuitions are sometimes mistaken. I said earlier that there clearly is an evidence asymmetry when we are asked how we know that we believe something, or when we are asked how we know that a third-person believes something. And this is, in deed, the way in which

the question is usually asked. But questions of this form assume that a belief already exists and they request information on the basis of which it can be discovered. The first-person is treated as if she does not rely on any evidence to know what her already established beliefs are, while the third-person is treated as if he cannot know what those already established beliefs are without making use of such evidence.

This asymmetry can, however, be interpreted as implying that the first-person is treated as if she retrieves her own beliefs directly from her own mind, while the third-person is treated as having to rely on behavioural and verbal information in order to guess at what might be in that first-person's mind. And this is, unfortunately, the way in which the majority of people interpret the concept of a belief. But this is precisely the Cartesian picture of the mind to which everyone has objected so convincingly throughout the past four centuries. I possess a collection of beliefs which I discover directly through introspection, while others have to use my behaviour and try to infer what those beliefs are. This picture of beliefs is not only rejected by interpretationists, but by many other contemporary philosophers.

Many theories towards the end of the twentieth century defended a view of self-knowledge whereby I come to know what my beliefs are through looking outward to my environment, as opposed to inward to my mind.[10] I come to know that I believe that it is summer by looking at the sun, feeling the heat, and so forth. Such theories are criticised for neglecting to distinguish between the question of how I obtain knowledge of what I already believe and the question of how I form a new belief. And that is precisely the point about which they disagree. Theories that rely on introspection for knowledge of our beliefs hold that beliefs are fixed, already established entities, while those who subscribe to the outward-looking claim argue that beliefs are states that are re-examined and re-established every time I become aware of them.[11] So, for the latter, the process of coming to know what I believe is the same question of establishing exactly what I believe, or of subjecting a belief to rational scrutiny, at the very least.

Interpretationists also reject the idea that there are beliefs already in place which it is up to everyone to discover. They also think that Beliefs are not fixed, introspectable states that exist inside the individuals to whom they belong. And, according to them,

University of Cape Town

the process whereby third-persons discover beliefs is not like a treasure hunt in which the participants attempt to find hidden articles by means of a series of clues provided by a game leader. Beliefs are, instead, no more than statuses that are attributed to us in a social context. There are no elements of the content of beliefs that cannot be fixed during communication. And, if this is the case, then the question of how one knows that one believes something, and the question of how one knows that others believe something, once again make little sense. After all, if something is called a belief only when it is attributed, then there exist no fixed set of beliefs that one can obtain knowledge about. Beliefs simply do not exist independent of one's attribution of them. Thus, for an interpretationist, the question of how one knows that one holds a belief is a question of how one attributes beliefs to oneself. And the question of how one knows that someone else believes something just becomes a question of how one attributes beliefs to others.

So, how can we square the evidence asymmetry (which is clearly present in our practices) with this interpretationist conception of beliefs? Most philosophers will ask why we should evaluate theories of beliefs according to whether they can accommodate the Cartesian picture if it is widely believed to be wrong. But most laypersons will ask why we should evaluate such theories according to whether they can accommodate this deeply theory-laden conception of beliefs. After all, if the main advantage of describing the asymmetries on a practical level is to avoid relying on theories that are not shared by everyone, then it seems important to hold on to our practices as a guide to concepts that almost everyone intuitively accepts.

I will, as a result, retain the evidence asymmetry as something that is present in the way in which we treat one another. We expect that all theories of beliefs are able to accommodate the idea that we treat third-person beliefs as knowable only through behavioural and verbal evidence while we treat first-person beliefs as knowable independent of verbal and behavioural evidence. But, it is important to keep in mind that, for an interpretationist, the evidence asymmetry pertains to how we attribute, as opposed to discover, beliefs, with the result that this asymmetry is not a response to an epistemological question of how we discover existing articles, but how we decide on, and attribute, them. (More about this in section 3.)

University of Cape Town

This relationship can be further qualified by the transparency asymmetry which deals with the question of how we form or attribute beliefs which, as explained above, interpretationists will see as the same question as how they are known. Where the evidence asymmetry holds that we know what our beliefs are without making use of behavioural evidence, the transparency asymmetry suggests that we form or acquire our beliefs by forming opinions about states of affairs in the world. If one is comfortable with a picture of beliefs as fixed states that await discovery through, for example, introspection, one will think that the first question differs from the second. The question of how I know what I do believe is different from the question of how I form a belief. If one does not subscribe to this picture of beliefs and one thinks that beliefs are, for example, states that I fix by an investigation of my environment, then there is no room for an epistemological question. So, the form of the evidence asymmetry that an interpretationist will want to account for simply holds that we do employ behavioural evidence when attributing beliefs to others, while we do not employ behavioural evidence when attributing beliefs to ourselves. And, in addition to this, they will have to be able to explain the transparency asymmetry, according to which we attribute beliefs to others in the service of an explanation of their behaviour, while we attribute beliefs to ourselves only if they are consistent with our view of the world.

University of Cape Town

2.5 THE RELATIONSHIP BETWEEN THE ASYMMETRIES

One way of thinking about the relationship between the three asymmetries is that we treat a first-person as if she usually knows what she believes (the authority asymmetry) because she self-attributes beliefs that are consistent with her view of the world (the transparency asymmetry) and does not have to draw conclusions from her own behaviour (the evidence asymmetry).

This relationship is clearer in situations in which we realise that the asymmetries are not present. A case in which we assume that the authority asymmetry is absent is usually a case where the speaker is unsure about what she believes, or where there is some evidence that suggests that she may be wrong about what she thinks she believes. Then we refrain from assuming that she is, by default, right about what she believes since there are reasons to believe that she is not. Many such cases are also

examples of instances where we assume that the evidence asymmetry is absent. The speaker has somehow lost her position of authority over her beliefs and we now treat her as if it is appropriate to engage in a discussion about what she believes by referring her to her behaviour. If the speaker turns out to be wrong about what she believes, and we convince her that she actually believes whatever we think her behaviour suggests, then we also have a situation where we assume that the transparency asymmetry is absent. She is wrong about what she believes, she accepts our behaviour-based suggestion of what she believes and we accordingly realise that she is self-attributing a belief in the service of an explanation of her behaviour, as opposed to one that is consistent with what she deems true about the world. The absence of the authority asymmetry in a specific case does, of course, not entail the absence of the other two, since a speaker can still reject our behaviour-based suggestions of what she believes in favour of whatever she thinks she believes without employing her behaviour as evidence. The three often do coincide, though.

For the reasons above, the link between the evidence and transparency asymmetry is even stronger. A case where a speaker has to employ her behaviour to know what she believes will almost certainly be a case in which she ends up self-attributing a belief in the service of an explanation of her behaviour. It might coincide with what she holds true, of course, but in such cases the truth of the belief that she ends up self-attributing is incidental since she is self-attributing it specifically in order to make sense of her behaviour.

University of Cape Town

3. THE IMPORTANCE OF THE ASYMMETRY THESIS

The position endorsed in this project is that theories of beliefs that can accommodate these practices are more attractive than those that cannot do so. This stance does require some justification, especially in the light of a scepticism regarding intuitions and intuition-reflecting practices that has been gaining popularity in recent years. I am sympathetic to the scepticism of the role of intuitions and intuition-reflecting practices (hereafter just intuitions) in Philosophy, but I still contend that they do have a role in this project.

There are at least three positions on the role of intuitions in Philosophy. First, any philosophising is constrained by intuitions. Second, it is desirable for Philosophy to cohere as much as possible with intuitions, unless there are good reasons for suspecting that they are mistaken, in which case it is acceptable for our theories to depart from them. Third, intuitions are irrelevant to Philosophy and our theorising does not have to take them into account.

I want to reject the first claim. Our theorising cannot be constrained by intuitions since they are difficult to interpret. We often appear to have conflicting intuitions and they are largely dependent on the specific circumstances that are sketched to draw them out. Intuitions are the products of socio-historic circumstances and they often change. It is not uncommon for two persons' intuitions to differ, or for one person's intuitions to differ from one time to another. They are, moreover, probably not entirely theory-independent themselves since their development is not isolated from the development of an individual or a group's theories about the world. Intuitions should, thus, not be treated as the ultimate given in Philosophy.

I also want to reject the third claim. All theorising has to start somewhere, with the result that this approach's pretence not to take intuitions into account is a fiction. Moreover, if it had been possible not to take them into account, the theories resulting from such theorising will be wildly unbelievable and largely irrelevant to our lives and how we understand them.

I, thus, want to embrace the second claim, in accordance with which we are neither reifying (1), nor ignoring (3), our intuitions, but using them as a useful starting point for our theorising. We should treat them as potentially fallible, and if a specific intuition leads us to theories that we think are implausible, we should show why it is mistaken and why it should be discarded.

This approach is preferable to any other, since it is likely to be accepted by proponents of schools of thought as different as Philosophy as a substantive discipline and philosophical quietism. According to proponents of Philosophy as a substantive discipline, the aim of Philosophy is to arrive at theories that resolve things that we find puzzling. Intuitions provide us with a starting point for our theorising, but our

University of Cape Town

conclusions are not constrained by them. Many theory theorists, for example, think that our common, everyday folk psychology is a good place from which to launch our theorising about beliefs (and other propositional attitudes), even if our resulting philosophical theories will modify such folk theories in the light of scientific evidence and further theorising. (See below). According to proponents of philosophical quietism, the purpose of Philosophy is not to theorise, but to describe. It should provide an explanatory narrative of how we, as a matter of contingent fact, came to be in a position where we found something puzzling. Our theorising is responsible for causing us to be in situations where we find things puzzling, and such explanatory narratives will serve as a therapy to return us to a situation where we did not find it puzzling. Philosophy should, in other words, describe both the practices and intuitions that allowed us to live without finding things puzzling, and describe the theorising that confused us. Proponents of both these conflicting schools of thought are likely to endorse my view on intuitions defended above, which is a good reason to accept it.

Theories of beliefs that can accommodate the practices in section 2 are, thus, more attractive than theories that cannot do so, unless they can show why our intuitions are mistaken. I am, accordingly, amenable to the possibility that a theory of beliefs does not have to leave room for all the asymmetries. As I explained above, I find it acceptable to allow interpretationism, for example, to modify the evidence asymmetry to apply to the attribution, as opposed to the discovery, of beliefs, since I find the arguments against beliefs as internal states persuasive. The evidence asymmetry, as it occurs in practice, relies on some intuitions that there are many good reasons for rejecting. And this is a strategy that I stand by in the case of all theories of beliefs. If they can justify an alternative description of one or more of the asymmetries by, for instance, persuading us that the people who participate in the practices are wrong to do what they do, or to think what they think, then I will be more than accepting of the resulting re-formulated asymmetry thesis, or even of a rejection of it.

4. FURTHER CONSTRAINTS

Not just any explanation of the asymmetry thesis is satisfactory, however. In this section I want to point out some characteristics of the asymmetry thesis that are likely to be overlooked but that are, nevertheless, worth accounting for, and argue that there

University of Cape Town

should be certain constraints on such explanations. Once again, I will adopt the approach above according to which theories that can account for, or that are constrained by, these characteristics, are preferable to those that cannot do so.

Firstly, it is desirable for theories to be able to allow for the idea that we treat others as if they usually do not employ their own behaviour and speech as evidence to know what they believe, as if they usually self-attribute beliefs that are consistent with how they see the world and as if they are usually right about what they believe. These three general assumptions hold during all interactions. But a theory that is genuinely capable of allowing for the asymmetries should also be able to accommodate situations in which, even though these general assumptions hold, we still realise that someone is using her own behaviour to know what she believes, self-attributing a belief in the service of an explanation of her behaviour or possibly wrong about what she believes.

Secondly, it is desirable for theories to be able to account for the asymmetries in practice, as opposed to only in theory. Theories that propose that something can be explained once we place ourselves in a possible world or in a fabricated set of conditions are useful to extract our intuitions regarding that thing, but they are less useful when we want a theory that explains what we, as a matter of fact, do in practice. Davidson's system of radical interpretation, as will become evident in chapter 3, is such a theory. A theory that is able to explain what we do in practice is preferable to one that can only set out a list of conceptual requirements and constraints that can never obtain.

Thirdly, a theory that can allow for a unified explanation of the asymmetries is preferable to one that accommodates only a haphazard or clumsy explanation of them. We prefer explanations that are intelligently designed to explanations that are forced into spaces where they just do not quite fit, even if they are able to clumsily explain whatever needs to be explained.

Lastly, explanations of the asymmetries that are constrained by univocality are preferable to those that are not. We want there to be only one concept of belief, with two different perspectives from which we can relate to it. Theories that deny

University of Cape Town

univocality are theories that suggest that there are two different concepts of belief, one that gets its meaning from the way in which a first-person relates to it, and another that gets its meaning from the way in which a third-person relates to it. Thus, when I refer to my own belief that reality television is tedious, I am talking about something different from my reference to someone else's belief that contains the same proposition. However, such theorists are going to have to explain our strong intuition that, in such cases, we are talking about the same sort of state that constitutes a commitment to the same sort of thing. Furthermore, the first-person and third-person often reach agreement and ascribe the same belief to one individual. But if univocality is denied, then we can never be said to ascribe the same thing, not even after discussion and apparent agreement between two people. Most theorists would want to avoid these types of problems.[12] Hence, my suggestion that explanations of the asymmetries that are constrained by the univocality of beliefs are preferable to those that are not.

I want to briefly illustrate these desired requirements and constraints by questioning two popular theories' ability to account for them. My aim is not to refute these theories, to claim that they cannot account for these points or even to give a detailed account of how their proponents have attempted to account for them. My aim is to suggest that the surface appearance of these theories make it appropriate to express some doubts regarding their ability to satisfy these points.

Cartesian dualism, according to which beliefs are non-physical states housed in a non-physical mind, maintains that the asymmetries are a product of the mind's ability to monitor and be aware of its own states. The fact that the mind is not a physical entity means that it is not accessible to a third-person, with the result that the first-person's judgements alone constitute her beliefs. That is, something is a belief if the first-person understands herself as having it. We treat a first-person as if she usually knows what she believes because, being her own beliefs in her own first-person accessible only mind with its monitoring mechanism, whatever she thinks she believes is, as a matter of fact, what she does believe. We treat her as if she does not have to use her own behaviour as evidence to know what she believes because the mind's monitoring mechanism gives her immediate access to her beliefs. The monitoring mechanism allows for a unified explanation of these two asymmetries since it allows the theory to

University of Cape Town

use the same mechanism to explain both. It, lastly, can explain the asymmetries in practice, since it takes the fact that we treat each other as such as a result of the fact that we are always right about what we believe, which is explained by the fact that the monitoring mechanism allows for immediate, infallible and omniscient access to our beliefs.

Cartesian dualism does, however, seem to be defeated by the other points. It, firstly, does not seem able to explain the transparency asymmetry, since it seems to commit itself to the view that my identification of my belief is an identification of a state that is present in my mind. The transparency asymmetry, however, requires that my self-attribution of a belief commits me to the truth of something outside myself. So Cartesian dualism is challenged to explain how an identification and description of an internal state through a self-monitoring mechanism can commit me as such. Even more problematically, if this mechanism is geared towards identifying internal states, then why does it not make it possible for a speaker to describe one of her beliefs without taking a stance on the truth of something outside herself? The monitoring mechanism just does not seem to fit with the transparency asymmetry and no Cartesian dualist, to my knowledge, has produced a reason for rejecting such practices.

Secondly, many philosophers have criticised Cartesian dualism for its inability to accommodate cases where speakers are shown to be wrong about what they believe. If a speaker's judgement of what she believes just is what a belief is, then whatever she judges herself to believe must be what she genuinely believes. I do not intend to labour this point any further here, except to add that if it has a problem accommodating such cases, it is likely to have a problem to accommodate cases in which speakers have to use their own behaviour as evidence to know what they believe. Even if they can produce an explanation for occasional failures of the monitoring mechanism, her mind, since it is not physical, is not accessible via any other means. If a third-person's observations of the speaker's behaviour can only be accurate if she confirms them to be correct, then it does not seem as if she can know whether she is making accurate judgements of what she believes when employing her own behaviour.

University of Cape Town

Thirdly, Cartesian dualism is not constrained by univocality, due to its commitment to the mind as a non-physical entity. Since this renders a speaker's mind inaccessible to her interlocutors, each person will have to fix the meaning of the concept of belief for herself.

Thus, Cartesian dualism, on the surface, seems to rank low on our scale of desirability since it, firstly, fails to explain some of what I want theories to be able to explain and, secondly, fails to give reasons for thinking that those things do not have to be accounted for.

The second theory that I want to put through the superficial test of desirability is the theory theory, held together with the background assumptions of functionalism and the representational theory of mind.

According to the theory theory, our ability to attribute "mental states" is an ability to draw inferences from a Theory of Mind. If our Theory of Mind, for example, includes the datum, "ceteris paribus, those who believe that leopards eat humans will run when they encounter a leopard", I will attribute the belief that leopards eat humans to a person that I spot running away from one. I know that this is what she believes because that information is contained in my Theory of Mind.

Many theory theorists, in addition, believe that our folk Theory of Mind is reasonably accurate, if incomplete, and that a scientific investigation of the mind will build upon this folk Theory of Mind to show where it is scientifically correct and incorrect. Most such theorists think that such scientific investigations should proceed via an examination of the functional roles of mental states. They, in other words, defend functionalism, according to which something is a belief in virtue of its functional role in the system of which it is a part. This functional role is determined by what they are caused by, their patterns of relations with other mental states and their typical outcomes. The belief that leopards eat humans, for example, is typically caused by relations with certain environmental events, by certain types of conversations with others, and so forth. It typically relates to other mental states by, for example, making the speaker more likely to have certain desires, fears, hopes, and so forth. Together

University of Cape Town

with such relations with other mental states, it typically has certain behavioural consequences like the running away from leopards.

Many theory theory-functionalists, lastly, defend a representational theory of mind according to which beliefs are functional states that relate a speaker to the symbolic representations of the content of those states. "Leopards eat humans" is the semantic content of a symbolic mental representation. If a speaker has the belief that leopards eat humans, he is in the functional relation (that is typical of a belief) to a symbolic mental representation with the semantic content "leopards eat humans". He can have another symbolic mental representation with that same semantic content (leopards eat humans), but with the functional role of, say, fear. In the popular literature these functional roles are usually referred to as one's belief box, one's fear box, etc., not to indicate a physical location in the brain, but to identify the functional role of everything in that box.

To impose order on this theory stew, theorists who defend all three these theories (Jerry Fodor, Shaun Nichols, Stephen Stich, etc.), usually believe that a scientific psychology will show that most of our commonsensical folk theory of mind is correct, that it will show that something is a belief in virtue of the specific functional role that it has and that it will show that the content of such functional role states is symbolic mental representations. Now, if a belief just is a specific type of functional role (a belief box) with a symbolic mental representation as its semantic content, and if the only way of non-scientifically accessing it is via our Theory of Mind, then, whether it belongs to ourselves or to others, it must be so accessed. We know what we believe in the same way as we know what others believe, namely, by drawing theory-mediated inferences from our behaviour (which includes our speech).[13][14]

We treat a first-person as if she is usually correct about what she believes since she has observed more of her behaviour than anyone else has, and can thus draw more accurate conclusions about her beliefs via her Theory of Mind. We treat her as if she can know what she believes without having to use her behaviour as evidence because the information on the basis of which she knows what she believes includes her memory of past behaviour. She, thus, can often know what she believes without employing her behaviour as evidence since she employs memories of her behaviour

University of Cape Town

instead. The theory theory can explain situations where we judge that the asymmetries are absent, because the theory-mediated inferences that we often have to rely on are far from infallible. Moreover, since our folk theory of mind may not always cohere with scientific evidence, the correct answer of what we believe will sometimes not be available to us. The theory theory can explain the asymmetries in practice, since it treats our practices as a result of the fact that the first-person will always be more closely acquainted with herself than anyone else can ever be. It gives a unified explanation of the asymmetries since it relies on our Theory of Mind to explain both asymmetries, and it respects univocality since both first- and third-person beliefs are defined by their role in a theory.

It, however, cannot improve much upon Cartesian dualism's attempt to explain the transparency asymmetry since it is committed to the view that our Theory of Mind correctly picks out functional states with symbolic representations as semantic content. The same questions as in the Cartesian case arise. The transparency asymmetry requires that my self-attribution of a belief commits me to the truth of something outside myself. But how can theory theorists explain that an identification and description of a functional state through a Theory of Mind can commit me as such? Once again, if this Theory of Mind is geared towards identifying functional states, then why does it not make it possible for a speaker to describe one of her beliefs without taking a stance on the truth of something outside herself? Since the functional role of a belief is partly defined in terms of what causes it, it has a slight advantage over Cartesian dualism. A functional state is not as wholly internal as a Cartesian non-physical mental state. But it is still not clear how describing a functional role commits me to the truth of something, as opposed to simply being a description of, amongst other things, the thing that caused it. The identification of a functional state with a symbolic mental representation as content just does not seem to fit with the transparency asymmetry and no theory theorist, to my knowledge, has produced a reason for rejecting such practices.

University of Cape Town

It is also important to point out that the explanation of the evidence asymmetry is inadequate since it, firstly, subscribes to the questionable claim that our memories contain sufficient information of past behaviour to derive our beliefs from and, secondly, will as a result have to admit that we quite frequently do rely on our

behaviour to know what we believe. Theory theorists have not produced a reason for rejecting the evidence asymmetry, and my modification of it cannot help since they insist on beliefs as functional states with mental representations as content.

Therefore, the popular combination of theories described above, on the surface, seems not to rank much higher than Cartesian dualism on my scale of desirability unless it is supplemented, either with a persuasive rejection of our practices or with an explanation that can be constrained by the above points.

5. CONCLUSION

In section 1, I provided reasons for deriving the asymmetries from our practices. In section 2, I described such practices and derived from them an evidence, a transparency and an authority asymmetry. In section 3 and 4, I proposed some desired requirements for, and constraints on, explanations of the asymmetries and provided reasons for preferring theories that can account for them over theories that cannot. In the remainder of this project I will consider which of these conditions Donald Davidson's system of radical interpretation can satisfy.

University of Cape Town

CHAPTER 3
DAVIDSON'S SYSTEM OF RADICAL INTERPRETATION

In this chapter I intend to set out the problem at the heart of the rest of this project. I shall do this by, firstly, providing a sympathetic exegesis of Davidson's arguments for the social constitution of meaning and belief content, secondly, describing his system of radical interpretation, thirdly, outlining the prima facie incompatibility between radical interpretation and the asymmetry thesis and, fourthly, explaining the aspect of an interpretationist system that, according to him, makes it compatible with the three attitude asymmetries.

1. TWO ARGUMENTS FOR THE SOCIAL CONSTITUTION OF BELIEF CONTENT

Donald Davidson's theory of interpretationism is one of the most complete and influential interpretationist accounts. His project was motivated by his conviction that semantic concepts like meaning and belief could never be made sense of outside a context of interpersonal interaction. That is, something can qualify as the meaning of a person's utterance and it can be correctly termed a belief only if it is established as such during the process of one person's interpretation or understanding of another.

University of Cape Town

Before proceeding to his arguments for this claim, I want to clarify a few definitions. The heading of this section is termed "the social constitution of belief content". By "social", Davidson has in mind that belief content needs to be established during some interaction between two people. He wants to deny that such content can be established by one person in isolation. By "constitution", Davidson literally means that the content is developed, or takes shape, during such social interactions. In the paragraph above, I somewhat misleadingly used the word "establish", which can be understood in a constitutive or in an epistemic sense. As explained in chapter 2, the epistemic sense, which is the more common way of thinking of beliefs, relates to how we know what someone believes while the constitutive sense relates to what exactly beliefs are. Davidson rejects the epistemic sense, since he does not think that beliefs already exist that are up to an interpreter to discover. He firmly supports the constitutive sense, in

which belief content literally is something that takes shape during social interactions. In the remainder of this project, terms like "establish", "determine", "settle", and the like, should, accordingly, be understood in a constitutive way. Lastly, the reader should note that we are here talking about the constitution of belief content and thereby about the constitution of beliefs themselves. Some philosophers draw a distinction between the two and may think that Davidson's interpretationism is aimed merely at establishing or individuating belief content. Davidson, however, ties content individuation fundamentally to beliefs themselves, with the result that the social constitution of belief content is also an argument about the social constitution of beliefs themselves.

He offered two arguments in support of the social constitution of belief, one that stems from considering our folk psychological explanations of behaviour as a metaphysical account of mind and one that goes via the publicity of meaning and its intricate connection with belief content.

University of Cape Town

1.1. THE ARGUMENT FROM FOLK PSYCHOLOGICAL EXPLANATION

Davidson's more straightforward argument for the idea that belief contents need to be established socially is an argument from folk psychological explanation. His point of departure was the observation that the way in which we usually make sense of the behaviour of others is through a process of interpretation whereby we attribute propositional attitudes that rationalise their behaviour. The Cartesian account of mind can allow for the view that beliefs rationalise behaviour, but according to this traditional picture, the issues of whether a belief is held, what belief is held and what belief can render which behaviour rational, are settled from the first-person's point of view. A belief has the content that it has because the first-person understands herself as having it.

Wilfred Sellars is commonly understood as being one of the first philosophers to question the identity of a belief as something established by the first-person's introspective knowledge of her own mind.[15] He calls the idea that beliefs are directly given to us through introspection the myth of the given, and proposes an alternative story that can account equally well for the mental and our intuitions regarding first-

person access to the mental.[16] He asks us to imagine a situation in which our ancestors, who initially understood behaviour in purely behaviouristic terms, acquire a new theory of behaviour that posits inner episodes as the causes of overt behaviour. At first they apply this new theory only to make sense of the behaviour of others, but then they discover that the theory can be extended to make sense of themselves as well by attributing beliefs to themselves upon observing their own behaviour. Their achievement is complete when they realised that their application of the theory is becoming so skilled that they often manage to self-attribute beliefs without being aware of applying a theory at all, even when they are unaware of applying a theory when attributing (or self-attributing) beliefs, such beliefs remain the postulates of a theory of mind. This Sellarsian myth gave rise to what is now popularly known as the theory theory of folk psychology. Beliefs, on this account, acquire their identity from their role in the rationalisation of behaviour.

Davidson does not accept the whole of what has come to be known as the theory theory. He denies that folk psychology is the same sort of theory as physics and biology, for instance, because he thinks that no theory that involves psychological terms like desires and beliefs can contain the type of law-like relationships found in other theories. [17] But this does not prevent him from accepting one of the main themes in the Sellarsian myth: the conception of a belief as a construct defined by its place in a commonsense framework for explaining (and even predicting) behaviour. He rejects the intuitive notion that, while others interpret our behaviour and attribute beliefs and desires to us to rationalise it, our genuine attitudes can be established in some other way such as through introspection or a neurophysiological study of the brain. Consequently, if the concept of a belief is exhausted by its role in making sense of people's behaviour, then there can be no element of a belief that is not available to a third-person interpreter that is observing and rationalising their behaviour. That is, if beliefs are no more than the sort of things that are attributed in order to make actions intelligible, then an investigation of the best method of attributing them will reveal all there is to know about them.[18] In Davidson's words, "talk apparently of thoughts and sayings ... belongs to a familiar mode of explanation of human behaviour and must be considered an organized department of common sense which may as well be called a theory. One way of examining [the nature of thought and language] is by inspecting the theory implicit in this sort of explanation."[19]

University of Cape Town

1.2. THE ARGUMENT FROM THE NECESSARY PUBLICITY OF MEANING

Davidson's second argument for the social constitution of belief content is an argument from the publicity of meaning. According to Davidson, "meaning, and by its connection with meaning, belief also, are open to public determination."[20] He probably decided on this approach because it is often easier to convince people of the public determination (or more precisely the social constitution) of meaning than of belief.

1.2.1. THE NECESSARY PUBLICITY OF MEANING

University of Cape Town

According to Davidson, meaning is necessarily public because the only way in which language can have semantic content is for that content to be determined during interaction with other language speakers with whom one can agree or disagree about the existence and nature of the environment that speakers claim to talk about. So, without communication between at least two people who are talking about their mutual environment, language cannot have semantic content at all because it would have no way of establishing the objectivity of the things it talks about. Consequently, if there are objects that are meant to give phrases their semantic content (or elements of the meaning of a phrase), that cannot be understood by an adequately equipped and informed person with whom we interact, then such objects (or elements) cannot play any role in the meaning of that phrase.[21,22,23] Thus, if I use a phrase to talk about something in my environment, it will always be possible for someone in favourable circumstances who is equipped with sufficient information to settle the matter of what I mean by that phrase.

More precisely, Davidson employs the notion of a Tarskian truth theory, rather than that of reference, in his theory of meaning. He believes that the notion of reference cannot be used in a theory of meaning, since an explanation of it would have to be given in non-linguistic terms, which is impossible. In his own words … "If the name 'Kilimanjaro' refers to Kilimanjaro, then no doubt there is some relation between English (or Swahili) speakers, the word, and the mountain. But it is inconceivable that one should be able to explain this relation without first explaining the role of the

words in the sentences; and if this is so, there is no chance of explaining reference directly in non-linguistic terms."[24] His dismissal of reference as the source of meaning makes him reject any theory of meaning that gives the meaning of a word or phrase by relating it to an object to which it is meant to refer. Instead, he proposes that a theory of meaning should consist of theorems that relate a speaker's sentences to the sentences her interpreter uses to specify the conditions under which her sentences are true.

This means that the meanings of a speaker's phrases have to be specified by sentences in her interpreter's idiolect that describe the truth conditions of her sentences as he sees it. Take, for example, a situation in which an English-speaking teacher is communicating with a Xhosa-speaking student. The teacher says, "It is sunny" while pointing up to the sun. The student tries to discover what the teacher means by this phrase by judging which environmental conditions will make the sentence true and translating those environmental conditions into his own idiolect. The student says, "It is sunny if and only if it is cloudy" while pointing skyward on a cloudy day. The teacher will indicate that the student has misunderstood what he means by the phrase, and will continue to do so until the student, according to the teacher, applies the phrase in the presence of the same type of environmental event that the teacher applies it in. When the student, for example, says, "it is sunny if and only if it is sunny" while pointing up at the sun, the teacher will judge that the student understands his sentence correctly and has translated it correctly into Xhosa. What the teacher means by this phrase is, thus, determined by the environmental conditions under which the teacher's sentence has been judged to be true by his interpreters. What the teacher means when uttering the phrase is conveyed to the student by the environmental conditions under which the student judges his sentence to be true. This picture applies even when the student misunderstands what the teacher means when uttering this phrase. Then what the student means when uttering this same phrase is determined by the environmental conditions (such as a cloudy sky) under which the teacher judges his sentence to be true.

This approach serves, not only as a way to avoid the concept of reference, but also to respect the holistic nature of meaning. Davidson does not think that the meanings of sentences can be attributed one by one on the occasion of their use. He holds firmly

University of Cape Town

onto the idea that meaning is holistic; that is, some sentences depend for their meaning on their connections with other meaningful sentences that, in turn, depend for their meaning on other meaningful sentences, and so forth, until a whole language with its interconnected meanings have been attributed. Clearly, the only way of attributing a whole language with interconnected meanings is to map a language onto another language with its own interconnected meanings. That is why the only way of getting to the meanings of a speaker's sentences is for an interpreter to map it onto his language so that their meanings can be holistically established by relating them to sentences, and the relationships between those sentences, in his language. (Much more about this in chapters 4 and 6.)

If what someone means when uttering a word is determined in any way other than by sentences in an interpreter's idiolect that specify the environmental conditions under which he judges the speaker's sentence to be true, it will lead to the same type of problems encountered by the traditional Cartesian theories. Davidson worried, in particular, about scepticism of both the existence, and the nature, of an external world. If a part of what the teacher means by his phrase is determined by, for instance, a picture in his subjective, first-person accessible-only mind, a physiological design in his brain or a pattern of stimulation in his sensory receptors or in his nervous system, then questions like the following are always appropriate: is he talking about real objects/events in an external environment? Or, is he talking about the correct objects/events in his environment?[25,26]

If we claim that the question of what a person means by a phrase can be settled by asking him which essentially first-person accessible mental picture he entertains when he utters the phrase, then we run into the question of how it can ever be established that such an insulated, private and subjective mental picture is about something in the world at all. If it is before the mind in the sense that it is immediately graspable by it, then there is no way of knowing, firstly, whether that picture relates the person to conditions in the world at all and, secondly, whether the words of others are uttered in approximately the same conditions as our own. We speak as if our words are applied to something objective, but if what we mean by a word is derived from a mental object, then its subjectivity prevents us from taking the first step in determining whether it corresponds to what it purports to represent. The problem with deriving

University of Cape Town

what a person means by her words from a mental picture does not end here, though. Davidson also points out that the very notion of such mental objects/pictures is implausible. If they are meant to be inner objects that are private and immediately accessible to the person to whom they belong, then they cannot also claim to be objective in the sense of representing objects in a shared environment. In his own words, "The only object that would satisfy the twin requirement of being 'before the mind' and also such that it determines the content of a thought must, like Hume's ideas and impressions, 'be what it seems and seem what it is'. There are no such objects, public or private, abstract or concrete."[27] Thus, theories that allow that what we mean by our words is determined by asking the first-person which mental object she is entertaining, give a highly implausible account of the source of the semantic content of language.

If, on the other hand, we claim that the question of what someone means by a word can be settled by examining the pattern of stimulation of their sensory receptors or nervous systems, then we run into the problem that it simply wouldn't be possible to determine who is right or wrong about the world.[28] To illustrate, let us imagine a situation in which one person, when a horse canters by, has the same pattern of stimulation that another person has when a cow is in view. In the presence of a horse, person 1 will say "there's a cow", which person 2, since their patterns of stimulation are identical, will interpret as meaning "there's a cow", even though she is sure that she is seeing a horse and almost always responds "there's a horse" when under those environmental conditions. Both people will think that they are right, especially when all their other perceptions and responses confirm this. Introducing a third person into the scenario to determine who is right cannot help for, if person 3, in the presence of horses and cows, has the same patterns of stimulation as person 2, the best we can then do is to say that person 1 has unusual or abnormal patterns of stimulation. We cannot say that her patterns of stimulation are wrong or inappropriate under those environmental conditions, because then we are implicitly appealing to the nature of something in the environment (as agreed upon by persons 2 and 3) to verify that their patterns of stimulation are more appropriate to that thing than person 1's. Thus, theories that allow that what we mean by our words is determined by investigating patterns of stimulation in our sensory organs or nervous systems cannot give a plausible account of how language acquires its semantic content, because they can

University of Cape Town

provide us, firstly, with no hint of the nature of the conditions that are meant to give language such semantic content and, secondly, with no acceptable notion of truth and falsity.

Instances of astigmatism, colour blindness, deafness and other sensory abnormalities together with, of course, evil demons and brains in vats controlled by mad scientists all ground the argument for the necessary publicity of language. The moment we try to derive semantic content from something other than the conditions in a common environment that people derive the truth conditions of each other's sentences from, we are, in effect, inserting additional information between the world and the semantic content; information that is meant to serve as evidence for what our sentences derive their semantic content from. And, if we do this, then we cannot extricate ourselves from the type of sceptical questions above. We would not be able to say whether the extra information is evidence for what it purports to be evidence for, namely, events in an external world. Thus, in Davidson's own words, "Without other people with whom to share responses to a mutual environment, there is no answer to the question what it is in the world to which we are responding."[29] And if we cannot even establish the existence and the nature of the objects to which we are meant to be responding, then our language simply cannot have semantic content.

University of Cape Town

1.2.2. THE CONNECTION BETWEEN MEANING AND BELIEF CONTENT

The next stage of Davidson's second argument for the social constitution of belief content is his claim that belief is inextricably linked with meaning. If we accept the above argument for the publicity of meaning, and we accept the argument that follows for the link between meaning and belief, then we should also acknowledge that belief contents are socially established. Belief content is inextricably linked with meaning. According to Davidson, a belief just is a meaningful sentence that the speaker holds true. Alternatively, a belief is a state of mind that is attributed by making use of a sentence that the speaker holds true. And, from the above argument, we know that the sentence that the speaker holds true will have to rely on the conditions in the environment that a third-person interprets the sentence as being about for its meaning. Now, if a belief is a meaningful sentence that the speaker holds true, or a belief is the sort of state that is attributed by giving a sentence that the speaker holds true, and the

meaning of the sentence that the speaker holds true is given by the conditions in the environment that they are interpreted as being about, then the meaningful content of the speaker's belief is given by the same correspondence between her application of her words, her interpreter's application of his words, and their reactions to each others' applications. If the semantic content of our sentences is derived from the conditions in our environment under which our interpreters judge our sentences to be true, and our beliefs have the same semantic content as our sentences, then the semantic content of our beliefs has to be derived from the same conditions in the environment under which our interpreters judge our sentences to be true. And if the semantic content of our beliefs is derived from the conditions under which others interpret our sentences as being true, then it means that our belief contents are publicly determined.

In the argument for the publicity of meaning, we want to say that different people talk about the same objects, namely, objects in the environment. If the objects are not in the environment, or somehow constituted socially, then we violate the requirement that language has to be public without which, as seen above, language can have no semantic content. And the only way of confirming that we are talking about something in the environment, as opposed to something in our bodies or minds, is to know that we are talking about the same things that others are talking about. This same argument applies to belief. We want to say that different people hold beliefs about the same things, namely, things in the environment. If they were not in the environment, or somehow constituted socially, then we violate a requirement that the objects that we hold beliefs about have to be public. For exactly the same reasons as in the case of the semantic content of language, without the public determination of the objects that we hold beliefs about, beliefs can have no semantic content. And the only way of knowing that our beliefs are actually about something in the environment, as opposed to something in our bodies or minds, is to know that we hold beliefs about the same things that others have beliefs about. As Davidson wrote, "The answer to what object a person is applying a word to, or having a belief about, cannot be given outside a context where a second person is responding to the first."[30]

University of Cape Town

2. RADICAL INTERPRETATION

2.1. PRELIMINARIES

The above two arguments support Davidson's conclusion that contentful beliefs are the sort of things that can only be socially established. If there are no more to the concept of belief than what others can attribute to us based on what we say and do, then it will not be possible to have a belief that someone with the right sort of information who follows the right sort of method of attribution cannot attribute to us. Hence, an investigation of the appropriate method by which they are attributed can tell us all there is to know about them. The problem is that, in ordinary circumstances, we share at least some language with the people that we are trying to understand. So, if it is true that our beliefs involve the same objects in the environment as our language, then our interpretive practices that involve a shared language will give us the sort of semantic knowledge which serves as a short-cut to people's beliefs. But Davidson's system of attribution (or interpretation as he termed it) is supposed to tell us how such semantic knowledge is obtained. This means that we cannot use our everyday interpretive practices as a description of an appropriate method of interpretation, because such everyday practices already assume knowledge of beliefs and meaning.

This is why Davidson wants us to imagine the predicament of an imaginary character, called a radical interpreter, who wants to make sense of the behaviour (verbal and non-verbal) of someone with whom she shares no language. Without ascribing meaning and beliefs through a process of interpretation, she cannot know what any of his behaviour means, with the result that everything he does will be completely mysterious to her. She, hence, needs to employ a method of interpretation that will give a satisfactory explanation of his behaviour and, because belief and meaning are the sort of things that rationalise behaviour, a method of interpretation that explains behaviour adequately will give a satisfactory account of what the foreigner means and believes. Thus, nothing other than third-person attributions can fix the beliefs that we hold, but this method of belief constitution can be accurate only if the process of interpretation whereby they are attributed is appropriate.

University of Cape Town

Let us, thus, examine Davidson's suggestions for an appropriate method of interpretation in order to clarify the constraints placed upon our attributions; that is, the rules that we have to apply in order for our third-person attributions to qualify as fixing someone's propositional attitudes accurately.

When I interpret, say Jane, I observe her utterances, her actions and her surroundings, and employ this information to assign meaning to her utterances and attribute propositional attitudes to her that render her behaviour intelligible. My evidence, hence, consists of her utterances, her behaviour and her environment, which I probably share. Davidson believes that I can use this evidence to know what sentences she holds true. I hear an utterance (or string of meaningless sounds), observe her behaviour and assign to her the attitude of holding true towards that sentence. Then I still do not know what the sentence means, but I know which sentence she holds true. To know what the sentence means, I, once again, observe her behaviour and environment. If I, for example, see Jane, a Xhosa speaker, run from what appears to me to be a big dog that is barking while chasing her, and I hear her yell "lumkela inja" or "ndicla uncedo" I can assign meaning to her utterances by employing my own perceptual experience to offer a hypothesis as to what in our environment is prompting her to hold them true. So, I will imagine that she yelled "lumkela inja" or "ndicla uncedo" because of the aggressive-looking dog. And, knowing this, I assign that truth condition to that specific utterance. Jane's utterance "lumkela inja" is true if and only if, from my own perspective, there is an aggressive dog. Meaning just is the truth condition of a sentence. And, as seen above, the only way in which the truth condition of Jane's sentence can be specified, is by making use of myself, as her third-person, who understands her utterance as prompted by something in our shared environment. Thus, at this point, as the interpreter, I will know what sentence Jane holds true, and will know what that sentence means. And, according to Davidson, a belief is no more than a sentence that a speaker holds true. So, once I know what sentence Jane holds true, and I know what it means, then I know what she believes.

Thus, interpretation proceeds as follows: I, firstly, have to assume that almost all Jane's utterances indicate the attitude of holding true towards the sentence she is uttering, secondly, observe the objects in our environment and offer hypotheses as to what her sentence means by referring to the objects that prompt her to speak the

University of Cape Town

sentence that she holds true and, finally, derive her belief from my knowledge of the meaning of the sentence that she holds true.[31]

2.2. CHARITY: BEYOND A ROUGH SKETCH

As mentioned above, Davidson proposed two constraints that need to be adhered to when interpreting someone else. He derived these constraints, not from anything theoretical, but rather from the way in which a radical interpreter will have to go about making sense of someone with whom she shares no language. Both constraints are aspects of what Davidson calls the principle of charity.

2.2.1. CORRESPONDENCE

The first constraint, which I already employed above, is one without which interpretation cannot get off the ground at all. In my interpretation of Jane above, I assumed that she was responding to the same aggressive-looking dog that I perceived. And this is what allowed me to judge that her utterance meant "aggressive dog" and her belief was about an aggressive dog. I, in other words, assumed that her perceptual experiences were similar to mine and that, therefore, she held beliefs that were true from my perspective, true. Without making these assumptions, I would not have been able to assign meaning to Jane's utterances at all and, consequently, I would not have been able to attribute a belief. This is Davidson's principle of correspondence, and it encourages the interpreter to take the speaker to be responding to the same features of the world that he (the interpreter) would be responding to under similar circumstances. It, hence, allows the interpreter to charitably credit the speaker with a degree of what the interpreter takes to be true belief about the world.[32]

2.2.2. COHERENCE

An assumption of correspondence, however, is a crude method of fixing someone's attitudes, unless we add Davidson's second constraint. When interpreting someone, I have to assume that she is rational, on my understanding of rationality, and attribute to her a collection of logically consistent beliefs.[33] I should not simply ascribe those beliefs that I judge to correspond with mine under one or two specific conditions.

University of Cape Town

Instead, I should discover as much as possible about the speaker before I attribute attitudes that are all consistent with one another.[34] If, for example, I am aware of the fact that Jane enjoys running, loves dogs and believes that the pursuing dog belongs to her, I cannot attribute a belief that the pursuing dog is dangerous, just because I am afraid of dogs and would have held that belief while being chased by it. I will have to attribute to her attitudes that are all consistent with each other and that are based on my entire body of Jane-related knowledge. For instance, a belief that the dog is harmless, a belief that she is playing a game, a desire to entertain both herself and the dog, and so forth. This is Davidson's principle of coherence, and it encourages the interpreter to detect a degree of logical consistency in the attitudes of the speaker. It, in other words, allows the interpreter to charitably credit the speaker with a degree of what the interpreter takes to be rationality.

2.2.3. THE PLACE OF CHARITY

The usefulness of the two aspects of the principle of charity is, thus, that I can use my own attitudes, utterances and perceptions as a guide to the attitudes, utterances and perceptions of others, unless there is a good reason not to do so. This is why theorists often say that, on Davidson's system of radical interpretation, my role, as interpreter, is to map a speaker's sentences onto my own by employing my own perceptions of our environment, together with my response to that environment, as a guide to what she is saying and, thereby, believing.

We should always obey these two aspects of charity when assigning meaning and attributing beliefs. This still does not guarantee an accurate interpretation, however, because it is possible for an interpreter to apply correspondence and coherence to an insufficient amount of evidence about the speaker. A genuinely accurate interpretive project is one in which an entire set of consistent attitudes are attributed based on everything one can know about her. In the situation above, if I possessed all the information that there was to be had about Jane, my interpretation and attributions would have fixed her entire range of meanings and attitudes accurately, because then I would have been what has often been called a fully informed (or omniscient) interpreter. This is Davidson's reason for claiming firstly, that correct attitude

University of Cape Town

determination requires such an omniscient interpreter and, secondly, that what such an omniscient interpreter cannot discover is simply not mental.

Critics may want to argue that we can never possess sufficient knowledge of any person to allow our interpretations to comply fully with, especially the coherence constraint, and that, as a result, our attributions can never be appropriate to fix a person's beliefs accurately. Still, the fact that such an omniscient interpreter does not actually exist, does not defeat Davidson's primary claim that beliefs, generally speaking, are states that are attributed to us by others during a process of interpretation that satisfies certain important constraints. If I interpret someone without being able to obey those constraints fully, such as when I lack sufficient information about her, I am still fixing her beliefs, but whether I am correct can only be established by making use of a complete method of interpretation that does include all information and, as a result, obey the constraints completely. Just so long as there is no talk of a belief that can be discovered independent of interpretation, or of something other than interpretation fixing beliefs, Davidson's theory remains in tact.

Now that all the main features of Davidson's theory of interpretationism have been presented, it should be clear enough that Davidson himself never intended his system to describe what we usually do when we communicate. (More about that in chapter 5.) For now it is important to appreciate that Davidson was trying to create a conceptual picture of the mental. He never suggested that situations of radical interpretation were common, or that individuals like omniscient radical interpreters existed (or even ever could exist). His aim was to use some of our intuitions, such as the procedures we would have to follow when trying to make sense of someone with whom we shared no language, to paint a picture of what would have to be in place if we wanted the whole of someone's mental life explained. Accordingly, in the remainder of this project, when I refer to "a situation of radical interpretation", I am talking, not about a situation that actually occurs, but about a conceptual process whereby meaning and beliefs have to be constituted with the necessary constraints in place. Similarly, when I refer to an omniscient interpreter, I am not talking about an actual person that can or does exist, but about an imaginary character that illustrates what would have to be known for a person's mental life to be explained completely and accurately.

University of Cape Town

3. THE INCOMPATIBILITY OF RADICAL INTERPRETATION AND THE ASYMMETRY THESIS

The following is just a loose and first take on exploring the incompatibility between radical interpretation and the asymmetry thesis and will be developed in more detail as the project proceeds.

3.1. RADICAL INTERPRETATION AND THE EVIDENCE ASYMMETRY

An objection that many philosophers have raised to Davidson's radical interpretationism is related to the first asymmetry identified in Chapter 2. If the very identity of belief is provided by the role it plays in third-person attributions and third-person attributions are almost never made without behavioural evidence, it implies that most beliefs are attributed based on behavioural evidence. And if belief is mostly attributed based on behaviour then, whether they belong to us or to others, they are so attributed. But, as argued in chapter 2, I do not usually treat myself, and others do not usually treat me, as if I learn of my own beliefs by observing or interpreting my own behaviour. It does not appear, however, that Davidson's radical interpretationist model can accommodate this view for, if he admits that we can learn of our own beliefs in a way that is independent of our behaviour, he is contradicting one of the most fundamental interpretationist tenets, namely, that there is no way of identifying beliefs outside a project of interpretation.

University of Cape Town

3.2. RADICAL INTERPRETATION AND THE TRANSPARENCY ASYMMETRY

The system of radical interpretation also seems to encounter problems when trying to account for the second asymmetry, which holds that my self-attributions of beliefs are usually consistent with my view of the world, whereas my attributions of beliefs to others are made in the service of an explanation of their behaviour. If the very identity of belief is provided by the role it plays in third-person attributions and third-person attributions are made in order to rationalise behaviour, it implies that beliefs can only be attributed on the basis of rationalising behaviour. And if belief can only be attributed in the service of an explanation of behaviour then, whether they belong to us or to others, they must be so attributed. The problem for radical interpretationism

stems from Davidson's claim that, for something to be a belief, an interpreter needs to attribute its content during a process of interpretation. From the third-person perspective, which is then how all beliefs are attributed, the acquisition and revision of beliefs essentially involve explanatory power whereas the way any one person views the world is of secondary importance. But, as argued in Chapter 2, from a first-person perspective one's view of the world plays a fundamental role in the acquisition and revision of beliefs while their explanatory power is purely incidental.

3.3. RADICAL INTERPRETATIONISM AND THE AUTHORITY ASYMMETRY

The asymmetry related to the authority that we assign first-person belief-claims poses a problem for Davidson's radical interpretationism because, if a third-person fixes one's beliefs by attributing their content, it seems to imply that such a third-person is actually in a more authoritative position regarding the identity of such beliefs. Once Davidson removes the first-person's privilege to establish the identity of her own beliefs by leaving it up to the third-person's attributions, he seems to deny the first-person's claim to be generally correct about what she believes. The one who fixes the identity of a belief is more likely to be correct about it than the one that is not in a position to do so.

University of Cape Town

4. DAVIDSON'S RECONCILIATION OF RADICAL INTERPRETATIONISM AND THE ASYMMETRY THESIS

4.1. DAVIDSON'S ASYMMETRIES

Davidson recognised the evidence and authority asymmetries discussed in chapter 2 and never denied the transparency asymmetry. In his writings on self-knowledge, he focused primarily on addressing the authority asymmetry, and his explanation of the evidence asymmetry is directly involved in this. I will, hence, set out his solution to the authority asymmetry and offer an explanation of the evidence and transparency asymmetries that is consistent with his theory of self-knowledge. I, however, need to address two important qualifications before moving on to his explanation of the asymmetries.

Firstly, Davidson defends a version of the evidence asymmetry that differs from the one endorsed in this project in two closely connected ways. Take the evidence asymmetry In Davidson's words, "Because we usually know what we believe (and desire and doubt and intend) without needing or using evidence (even when it is available), our sincere avowals concerning our present states of mind are not subject to the failings of conclusions based on evidence."[35] In this quote, it is clear that he actually claims that the first-person knows what she believes without making use of evidence of any kind, while I wish to remain neutral on the nature of such first-personal evidence, if any, and merely claim that we do not treat the first-person as if she knows what she believes by employing the same sort of evidence employed by a third-person. I, in fact, think that environmental evidence plays a large role in knowing what we believe, precisely because I view beliefs as states that are attributed via their transparency to how a person sees the world, as opposed to already established states that are waiting to be discovered. My disagreement with Davidson regarding the existence of first-person evidence, consequently, seems to be a result of another difference between our evidence asymmetries, namely, the Cartesian language that Davidson's evidence asymmetry is couched in. As I argued in Chapter 2, our intuitive notion of the evidence asymmetry, which is exactly what Davidson adopts, is probably mistaken and needs to be replaced by the transparency asymmetry (that is a much more plausible way of thinking about the traditional notion of immediacy). It is, in fact, surprising that Davidson holds onto the traditional version of the evidence asymmetry, considering his commitment to the idea that beliefs, qua contentful states, are not already established internal states about to be discovered by us.

Still, even if the evidence asymmetry is re-phrased with this in mind as, say, the role that behavioural evidence plays in our attribution (as opposed to our knowledge) of first-person beliefs, Davidson's evidence asymmetry nonetheless differs from mine since I only claim that behavioural evidence plays a much smaller role in first-person belief-attributions, while he eliminates all evidence from the first-person case.

Secondly, Davidson's version of the authority asymmetry also differs from mine. He argues that I still speak with authority about a belief-claim that someone else has legitimately overturned. In his words, "even when a self-attribution is in doubt, or a

University of Cape Town

challenge is proper, the person with the attitude speaks about it with special weight."[36] This claim seems to be stronger than my idea that we are treated as if we are usually right about what we believe, which implies that, once someone else has proved me to be wrong about what I believe, I am not treated as if I am right about that specific belief anymore.

Notwithstanding these differences in detail, his asymmetries are indisputably versions of my evidence and authority asymmetries. So, it is worth examining his reasons for thinking that they can be included in his system of radical interpretation. Considering that his versions of the asymmetries are "stronger" than mine, if his incorporation of his asymmetries into a system of radical interpretation works, then it is likely to work as an incorporation of mine.

4.2. DAVIDSON'S RECONCILIATION STRATEGY

In his paper, First-Person Authority, Davidson explains that the first two asymmetries can be accounted for once we understand that, within an interpretationist system, a speaker generally knows, while an interpreter may not know, what a speaker's words mean.[37] Recall that, for a belief to be attributed, an interpreter has to know both which sentence a speaker holds true and what she means by that sentence. Consequently, even when both the interpreter and the speaker know what sentence the speaker holds true, only the speaker knows what she believes because only she knows the meaning of her words. The interpreter can, of course, assign meaning to her words via interpretation, but the fact that he has to interpret her words, while she does not have to interpret her own words, makes it possible for him, but not for her, to be generally or largely mistaken about what her words mean. After all, the meaning of the words of a speaker involves numerous elements like … "her actions and other words, her education, birthplace, wit and profession, her relation to objects near and far, and so forth".[38] There, as a result, can be no general guarantee that an interpreter is interpreting a speaker correctly whereas the general guarantee that the speaker is getting the meaning of her words right is that whatever she consistently applies her words to is what gives them the meaning they have.[39] The asymmetries between first- and third-person perspectives of beliefs, therefore, rest on the presumption that speakers usually know the meaning of their words without interpretation (or the use of

University of Cape Town

behavioural evidence), while it is possible that their interpreters, even after interpretation (or evidence), are still mistaken about what the speaker's words mean.

There are, thus, two aspects to the reconciliation. Firstly, the speaker knows, while her interpreter may not know, what she means, and, secondly, the speaker knows what she means without using her own behaviour as evidence, while her interpreter cannot know what she means without using such evidence.

What is critically important in this project, is that Davidson admits that there may be an asymmetry between the speaker's knowledge of which sentence she holds true and her interpreter's knowledge of which sentence she holds true, but he does not think that this can help explain the attitude asymmetries. To say that a speaker authoritatively and immediately knows which sentence she holds true is no less mysterious than to claim that the speaker knows authoritatively and immediately what she believes. There does not seem to be any way other than the identification of some essentially first-person accessible thing to ground the speaker's knowledge of what she believes and of which sentence she holds true. But, since we already know that a belief and a sentence that is held true have to be socially constituted in order to have content at all, a speaker's authoritative and immediate knowledge of what she believes and of which sentence she holds true cannot proceed via the identification of something that is not accessible to her observer. Meaning, as will be shown below, is the only step in the interpretive process that can ground whichever asymmetries exist, so a meaning asymmetry is what we should focus on.[40]

University of Cape Town

Before proceeding to the details of Davidson's argument for this reconciliation, we can become familiar with the general form of his account of self-knowledge by spelling out how he accommodates the asymmetries. I will start with the authority asymmetry, because that is essentially the only one that he explicitly accounts for, and is the one that he thinks is more basic and interesting than the others are.

4.2.1. THE AUTHORITY ASYMMETRY

1. If both the interpreter and the speaker know which sentence the speaker holds true, then

2. There is a presumption that a speaker generally knows the meaning of her sentence because the consistent use of her sentences fixes their meaning, and
3. There is no presumption that an interpreter generally knows the meaning of her sentence because he has to conduct the complex task of interpretation in order to assign meaning to her utterances.
4. A belief is attributed by giving a meaningful sentence that the speaker holds true.
5. There is a presumption that, if both the interpreter and the speaker know which sentence the speaker holds true, the speaker generally knows what she believes while there is no such presumption in the case of her interpreter.

4.2.2. THE EVIDENCE ASYMMETRY

1. If both the interpreter and the speaker know which sentence the speaker holds true, then
2. There is a presumption that the speaker does not generally rely on behavioural evidence when assigning meaning to her own utterance because the consistent use of her sentences alone fixes their meaning, and
3. There is a presumption that an interpreter relies on behavioural evidence when assigning meaning to her utterance.
4. A belief is attributed by giving a meaningful sentence that the speaker holds true.
5. There is a presumption that, if both the interpreter and the speaker know which sentence the speaker holds true, then, generally, the only behavioural evidence employed in attributing beliefs to a speaker is employed by the interpreter in the process of assigning meaning to her sentence, and, thus,
6. There is a presumption that a third-person has to rely on behavioural evidence to acquire knowledge of the beliefs of another while a first-person's knowledge of her own beliefs does not usually involve such evidence.

4.2.3. THE TRANSPARENCY ASYMMETRY

1. If both the interpreter and the speaker know which sentence the speaker holds true, then

University of Cape Town

2. There is a presumption that the speaker generally knows the meaning of her sentence because the consistent use of her sentences fixes its meaning, and
3. There is no presumption that an interpreter generally knows the meaning of her sentence because he has to conduct the complex task of interpretation.
4. There is a presumption that, if both the interpreter and the interpretee know which sentence the interpretee holds true, then, generally, only the interpretee really knows what she holds true because she generally knows, while her interpreter may not, the meaning of the sentence that she holds true.
5. A belief is attributed by giving a meaningful sentence that the speaker holds true.
6. There is a presumption that self-ascriptions of beliefs are generally consistent with what a person holds true about the world while third-person attributions are made in the service of an explanation.

4.3. DAVIDSON'S THREEFOLD ARGUMENT FOR HIS RECONCILIATION STRATEGY

Davidson offered an argument for his reconciliation strategy that consists of three parts. I will call the first component his transcendental argument from interpretability, the second his argument from disquotation and the third his argument from the speaker's status as an interpreter. In chapter 4, I intend to argue, against the common misconception that Davidson did not in fact offer three separate arguments for his reconciliation strategy.

4.3.1. THE TRANSCENDENTAL ARGUMENT FROM INTERPRETABILITY

In the first stage of his argument for his reconciliation strategy, Davidson asks us to imagine a situation in which two people attempt to communicate, even though they do not share a language and, thus, can make no sense of each other without engaging in a project of radical interpretation. If the speaker wishes to communicate with the interpreter, she will have to try to be interpretable. In other words, as Davidson puts it, "the best she can do is to use a finite supply of distinguishable sounds applied consistently to objects and situations she believes are apparent to her hearer".[41] The interpreter has only the speaker's utterances and non-verbal behaviour as evidence to interpret as well as, of course, his own perceptions of the environment that they share.

University of Cape Town

Davidson argues that, in this situation, it makes no sense to think that the speaker is generally getting her language wrong. This is because the consistent use of her sentences gives them their meaning. If she uses her sentences inconsistently, she will not be interpretable. In other words, if she fails to employ the same sentence in the same environmental conditions every time (or at least most of the time), her interpreter will not be able to make sense of what she is saying. On the other hand, if she uses her sentences consistently, she will be interpretable and, because the consistent use of her sentences will render her interpretable, we can say that she is getting her language right. No sense can be made of the notion of someone who generally gets her language wrong for, if she was getting her language wrong, then she wouldn't be interpretable as applying her words consistently and would then, by definition, not be speaking a language at all.

Davidson has often warned against a conception of the consistency of the use of sentences as simply a collection of noises that an interpreter can find meaningful. We should understand the consistency of use as stemming from an intention of the speaker to use her sentences consistently, not from whatever the interpreter can find meaningful.[42] Therefore, if a speaker is interpretable, she is necessarily getting her language right. And if she is getting her language right, we can assume that she is applying her sentences consistently with the intention of providing clues about their meaning to her interpreter.[43] And if she uses her sentences consistently with the intention of providing clues about their meaning, then one can say that she knows the meaning of her sentences.

Davidson then generalises this argument to all language speakers to show that his reconciliation strategy, and thereby the asymmetries between first- and third-person perspectives of beliefs, apply to all language speakers. All language speakers are interpretable because, for something to qualify as a language, one needs to apply the same sentences consistently in the same environmental conditions. Therefore, all language speakers are mostly getting their language right. And the fact that all language speakers mostly get their language right implies that they are necessarily applying their sentences consistently with the intention of providing clues about their meaning to their interpreters. And the fact that they use their sentences consistently with the intention of providing clues about their meaning implies that they know the

University of Cape Town

meaning of their sentences. And the fact that all speakers, unlike their interpreters, generally know the meaning of their own sentences implies that all speakers, unlike their interpreters, generally know what they believe (the authority asymmetry) without making use of behavioural evidence (the evidence asymmetry) and ensures that their beliefs are consistent with what they hold true (the transparency asymmetry).

4.3.2. THE ARGUMENT FROM DISQUOTATION

The point of the transcendental argument from interpretability is to show that, if one is interpretable, then one is using one's language consistently with the intention of providing clues to one's interpreters about the meaning of one's sentences. Disquotation is a way in which Davidson proves that, if someone is interpretable and applying her sentences consistently, she can actually state their meaning. This is Davidson's account of what we would want to call first-person perspective.

He explains this stage of the argument for his reconciliation as follows: "The speaker, after bending whatever knowledge and craft he can to the task of saying what his words mean, cannot improve on the following sort of statement: 'My utterance of "Wagner died happy' is true if and only if Wagner died happy'. An interpreter has no reason to assume this will be his best way of stating the truth conditions of the speaker's utterance."[44]The sentence on the right-hand side of the biconditional is understood to be the truth conditions of the sentence on the left. The argument is, accordingly, that a speaker can always correctly state the truth conditions of her sentences by using the same phrase on the right hand side of the biconditional that she uses on the left, because, assuming that she is applying her words consistently, the environmental conditions specified by the phrase on the right-hand-side of the biconditional will be identical to the environmental conditions specified by the phrase on the left. An interpreter, of course, can state the same biconditional. The only problem is, however, that he will be stating it in the speaker's idiolect whose meanings he does not yet know. So, when stating the biconditional in the same way as the speaker, he has no guarantee that disquotation in the speaker's idiolect will yield an accurate account of the truth conditions of the utterance.

University of Cape Town

If, on the other hand, the interpreter tries to state the biconditional in his own idiolect, which will involve interpreting the meaning of the sentences on both sides of the biconditional, he still does not know whether the biconditional in his own idiolect correctly states the truth conditions of the sentence because he does not know whether he correctly interpreted the meaning of the speaker's sentence. He will have to start by stating the phrase on the left-hand-side of the biconditional (which is the speaker's utterance that he is trying to interpret), then employ all the speaker's behavioural evidence in addition to his own perception of the environment to interpret the phrase on the right, and only then will he be able to say that, in his own idiolect, the meaning of the phrase on the right is identical with that on the left, since, according to him, the environmental conditions specified by the phrase on the right are the same as those specified by the speaker's original phrase on the left.[45] Therefore, a speaker is always in a position to state the truth conditions of her sentences correctly while there is no guarantee that her interpreter can.

Disquotation, however, can only work if the speaker is applying her sentences consistently under the same environmental conditions. If she fails to do this, it is conceivable that, when she states a biconditional like the one above, the phrase on the right-hand-side of the biconditional does not specify the same environmental conditions as the phrase on the left. The transcendental argument from interpretability is supposed to ensure that, if the speaker is interpretable, she is applying her sentences consistently and, thereby, that she is getting her language right. And the argument for disquotational knowledge is supposed to provide a way in which the first-person, if applying her sentences consistently, can state their truth conditions.

We should not understand the transcendental argument from interpretability independently of the argument for disquotational knowledge because interpretability alone does not explicitly allow us a way to state the truth conditions (and thereby the meaning) of our own utterances. And we should not interpret the argument for disquotational knowledge independently of the transcendental argument from interpretability because disquotation alone does not guarantee the consistent application of our sentences, which is required to get the truth conditions of our sentences right.

University of Cape Town

4.3.3. THE ARGUMENT FROM THE SPEAKER'S STATUS AS AN INTERPRETER

This third argument for the reconciliation is not one that Davidson himself ever offered in support of it, probably because it only dawned on him in the latter stages of his career that his system of radical interpretation implied it.[46] It does, however, strengthen the case for a reconciliation between radical interpretation and the attitude asymmetries considerably and it confirms that my interpretation of Davidson's reconciliation strategy is more accurate than the standard interpretation offered by other philosophers.

If the speaker is interpretable, it implies that she is speaking a language. If she is speaking a language, then she is capable of being an interpreter of others, since she has a language onto which she can map their idiolects. In order to assign meaning to their utterances, she needs to observe which environmental conditions they are uttering the sentences that they hold true in the presence of, figure out which sentence she would have held true under the same environmental conditions and, thus, via her own perception of their environment and her knowledge of her own language, translate their sentences into her language to interpret their meaning. The only way in which she will be able to know what they mean by the sentence that they hold true, in other words, is by knowing which sentence she would have held true in the same environmental conditions. And, if she knows which sentences she holds true when encountering the environmental conditions in which she holds them true, then she clearly knows the meaning of her sentences. Speakers, thus, know what they mean since their status as interpreters of others requires their knowledge of their own language, while their interpreters still have to work out what they mean. (Much more about this in chapter 6).

The argument from the speaker's status as an interpreter cannot work independently of the transcendental argument from interpretability and the argument from disquotation. A speaker can be an interpreter of others only if she speaks a language, and we can only say that she speaks a language if she is interpretable. It also requires the argument from disquotation because, to be an interpreter of another, our speaker needs to be able to state the truth conditions of her own sentences. If she cannot state

University of Cape Town

the truth conditions of her own sentences accurately (which will be the case if she is not interpretable), then she cannot state the truth conditions of someone else's sentences accurately since getting right the truth conditions of someone else's sentences is parasitic upon getting right the truth conditions of one's own.

5. CONCLUSION

Once again, the reader should bear in mind that Davidson was trying to create a conceptual picture of the mental, as opposed to proposing that all the above is what we in fact do in practice. Situations of radical interpretation are not common, and individuals like omniscient radical interpreters do not (and probably cannot) exist. The aim of this chapter was to give a sympathetic exegesis of Davidson's system of radical interpretation, which describes some methods and constraints that he proposes have to be in place for beliefs to be attributed accurately. Beliefs have to be constituted socially (section 1). They have to be attributed by an interpreter that is observing and making sense of a speaker's speech and behaviour by attempting to charitably map her language and behaviour onto his own language and his own behaviour (section 2). Radical interpretation demonstrates that first-person speakers know what they mean and that they do not have to use their own behaviour as evidence to know what they mean. This shows why we treat first-person speakers as if they know what they believe, as if they do not have to use their own behaviour as evidence to know what they believe and as if they self-attribute beliefs in the service of an explanation of their behaviour (section 4).

In the remainder of this project I will oppose the popular view that Davidson's reconciliation strategy cannot give a satisfactory account of the three attitude asymmetries. Theorists are typically drawn to this conclusion by their misunderstanding of what Davidson was saying. I want to suggest that the standard interpretation of Davidson's reconciliation (called the meaning asymmetry) cannot explain the asymmetries, but that an improved interpretation of this reconciliation (called the sentence held-true asymmetry) can.

University of Cape Town

Before proceeding to chapter 4, it is important to outline my approach to the problem so that readers can understand the road I want them to travel in the remainder of this project.

The first claim (1) is that interpretation in the radical case necessarily involves a general meaning asymmetry between a speaker and her interpreter since interpretation implies that the former, but not the latter, generally knows the meaning of her words. (Note the word generally.) We, firstly, assume that a first-person usually knows what she means because the consistent use of her words, as interpreted by her interpreter, fix their meaning. We, secondly, assume that a first-person does not usually use her own behaviour to know what she means since the consistent use of her language (as opposed to her speech and behaviour) gives her her knowledge of what she means. According to Davidson, radical interpretation necessarily leads to the general form of the meaning asymmetry since, if a speaker is interpretable, she is necessarily applying her words consistently and can thereby qualify as speaking a language. Language use, hence, brings about the general form of the meaning asymmetry, since the consistent use of our words and the consequent success of our disquotational statements explain the first-personal aspects of meaning.

The second claim (2) is that the general form of the meaning asymmetry leads to the general form of the attitude asymmetries. If radical interpretation can show why the speaker usually knows what she means, and why she usually knows this without using her own behaviour as evidence, then it can show why we are treated as if we usually know what we believe, why we are treated as if we usually know this without using our own behaviour as evidence and why we are treated as if we usually self-attribute beliefs that are consistent with what we hold true. This step from meaning to belief is acceptable, firstly, because a belief just is a sentence that is held true and, secondly, because the content of our beliefs is the same as the content of our language. If the fact of speaking a language cannot explain our asymmetrical treatment of first- and third-person meaning, then it clearly cannot explain our asymmetrical treatment of first- and third-person beliefs, and if it can account for the asymmetrical treatment of meaning, then it can also account for our asymmetrical treatment of belief.

University of Cape Town

The third claim (3) is that, even though the general form of the attitude asymmetries are always present, there are situations in which we realise that the attitude asymmetries are absent and a speaker is wrong about what she believes, has to use her own behaviour as evidence to know what she believes or has to self-attribute a belief in the service of an explanation of her behaviour.

The fourth claim (4) is that, during communication in our linguistic communities with their shared conventions, claims 1, 2 and 3 have to persist.

Davidson himself does not lead us far through the challenge of the asymmetry thesis. He explicitly argues that defending claim 1, and arguing that claim 2 follows from claim 1, is what he needs to do to defend his system of radical interpretation from the challenge of the asymmetry thesis. I, on the other hand, propose that he also needs to be able to explain how claim 3 follows from claims 1 and 2, and how claim 4 follows from claims 1, 2 and 3. Davidson's tendency to overlook the last two claims made him formulate his reconciliation strategy very roughly to give a clear account of only claims 1 and 2. This rough formulation has since become the standard interpretation of Davidson's reconciliation strategy and is taken directly from how he formulated it. It is what I term the meaning asymmetry and what I intend to argue gives an unsatisfactory account of the attitude asymmetries.

Accordingly, the remainder of this project will proceed as follows: In chapter 3, I explained why claim 2 follows from claim 1. In chapter 4 section 2, I intend to argue that the meaning asymmetry interpretation can explain claim 1 and concede, for the reasons given in chapter 3, that claim 2 follows from claim 1. In chapter 4 section 3, I will argue that the meaning asymmetry interpretation cannot show how claim 3 follows from claims 1 and 2. In chapter 5, I will argue that the meaning asymmetry interpretation cannot show how claim 4 follows from claims 1, 2, and 3. In chapter 6, I will give an alternative interpretation of Davidson's reconciliation strategy, termed the sentence held-true asymmetry, that is based on claims 1 and 2, but that can, in addition, account for claims 3 and 4. Since my sentence held-true asymmetry builds on the successes of the meaning asymmetry interpretation, my agreements with it in chapter 4 should accordingly be read as defences of my sentence held-true asymmetry interpretation.

University of Cape Town

CHAPTER 4
THE MEANING ASYMMETRY AND A SITUATION OF RADICAL INTERPRETATION

In this chapter I intend to consider the standard interpretation of Davidson's reconciliation of radical interpretation and the asymmetry thesis. It is directly based upon the description in the previous chapter and is termed the meaning asymmetry. I, firstly, intend to argue that the meaning asymmetry can show why the attitude asymmetries, when understood as general assumptions about beliefs, are always present in a situation of radical interpretation. I, secondly, will argue that the meaning asymmetry leads to peculiar consequences in situations where the attitude asymmetries are absent. My first argument is, thus, that the meaning asymmetry is more plausible than is often supposed, while my second argument is that, notwithstanding a degree of plausibility, there is a feature of the attitude asymmetries that it cannot explain. It should be noted that this chapter is concerned with the meaning asymmetry's ability to explain the attitude asymmetries only in a situation of radical interpretation, and not in everyday interactions in linguistic communities. Since a situation of radical interpretation does not arise frequently, and since a fully informed radical interpreter is an imaginary character that illustrates a point about the mental, rather than a person that can ever exist, the question in chapter 4 should be understood as a question about the ability of the meaning asymmetry to provide a conceptual account of the attitude asymmetries.

University of Cape Town

1. THE MEANING ASYMMETRY INTERPRETATION

The meaning asymmetry interpretation is taken directly from Davidson's explanation of his reconciliation strategy, with little awareness of its origin or implications. A speaker can only be said to speak meaningfully if she is interpretable to an omniscient radical interpreter. Since the consistent use of her sentences is what renders her interpretable, the fact that she is interpretable means that the consistent use of her sentences fixes their meaning. If the consistent use of her sentences fixes their meaning, and she is applying them consistently, then she is getting the meaning of her sentences right. A speaker, thus, knows, without evidence, what she means. Her

interpreter has to conduct the complex task of the interpretation of her speech in the context of their environment and her history to know what she means. Accordingly, the speaker knows, while her interpreter may not know, what she means, and the speaker knows without evidence, while her interpreter cannot know without evidence, what she means. (The arguments were set out in detail above and will not be repeated here.)

This is the form that the meaning asymmetry takes as an explanation of the attitude asymmetries.

The Authority Asymmetry as explained by the meaning asymmetry

1. Both the interpreter and the speaker generally know which sentence the speaker holds true.
2. There is a presumption that the speaker generally knows the meaning of her sentences because their consistent use fixes their meaning.
3. There is no presumption that an interpreter generally knows the meaning of the sentences of the speaker because he has to conduct the complex task of interpretation in order to assign meaning to her utterances.
4. A belief is attributed by giving a meaningful sentence that the speaker holds true.
5. There is a presumption that the speaker generally knows what she believes while there is no such presumption in the case of her interpreter.

University of Cape Town

The Evidence Asymmetry as explained by the meaning asymmetry

1. Both the interpreter and the speaker generally know which sentence the speaker holds true.
2. There is a presumption that the speaker does not generally rely on behavioural and verbal evidence when assigning meaning to her own utterances because the consistent use of her sentences alone fixes their meaning.
3. There is a presumption that an interpreter relies on behavioural evidence when assigning meaning to her utterances.

4. A belief is attributed by giving a meaningful sentence that the speaker holds true.
5. There is a presumption that, generally, the only behavioural evidence employed in attributing beliefs to a speaker is employed by the interpreter in the process of assigning meaning to her sentences.
6. There is a presumption that a third-person has to rely on behavioural evidence to acquire knowledge of the beliefs of another while a first-person's knowledge of her own beliefs does not usually involve such evidence.

The Transparency Asymmetry as explained by the meaning asymmetry

1. Both the interpreter and the speaker generally know which sentence the speaker holds true.
2. There is a presumption that the speaker generally knows the meaning of her sentences because their consistent use fixes their meaning.
3. There is no presumption that an interpreter generally knows the meaning of the words of the speaker because he has to conduct the complex task of interpretation.
4. There is a presumption that, generally, only the speaker really knows what she holds true because she generally knows, while her interpreter may not, the meaning of the sentence that she holds true.
5. A belief is attributed by giving a meaningful sentence that the speaker holds true.
6. There is a presumption that self-ascriptions of beliefs are generally consistent with what a person holds true about the world while third-person attributions are made in the service of an explanation.

The meaning asymmetry interpretation, whose advocates I will from now on label traditional interpretationists, maintains that Davidson attempts to explain the attitude asymmetries via a meaning asymmetry only, and that he denies, either the existence of a sentence held-true asymmetry, or the role of a sentence held-true asymmetry in an explanation of the attitude asymmetries. (Note how the first statements above differ from those in chapter 3.) Davidson does not explicitly use a sentence held-true asymmetry, but he leaves open the possibility of its existence and even its explanatory

University of Cape Town

role. (More about this in chapter 6). Traditional interpretationism takes Davidson as denying that a sentence held-true asymmetry has any explanatory role, which is why it starts from the assumption that the speaker and her interpreter generally know which sentence she holds true. Davidson, on the other hand, treated the meaning asymmetry as primary and argued that, whichever other asymmetries exist, and whatever explanatory role they have, need to be derived from the meaning asymmetry.

The meaning asymmetry interpretation is present in the writings of numerous philosophers. Bernhard Thele made it clear that he thought that Davidson should have explained the attitude asymmetries via both a meaning and a sentence held-true asymmetry, but that he explained them with reference to a meaning asymmetry only, without leaving room for a sentence held-true asymmetry. In his own words (with my terminology and emphasis):
"In order to give a complete explanation of the attitude asymmetries, it seems, we have to explain the sentence held-true asymmetry as well as the meaning asymmetry. But Davidson argues that we do not need the sentence held-true asymmetry to explain the attitude asymmetries."[47]

Kirk Ludwig also interprets Davidson as arguing for a pure meaning asymmetry as an explanation for the attitude asymmetries. In his own words:
"Although a speaker might know with better warrant what her words meant, she might not know with as great a warrant as her interlocutor what sentences she held true. And, without an asymmetry between one's own knowledge of which sentences one holds true and that of an interpreter, we are not guaranteed an asymmetry between one's knowledge of one's own and of others' attitudes."[48]

Or David Beisecker: "Since Davidson is reluctant to trace any asymmetry of knowledge between speakers and interpreters to a sentence held-true asymmetry, presumably the speaker knows which sentence she holds true in the same way as her interpreters would.[49]

These philosophers, together with some others, then go on to argue that the concept of the meaning asymmetry is heavily flawed and, if Davidson cannot explain our asymmetrical treatment of meaning, then he cannot explain our asymmetrical

University of Cape Town

treatment of belief. In the remainder of this chapter, I want to discuss the ability of the meaning asymmetry to explain the speaker's immediate and authoritative knowledge of what she means in a situation of radical interpretation.[50] If it can explain her immediate and authoritative knowledge of what she means in a situation of radical interpretation, then the connection between meaning and belief guarantees that it can explain the three attitude asymmetries. I will offer two arguments. In section 2, I want to defend the meaning asymmetry from some common misunderstandings and show that it is, in fact, able to explain our asymmetrical treatment of meaning (and thereby of belief). In section 3, I want to begin my argument against the meaning asymmetry interpretation by showing that the meaning asymmetry alone cannot explain situations in which the attitude asymmetries are judged to be absent. (See the schematic outline at the end of chapter 3).

2. THE MEANING ASYMMETRY AS AN EXPLANATION OF THE GENERAL FORM OF THE ATTITUDE ASYMMETRIES

2.1. OBJECTIONS TO THE TRANSCENDENTAL ARGUMENT FROM INTERPRETABILITY

The first challenge is based on many Davidsonian claims and can be spelled out as follows: Davidson takes the fact that utterances are interpretable to mean that they are meaningful, the fact that they are meaningful to imply that the speaker means something by those utterances and the fact that the speaker means something by her utterances to imply her actually knowing what she means by them. Bernhard Thele once accused Davidson of sliding from:

(a) A speaker's utterances are (in general) interpretable to
(b) A speaker's utterances are (in general) meaningful to
(c) A speaker (in general) means something by his utterances to
(d) A speaker knows (in general) what he means by his utterances.[51]

In a similar vein, Barry Smith objects that Davidson's transcendental argument from interpretability substitutes "meaning what I say" for "knowing what I mean".[52] Thele and Smith essentially have the same problem with this deflationary strategy towards knowing the meaning of one's utterances.[53]

University of Cape Town

It is easy to see why "meaning something by my utterances" should normally coincide with "knowing what I mean by those utterances", and that "meaning something by my utterances" usually depends on those utterances being meaningful. But Davidson does not think that "my utterances are usually meaningful" involves my "meaning something by my utterances" or that "meaning something by my utterances" usually involves "knowing what those utterances mean". He proposes that "my utterances are meaningful" implies that "I mean something by my utterances" and/or that "I know what I mean by my utterances". But remember that, according to Davidson, for my words to be meaningful (and thereby to mean something by my words) is no more than for others to interpret me as meaning it more often than not. Whether I can truly be said to use my sentences consistently depends upon whether I am interpretable; that is, upon whether others can attribute consistency of use to me. So, whether my sentences are meaningful, and whether I mean something by my sentences seems to be up to my interpreters. And if knowing the meaning of my sentences is the same as my sentences being meaningful and, thereby, as meaning something by my sentences, then Davidson is in effect arguing that, if others understand my speech to be meaningful (or interpretable), then I know the meaning of my sentences. And this is the point to which many Philosophers object since they think that it misses some aspect that is central to the first-person perspective; something akin to first-person involvement.

Barry Smith, for example, once remarked that, "we are missing something crucial to the first-person perspective of the language-user if we do not recognize a sense of comprehension beyond that of the speaker producing words the interpreter can find meaningful".[54]

But what exactly does Smith's objection mean? It intuitively endorses the idea that we enjoy some first-person comprehension of our language while we speak. But this is not an intuition that an interpretationist will deny. Smith then goes on to claim that interpretationism cannot accommodate this intuition, without attempting to discover whether interpretationism can, in fact, accommodate it. In what follows, I intend to show that interpretationism can accommodate a sense of first-person comprehension, but that such first-person comprehension is tied to third-person interpretation.

University of Cape Town

If critics try to reject interpretationism for its inability to make room for a sense of first-person comprehension of meaning independent of third-person interpretation, then they are unquestionably assuming the very aspect of mind that interpretationism denies, which dooms any further discussion to failure.

Bernhard Thele has a worry that is similar to Smith's, and that he manages to express without employing objectionably first-personal terms. Thele thinks that being interpretable does not imply (or cannot be substituted for) knowing what we mean. In his own words:

> Davidson goes so far as to identify 'knowing what she means' with getting her language right. This identification is certainly unacceptable: we certainly would not say that a speaker knows the meaning of two of her utterances if she were agnostic about whether or not they mean the same. But I see no reason why a speaker who is agnostic about sameness of meaning should be uninterpretable.[55]

Thele's problem is that it is possible for a speaker to be both interpretable and lacking knowledge of what she means, since it is possible for a speaker to be both interpretable and agnostic about whether two of her sentences mean the same.

In a case where a speaker is uninterpretable, a meaning asymmetry interpretation seems right. If she, for example, is agnostic about whether the meaning of the sentences "I deny this" and "I refute this" are identical or different, she will probably refrain from using them at all, especially because she knows that she is meant to make herself interpretable. Still, if she does use the sentences with deny and refute interchangeably, while admitting that she has no opinion on whether they mean the same, she will remain generally interpretable, but with regard to these two sentences, she will fail to apply them consistently in the same environmental conditions. She will apply both to describe a situation where she declares something untrue or refuses to believe that something is true. She will also apply both to describe a situation where she proves a theory to be false or where she overthrows a theory by argument or proof. As a result, I will not be able to map her language onto my own, or I will not be able to understand which environmental conditions her sentences are being applied in. In my own experience there are two events for which my language has two different sentences, while it seems to me that she understands the same two events, but that her

University of Cape Town

language has two sentences that are randomly applied to both. This will render her use of those two sentences uninterpretable to me even though I find the remainder of her language generally interpretable. It is beyond doubt that the speaker does not know what she means by those sentences, and that is why we can say that her use of them is uninterpretable, even though she is otherwise generally interpretable.

Thele did not provide us with an example of a situation in which a person is both interpretable and agnostic about sameness of meaning which weakens his objection considerably. If a speaker, for example, employs the sentences "This is a good result" and "This is a good outcome" to refer to the same type of event, and she does so consistently, then her use of these two words will obviously be interpretable. In my experience there is one type of event in a specific type of context for which my language has two different sentences, and it seems to me that her experience and her language resemble my own in this way. A strong case can be made, however, that, if she really does not know whether the two words mean the same or not, then something in her speech will reveal this fact.

Say she asks someone whether there is a difference between the sentences containing outcome and result, or whether there are different conditions that they can be applied in, or whether they are always applied in the same conditions, or whether there are other contexts in which they cannot be used interchangeably, etc. That would give her agnosticism about meaning away because, instead of applying them consistently, she will be asking questions about their application to ascertain what would make her use of these words interpretable to others. If Davidson's claim is accepted that almost all our assertions are of sentences held true, her questions will be interpreted as follows: she holds the sentence "I am unsure about whether these sentences containing outcome and result mean the same" true. Her disquotational statements will become questions that look something like this: "This is a good outcome if and only if what?" and "This is a good result if and only if what?" This is a situation in which the speaker is agnostic about meaning, and understood to be agnostic about meaning. Her interpreters will find her past use of those two sentences interpretable, because of her consistent application, but past interpretability alone does not imply true interpretability. Her interpreters will know that her consistent application does not imply that she is getting her language right, exactly because they understand her to be

University of Cape Town

unsure about whether they mean the same, and may at any time become uninterpretable. The evidence on which we judge whether someone is interpretable or not includes not only past assertions, but also future ones.[56]

My critics may object that it is conceivable that the speaker may be agnostic about the meaning of two of her sentences without ever asking another person whether they mean the same or not. But an interpretationist like Davidson does not claim that the agnosticism has to be revealed in the person's actual speech. All he claims is that the agnosticism about meaning has to be interpretable. The evidence from which an interpreter interprets includes not only the speaker's actual statements, but also her dispositional ones. If the speaker is ever asked whether she thinks that those sentences containing result and outcome mean the same, then she will admit that she does not know. Her speech will reveal her agnosticism and her application of the two sentences will be understood in that light. Davidson's claim thus stands since the agnosticism will be interpretable to an omniscient interpreter under favourable circumstances.

My critics may object that it is unlikely that someone will ever ask the speaker whether she knows whether those sentences containing outcome and result mean the same if she applies them consistently in the same conditions. But this objection commits the same mistake as the one above. Whether an interpreter, in fact, will or will not ask such a question is not important. All that Davidson needs to defend his claim is that, in principle, it will be possible for an interpreter to ask such a question and that, in principle, the speaker's agnosticism about meaning will be so revealed. If the interpreter never asks, then he is not interpreting her under favourable circumstances since he is not prompting her in a way that will reveal the potentially accessible information. If he does ask, then the circumstances are ideal for interpretation since he is prompting her in a way that will bring such information to light. (More about this in 2.2 below).

One of the mistakes philosophers make is to assume that the only information that is relevant to being interpretable is a speaker's actual application of her language. But, as seen above, the speaker's application of her sentences has to be interpreted in the light of everything else we know about her, which will include her potentially manifestable agnosticism. So, her interpreters, because they will have all this

University of Cape Town

information in mind, will know that she is not applying her sentences consistently precisely because she is asking questions about how to apply them, and thereby they will know that her use of them is not interpretable and that she is, as a result, not getting her language right. A convincing example of a situation in which a person is agnostic about the meaning of two phrases and still remains fully interpretable with respect to her use of them has not, to my knowledge, been produced.[57]

I presume that the worry behind both Thele's and Smith's objections is that it is possible for someone to be interpretable but yet not be able to explain what she means by one of her phrases.[58] But they aim this objection specifically at the transcendental argument from interpretability. And, as explained in chapter 3, the transcendental argument from interpretability should not be understood as an argument that operates independently of the argument from disquotation. Once the argument from disquotation is added, this objection no longer holds because then the speaker does have a way of stating what her phrases mean.

2.2. OBJECTIONS TO THE ARGUMENT FROM DISQUOTATION

This objection holds that disquotation cannot give us the knowledge of the truth conditions of our sentences. Those who advance this objection argue that there is a difference between knowing that a biconditional is stating the truth conditions of a sentence and knowing what truth conditions the biconditional states.[59] Disquotation can show why a speaker knows that she is stating the truth conditions of a sentence, but knowing that truth conditions are being stated is not the same as knowing those truth conditions. And surely merely knowing that one is stating the truth conditions of a sentence cannot be sufficient for actually knowing those truth conditions. If Davidson and I assume that the knowledge that truth conditions are being stated is sufficient for actually knowing what those truth conditions are, then our theory involves a peculiar notion of knowledge. A type of knowledge of truth conditions that may not be expressible or communicable to others, a type of knowledge of meaning where the individual in question may not be able to tell the difference between the meaning of two of her sentences in different terms from those contained in the sentence, knowledge of meaning where the individual may not be able to tell when two of her sentences mean the same, knowledge that she might not necessarily be able

University of Cape Town

to relate to whatever else she believes or to conditions in her environment. And if the knowledge that one is stating truth conditions cannot be sufficient for actually knowing those truth conditions, then the speaker cannot qualify as knowing the truth conditions of her sentences. And if she cannot be said to know the truth conditions of her sentences, then it cannot be claimed that she knows their meanings.

Here is why critics may think that stating truth conditions is not sufficient for knowing what those truth conditions are. When we assign meaning to the utterances of another speaker, we use behavioural and environmental evidence to judge what that speaker applies her words to. And then we state a biconditional, just in our own idiolect. "Jane's utterance that snow is white is true if and only if snow is white." And this qualifies as knowledge of what she means, precisely because we have identified the environmental conditions that give the truth conditions of her utterance and, as a result, the phrase on the right-hand-side of the biconditional that we use to explain what she means picks out the environmental conditions that give the truth conditions of the phrase on the left. This cannot apply to the first-person case where she states the biconditional that holds the truth conditions of her utterance as well, since such cases do not involve an interpretation of what she is uttering her sentence in the presence of. The third-person case, thus, involves a judgment of what the speaker is applying her words to, and a judgment of how her application relates to the interpreter's own language. Critics may argue that the first-person case involves no more than the mere application of phrases, the stating of the same phrase on both sides of a biconditional and the good fortune of appearing to others to be applying her phrases consistently. It might be this difference between the biconditionals that allows an interpreter to succeed in having knowledge of what the speaker means while the speaker's own biconditional can accomplish no more than conveying to the speaker that, if an interpreter manages to find her interpretable, then truth conditions are being stated.

The first mistake that this objection makes is to interpret Davidson as saying that disquotation is meant to give the speaker the knowledge of the truth conditions contained in the disquotational biconditional or, as claimed above, that stating the truth conditions is sufficient for knowing what those truth conditions are. This is, however, not the case since it neglects to take into account the role of the

University of Cape Town

transcendental argument from interpretability. She knows the truth conditions of her sentences, as evidenced by the fact that she is applying her language consistently. Given that she is applying her language consistently and that she accordingly knows what she means, disquotation gives her an accurate way of stating the truth conditions of her sentences. My critics, in other words, should not state the objection in terms of what knowledge the disquotational biconditional can give the speaker, but rather as what knowledge it allows the speaker to state. And if the objection holds that disquotation can allow the speaker to state truth conditions without the speaker's knowing which truth conditions she is stating, then there is a persuasive line of reasoning to show such critics to be mistaken.

My response appeals to interpretability to show that the speaker must know which truth conditions her disquotational biconditionals state. If the speaker is judged to apply a phrase consistently across many different contexts, she is judged to respond consistently to questions about the environmental events that her biconditional involves and she is judged to respond consistently to questions about the truth conditions as stated in her biconditional, it is difficult to see what can justify denying that her behaviour exhibits knowledge of the truth conditions contained in the biconditional. Like in the case of interpretability in 2.1 above, if she knows only that her biconditional is stating truth conditions, without knowing what those truth conditions are, her behaviour or speech will, if prompted appropriately, reveal her uncertainty about what such truth conditions are.

University of Cape Town

Moreover, interpretation proceeds holistically or via a judgment of coherence. As explained in chapter 3, meaning is holistic in nature since sentences in a language depend for their meaning on other sentences in that language. Her behaviour will, as above, reveal her uncertainty of what truth conditions her disquotational biconditional states And this point holds not only when the speaker is asked about that one specific biconditional, but also in other contexts across the whole of her language. If she lacks knowledge of the truth conditions stated in, for example, the biconditional "snow is white if and only if snow is white", such ignorance will be revealed when she talks about (or when she is specifically asked about), for example, sentences like "Snow is cold", "Milk is white", "Grass is not white", etc. We, thus, do not have to worry about the possibility that disquotation allows us to state the truth conditions of our sentences

without our knowing what those truth conditions are. If this had been the case, then we would have gotten large parts of our language wrong, evidenced by frequent inconsistent applications of our words and admissions of our ignorance of the truth conditions of our sentences.

The mistake that Philosophers tend to make is to assume that the only evidence for knowledge of meaning through disquotation is a statement of a biconditional like the one above.[60] But what sort of evidence for knowledge can that really be, considering that it is, as critics enjoy pointing out, merely a statement of the same phrase on both sides of a biconditional? The evidence for that knowledge (or lack of knowledge) is present throughout the whole of the speaker's speech.

In order to dispute this response, as well as the one in Section 2.1, a critic will have to construct an example of a case in which the following holds: First, the speaker lacks knowledge of something. Second, the speaker consistently behaves as if she has this knowledge. Third, even if interpreted in favourable circumstances, (by, say, prompting an honest, intelligent speaker who understands what is asked, who wants to provide the answers and who is not prevented from doing otherwise), her ignorance remains hidden. In other words, an example that involves a speaker that is willing to admit her ignorance, but who cannot do so since such ignorance is the kind of thing that can remain inexpressible in language and behaviour. Such an example cannot be produced.

University of Cape Town

Firstly, it may require a type of necessarily first-person only accessible space unreachable even to those with the right kind of information obtained in the right kind of circumstances. More importantly, it will have to deal with all the implications of such a view regarding the nature of such a space. This is exactly the type of picture that everyone has condemned Cartesian immaterialism for for centuries, without realising that their own way of thinking about the mind either conforms to it or entails the same consequences as it does.

At this point my critics may accuse me of assuming that the objection implies that knowledge of what the truth conditions are is necessarily inaccessible to observers; hence, my anti-Cartesian words of warning above. My opponents might claim that the

example above is consistent with materialist theories of mind (such as J.J.C. Smart's type identity theory, D.M. Armstrong's central state materialism, Hilary Putnam's machine-state functionalism or Jerry Fodor's psychofunctionalism) that do not imply that such knowledge is inaccessible to observers like Cartesian immaterialism does. Since they are materialist theories, they inevitably allow third-person access to mental states, but not necessarily the type of access that interpretationists demand. Now my opponents might claim that my response above is persuasive only if such materialist theories are disregarded. It is possible to construct an example of a situation with the three features I specified above if one bears in mind that the speaker's lack of knowledge of what truth conditions her biconditional states can be accessible without being evident from her behaviour, past, present and future, actual and dispositional.

Interpretationism restricts the information on the basis of which beliefs can be attributed to what any ordinary layperson has access to, namely, a speaker's speech and behaviour together with information about her environment. An examination of the physical structure or processes of a brain, which is the kind of information that the materialist theories listed above allow, goes way beyond what any layperson interpreter can find out and is, according to an interpretationist, irrelevant to the mental. Interpretationism holds that, if we discover that the attributions that an omniscient interpreter under favourable circumstances make are reflected in our physical brain processes, then we can still use a speaker's speech, behaviour and environment to attribute beliefs. It further maintains that, if we discover that there are brain activities that we cannot link with the beliefs attributed by an omniscient interpreter in favourable circumstances, then such activities do not give us any information about anything mental. No questions about the mental can be answered at any level other than a layperson's interpretations based upon information about a speaker's speech, behaviour and environment. If this, my critics may claim, allows for a situation where everything that is mental is potentially manifestable in speech and behaviour, and a picture where everything mental is potentially manifestable in speech and behaviour allows for a situation where disquotation (together with a third-person judgment of interpretability) can give a speaker knowledge of what she means, then one of the materialist approaches above is preferable.

University of Cape Town

Such strategies cannot give a plausible account of the speaker's ignorance with regard to the truth conditions that her biconditional states, however. What is required is a picture that can make sense of a situation where a speaker consistently gives the right truth conditions of a phrase by applying the relevant parts of her language consistently, where she is asked specifically whether she knows the truth conditions of a phrase, where she is willing to admit that she does not but where she is unable to since everything she says or is disposed to say indicates that she does know the truth conditions, etc. In such cases, the materialist theories above will have to hold that a scientist who is studying physical or functional brain processes will be able to detect that, even though everything we do know, and can know, about her indicates that she does know what those truth conditions are, she actually does not. This seems like something that few contemporary philosophers will endorse. It completely removes something that is meant to be mental from the realm of rationality. The "scientifically accessible only" ignorance fails to cohere with anything that the speaker does or can possibly say or do. It, further, fails to cohere with anything else that she knows or believes and even conflicts with some of what she knows and believes. (Knows and believes as judged both by her and by others, of course).

The most desperate of my critics can try to argue that it does cohere with some of her other knowledge and beliefs, and that the knowledge and beliefs that I think it conflicts with is not really knowledge and beliefs that she has, since they are things that her interpreters are mistakenly attributing to her. But such critics will overlook the enormity of the mistakes that her interpreter will then have to make. Say that such critics argue that she is ignorant of the truth conditions of her sentence "Snow is white", even though she is applying that specific sentence, together with sentences about snow and white consistently across all contexts. To deny that she knows the truth conditions of "Snow is white", they will have to argue that the scientist (who is detecting the physical or functional state of the ignorance of those truth conditions) is right, while her interpreters are wrong, about what she knows. But if her interpreters are wrong when attributing to her the knowledge of those truth conditions, then their attributions to her of the knowledge of the truth conditions of many other sentences will fall into doubt. One does not even have to subscribe to the holistic nature of meaning to see this. Straightforward compositionality is sufficient to show that, if the speaker knows the meaning of "snow", "is" and "white", and she can apply it

University of Cape Town

consistently across all contexts, then she knows the meaning of the sentence "snow is white". There simply is no justification to withhold knowledge of the truth conditions from her if everything she does and can say indicates that she knows them.

The idea that a speaker lacks knowledge of what truth condition her biconditional states can be plausible only if such ignorance is accessible either to herself or to her interpreters. And if it is accessible, then the interpretationist claim stands that what is not discoverable by an omniscient interpreter under favourable circumstances just is not mental. Then there is nothing that prevents us from accepting the claim that a disquotational biconditional (together with a third-person judgment of consistent word application), allows the speaker to accurately state the truth conditions that she knows her sentences to have.

Most of the arguments in this chapter are based on the claim that interpretability does not imply that the speaker genuinely knows what she means. It only implies that, since we need to make her interpretable, and since knowledge of what she means will make her interpretable to us, we have to attribute the knowledge of what she means to her. Davidson brought about this objection by misleadingly asserting: "There is a presumption - an unavoidable presumption built into the nature of interpretation - that the speaker usually knows what he means."[61] This encouraged philosophers to treat the speaker's knowledge of what she means as an instrumental tool that we use to make her interpretable, as opposed to as something that she genuinely has. In Sarah Sawyer's words:

> According to Davidson, if we do not assume a subject knows her thoughts, then we cannot begin the process of radical interpretation. This explains why the presumption is needed, but it is not clear that this in itself provides a justification for it.[62]

This objection, however, conceives of interpretationism as instrumentalism whereby we use whatever is useful to make sense of what others do. The metaphysical commitment of interpretationism is that, whatever an omniscient radical interpreter attributes in order to render a speaker interpretable, picks out what is already there, just from the perspective of that interpreter's interpretive scheme, just like the lengths of thirty centimetres or twelve inches are properties that a ruler already has, described by either the measurement scheme of centimetres or inches. (More about this in

University of Cape Town

section 2.3). The response to this objection is, thus, that the speaker knows what she means, and that this is why she applies her language consistently and is interpretable to her interpreters. Her interpreters, thus, attribute the knowledge of her language to her, not only because it is useful to render her interpretable, but because that is what is supported by the largest amount of evidence. We, in other words, treat others as if they know what they mean because radical interpretation demands that we pick out and attribute to them whatever all available evidence proves they have. (I will return to this point in chapter 7 to assess whether interpretationism can, accordingly, accommodate a stronger version of the asymmetry thesis than that described in chapter 2).

2.3. OBJECTIONS TO THE MEANING ASYMMETRY VIA DUAL INTERPRETIVE SCHEMES

University of Cape Town

According to this objection, radical interpretation cannot involve the meaning asymmetry, since such a meaning asymmetry will have to be compatible with dual interpretive schemes. It cannot be made to be compatible with dual interpretive schemes without peculiar consequences. Thus, the meaning asymmetry that radical interpretation implies is implausible and cannot explain the attitude asymmetries.

Davidson argues that, where there are two omniscient interpreters that are both employing a method of interpretation that satisfy the necessary constraints, there is no independent check on which interpreter's understanding is correct. His system of radical interpretation in fact entails this. As argued in the previous chapter, what an omniscient interpreter who employs an appropriate interpretive project cannot find out about a speaker simply is not mental at all. There is nothing other than interpretation that can settle the question of what a person means and believes. This opens up the possibility that two such fully informed appropriate systems of interpretation can yield different results, without the option of settling the matter in some other way. If interpreter I interprets a speaker as meaning I*, and interpreter N interprets a speaker as meaning N*, and they both possess all he information that there is to be had about her, and they both employ a system of interpretation that obeys the constraints of correspondence and coherence, then there is no further way of judging whether she means I* or N*; That is, there is no way of adjudicating between the two systems to

ascertain which one is giving the more accurate interpretation. And, since interpretationism holds that meaning is constituted by these interpretations, there is then essentially no one thing that the speaker actually means. In other words, if there is no more to meaning than what an omniscient interpreter who is employing an appropriate system of interpretation can attribute, then there is no more to what a speaker means than what two such interpreters will attribute. If they happen to attribute different meanings, then there is no further matter of fact as to what the speaker genuinely means.

One common objection to Davidson's meaning asymmetry as an explanation for the attitude asymmetries is, accordingly, that if two such omniscient interpreters understand a speaker as meaning different things by her words, then it will have to imply that she means both those things by her words and that she, as a result of Davidson's deflationary strategy to meaning, knows that she means both those things by her words.[63] But the idea that she can mean two different things by her words and, therefore, know that she means two different things by her words once again requires a fairly peculiar concept of knowing: a kind of knowledge that she will not be able to express or that she will not be able to relate to whatever else she knows. She again will not be able to tell the difference and similarities between the meanings of two of her sentences because the two different things that she supposedly means by them (and knows what she means by them) are attributed to her by two third-persons.

Attempting to solve the problem by using "intending to mean something by her words" instead of "meaning something by her words" cannot work either. In the original argument, Davidson phrases consistency of use in terms of an intention to use our words consistently.[64] But, an intention is just another propositional attitude, so it is going to have to be attributed to her by an interpreter to qualify as an intention. And then the problem is the same as in the paragraph above. The speaker's intention to use her words consistently will be attributed from the third-person stance. So, intending to mean something by her words, and thereby knowing that she means something by her words, will be attributed by a third-person via a project of interpretation. Then we will still have to accept the possibility that two omniscient interpreters can understand us as intending to mean different things by our words, which implies that we do know that we intend to mean those different things by our words. Thus, if knowledge of

University of Cape Town

what we mean has to be attributed from the third-person stance, our concept of knowledge will have to change considerably.

Davidson, however, also argues that interpretations are relative to specific interpretive schemes. Even if two omniscient interpreters interpret me as meaning two different things by my words, I do not mean, and know that I mean, two different things by my words. There are simply two equally accurate systems of interpretation that can make sense of what I say. A correct understanding of the way in which the objects of speech determine what we mean (and consequently the identity of our beliefs) can illustrate why indeterminacy does not imply that I know that I mean two different things by my words.

I will borrow Davidson's measurement analogy to show how this objection can be overcome.[65] The situation of indeterminacy explained above is found in the measurement of length, weight and distance just as much as in the attribution of meaning and belief content. One measurement system employs centimetres to assign numbers to a ruler that indicate its length, another system employs inches to assign different numbers to that same ruler to indicate its length. One omniscient interpreter (I) that obeys the right interpretive constraints assigns meaning and belief content to a speaker that make sense of what she says and does, another such interpreter (N) may assign different meanings and belief contents to her that make sense of her. Even though they assign different numbers, we cannot say that the system that uses centimetres is more right than the one that uses inches, since they make equally good sense of the length of the ruler. Similarly, even though they assign different meaning and belief contents, we cannot say that interpreter I is more right than interpreter N, since their systems make equally good sense of the speaker. Let us assume that a speaker is applying her words consistently. Interpreter I says that she means I* and interpreter N says that she means N*. Her language can thus be mapped onto two other people's languages (or her language can be measured by two measurement models).

Returning to the objection above, what exactly does she mean and know that she means? The objection assumes that she means, and that she knows that she means, everything that is attributed to her. This, however, once again neglects the role of

University of Cape Town

disquotation. She means, and knows what she means, whatever truth conditions her disquotational biconditional states. If it is slightly different from the truth conditions that her interpreter's translation of her biconditional states, it is simply because his language (or measurement model) is a slightly different model that is capable of making sense of what she says. And if there are two omniscient interpreters who are obeying the necessary constraints of interpretation that attribute different meanings to her, then she still knows that she means whatever truth conditions her biconditional states. And what truth conditions exactly her biconditional states is a question for which there cannot be only one answer, precisely because it is a matter that can be measured by different measurement models that are employed from different perspectives.

She knows what truth conditions her biconditional states because she applies her words consistently (even across different contexts). As seen above, if she is asked what she knows she means, and she is willing to say what she knows she means, and she does say what she knows she means, and there is nothing that she can say or do that suggests that she does not know what she means, then she does know what she means. Interpreter I and N know what truth condition her biconditional states because they have found measurement models that make complete sense of her by finding coherent relationships between all her statements (past and future, actual and dispositional). There may be more than one measurement system that can make equally accurate sense of her. But this does not mean that she means different things by her words or that she knows that she means different things by her words in the same way as we do not usually say that a ruler is both thirty centimetres and twelve inches. If it is measured in centimetres, then it gets assigned the number thirty. If it is measured in inches, then it gets assigned the number twelve. The numbers thirty and twelve are given relative to the systems that assign them. If a speaker is interpreted by omniscient interpreter I, she gets assigned a collection of truth conditions; if she is interpreted by omniscient interpreter N, she may be assigned a different collection of truth conditions. The truth conditions, and thereby meanings and belief contents, are assigned relative to specific interpretive schemes. So there is no room to say that she means, and knows that she means, two different things by her words. Since there are two omniscient interpreters that are finding her word application to be consistent, we can say that she is speaking a language. And the fact that she is speaking a language

University of Cape Town

allows her disquotational biconditionals to be accurate statements of what she means, and of what she knows she means. This constitutes her interpretive scheme for herself, so what she means, and what she knows she means, is given relative to her own interpretive scheme of herself.

Objecting to the measurement analogy by suggesting that different measurement models can be directly translated into each other and that there is consequently some kind of matter of fact as to the ruler's length, as Philosophers tend to do, is of no help in this context. Proponents of this view will say that, analogously, there must be a matter of fact as to what the speaker means (and what she knows she means), evidenced by the fact that she can say what she means and what she knows she means. Theorists may then object that a theory that allows for interpretive schemes that cannot be translated into each other cannot capture this matter of fact of what the speaker means.

Even if it is true that different measurement models can be directly translated into each other in a way that interpretive schemes cannot, it does not help the objection above. It is true that different models of interpretation (that is, interpreters' languages and systems of rationality), cannot be directly translated into each other since a translator will have to use another such system to relate the systems he is trying to compare, which simply leads to questions about the third system. But this observation does not change the key point that all measurement models are equally accurate. We don't have to compare centimetres and inches to verify that they make equally good sense of measuring the length of a ruler. All we have to do is to investigate each system to establish whether it can attribute coherent relations between different measurements. For example, we want four centimetres to be twice the length of two centimetres and we want one-hundred inches to be twenty times the length of five inches. We don't need inches to verify that centimetres can attribute such coherent relations and vice versa. The same holds for attributions of meaning and belief content. We don't need to be able to translate systems of interpretation into each other to be able to verify that a system makes accurate sense of what a speaker does. If it can make sense of everything the speaker does, and it can attribute coherent relations between what the speaker does, and it carries the additional benefit that many of the speaker's observations of her environment seem to coincide with what the interpreter

University of Cape Town

observes, then it is an accurate system of measurement for that speaker. Whether centimetres and inches can be directly compared, and whether this implies that there is a matter of fact as to the length of the ruler are matters to be settled elsewhere. The point here is that, since the interpretive systems cannot be directly compared, and since there is nothing over and above interpretation that can settle the question of what a speaker means, there cannot be a matter of fact as to what she means. Her language and its disquotational biconditionals, which serve as an interpretive system of herself, is one such system that gives her knowledge of what she means and believes. Her interpreters have other equally accurate systems that help them make sense of her. Whether disquotation does, or does not, assign the same truth conditions (and thereby meaning and belief contents) to her as her interpreters do, is not important, firstly, since we have already established that she knows that she means only one thing and, secondly, since the impossibility of translating systems into each other without employing further such systems will make it a question without a solution anyway. And why would we want such an answer if the speaker knows what she means, and if her interpreters can make an accurate judgment of what she means. Whatever the speaker is doing is right, and whatever her interpreters are doing is right. They can communicate successfully. There is simply no problem to be solved, no further questions to be asked.

University of Cape Town

3. THE MEANING ASYMMETRY AND THE ABSENCE OF THE ATTITUDE ASYMMETRIES

The meaning asymmetry can, accordingly, explain the presence of the attitude asymmetries as assumptions about how people usually relate to their own beliefs as well as to the beliefs of others, but a lot more is required of an acceptable theory of self-knowledge. Explaining our general assumptions about beliefs is important, but we also need an explanation of specific cases in which we treat a speaker as if she is wrong about what she believes, as if she is using her own behaviour and speech as evidence to know what she believes and as if she is self-attributing a belief in the service of an explanation of her speech and behaviour. And this is where the meaning asymmetry alone fails.

I will use the authority asymmetry to illustrate, since all three of the attitude asymmetries pose the same problem. Let us return to an example in chapter 2. Suppose that a speaker says that she believes that her husband is not a murderer, while someone with whom she communicates maintains that she actually believes that her husband is a murderer since a lot of her behaviour indicates that she fears him.

The meaning asymmetry interpretation will have to explain it as follows:

1. Both the speaker and her interpreter know that she holds true the sentence "My husband is not a murderer" since that is what she is saying.
2. Since her usual meaning of such a sentence is incompatible with the majority of evidence in this case, she seems to be applying her words inconsistently, which removes her position of first-person authority.
3. Both the speaker and her interpreter have to use everything they know about her as evidence to know what she means by the sentence "My husband is not a murderer".
4. Since the majority of evidence indicates it, they both know that, in this specific situation, by the sentence "My husband is not a murderer", she means what she usually means by the sentence "My husband is a murderer".[66]
5. A belief is attributed by giving a meaningful sentence that the speaker holds true.
6. They both know that she believes that her husband is a murderer.

University of Cape Town

While discussing her belief with her interpreter, she will use both the sentences "My husband is not a murderer" and "My husband is a murderer". She uses the former when saying what she thinks she believes and the latter when talking about what she thinks her interpreter is attributing to her, which she argues is exactly the opposite of what she does believe. According to the meaning asymmetry interpretation, by her sentence "My husband is not a murderer", in this specific situation, she actually means that her husband is a murderer. But in this description of the discussion between the speaker and her interpreter she is using the sentence "My husband is a murderer" to talk about her husband being a murderer, which is exactly what she argues she does not believe.

In this situation she understands herself as meaning two different things by the two sentences since she is applying her words inconsistently and is consequently wrong about what she means. But she, in fact, does not mean two different things by the two sentences as confirmed by the majority of evidence. From her interpreter's point of view, she means the same thing by the two sentences, even though she fails to realise it. He will understand that she understands herself as meaning two different things by the two sentences, and he will be able to work out what those two things are. Hence, from his perspective it is a discussion because he knows that she thinks that she means two different things by her two sentences, and from her perspective, it is a discussion because she thinks that she means two different things by her two sentences.

The problem with this explanation is that the meaning asymmetry interpretation then suggests that the discussion the speaker and her interpreter will be having is a discussion about what she means by her sentence "My husband is not a murderer", while it will actually be a discussion about what she holds true. The interpreter will not try to convince the speaker that she actually means the opposite of what she usually means by that sentence. He will try to convince her that she is wrong about what she holds true. Further, when she corrects herself, she will not admit that she has the meaning of her sentences wrong. She will say that she is wrong about what she holds true. It will be a discussion about what in the environment she holds true, and not about what she means.

Take Davidson's original argument for his system of radical interpretation. In a situation where two people share no language, the best a speaker can do is to apply her words consistently to things in her environment that she thinks are apparent to both herself and her interpreter.[67] Davidson derives the two constraints of the principle of charity from what our intuitions inform us such an interpreter will have to do to succeed in making sense of her. If we continue to describe this process, we will eventually reach a stage where the interpreter and the speaker are both familiar with her language; where they both know her typical responses to specific environmental situations. Once we reach this stage, there is nothing in interpretationism that rules out the possibility of judging that a speaker has misidentified environmental conditions. The interpreter will know that, even in this one situation, by the sentence "My

University of Cape Town

husband is not a murderer" the speaker means that her husband is not a murderer. He will also know that, by the sentence "My husband is a murderer" she means that her husband is a murderer. Her language will be held constant. But he will judge that she has misidentified which state of affairs she holds true.

This can be best illustrated by the type of discussion they will have. An interpreter is not typically going to say, "in this situation you mean 'your husband is a murderer' by the sentence 'My husband is not a murderer' and I will convince you that that is what you mean". He is most probably going to say, "Since you are having a discussion with me, and since I understand your point in the discussion and since you understand mine, you clearly understand what you mean by both these sentences. But there is behaviour that indicates that you have misidentified what you hold true". He is not going to say, "There is behaviour that indicates that you are applying your words inconsistently and you therefore think that you mean one thing but you actually mean the other". Our intuitions, in the form of a discussion between the speaker and her interpreter, indicate that a discussion about what someone believes will be a discussion about what environmental conditions the speaker takes to be true, as opposed to about what the speaker means by what she says.

Philosophers who believe that the meaning asymmetry alone can explain the attitude asymmetries (my traditional interpretationists), may respond that our intuitive understanding of the situation is right. It is true that, in a case where we do not know what we believe, it is because we do not know what we hold true. And in a case where we are wrong about what we believe, it is because we are wrong about what we hold true. Recall that the meaning of a sentence is its truth conditions; that is, the environmental conditions under which an interpreter judges the speaker to hold it true. If a speaker uses the sentence "My husband is not a murderer" in the presence of her husband being a murderer, she is getting the meaning of her sentence wrong precisely because she has misidentified the environmental conditions under which she is currently applying it. Getting the truth conditions of a sentence wrong is the same as getting its meaning wrong. Both the speaker and her interpreter know which environmental conditions she usually utters it in, but he realises, while she does not, that the current environmental conditions are not the ones that usually prompt her to utter that specific sentence.

University of Cape Town

This response is similar to the strategy that I defend in chapter 6. The speaker is wrong, while her interpreter is right, about the environmental conditions in which she is uttering her sentence. They both attribute to her the same sentence that she holds true. "My husband is not a murderer". But he is right, while she is wrong, about what she means by it since he is right, while she is wrong, that the current environmental conditions are not the same conditions that she usually holds it true in. The moment her interpreter convinces her of the nature of their shared environmental conditions, and that she usually holds a different sentence true in them, then she will change the sentence that she holds true.

The problem with this response in this context is that it appeals to the connection between meaning and truth to obtain a sentence held-true asymmetry which it then utilises to explain situations in which the attitude asymmetries are absent. And even though I agree that this is how Davidson should be interpreted, it is not consistent with the meaning asymmetry interpretation according to which Davidson allegedly argues that, firstly, a sentence held-true asymmetry either does not exist or plays no explanatory role in the attitude asymmetries and, secondly, the meaning asymmetry, in the absence of all other asymmetries, can explain the attitude asymmetries.

The meaning asymmetry alone, as a result, cannot explain our intuitive conceptions (that are backed up by the requirements and constraints of a situation of radical interpretation), of situations in which the three attitude asymmetries are absent.

We want an explanation of the authority asymmetry that can respect the following: When the speaker is unsure about what she believes, she consults the states of affairs in the world that she is trying to form a belief about. She can use two sentences, one to say what she thinks she believes and one to talk about the opposite of what she thinks she believes. She knows what both sentences mean. She can use them in the current situation and in other contexts, she has always used both sentences consistently as interpreted by others and she can relate both to the rest of her language. She can thus accurately state the truth conditions of both through disquotation. She just does not know which of the two sentences she holds true.

University of Cape Town

We want an explanation of the evidence asymmetry that can respect the following: When the speaker is unsure about, or does not know, what she believes, she knows the meaning of both of the sentences that she suspects she might hold true. She can use them in the current situation and in other contexts, she has always used both sentences consistently as interpreted by others and she can relate both to the rest of her language. She can thus accurately state the truth conditions of both through disquotation. She just does not know which sentence she holds true since there is evidence that suggests both. She now has to try to figure out which case is supported by the largest amount of evidence. But then she is not using her own behaviour and speech to figure out what she means, she is using it to know which sentence she holds true.

We want an explanation of the transparency asymmetry that can respect the following: When the speaker self-attributes a belief in the service of an explanation of her behaviour and speech, it is because she does not know what she holds true. She knows the meaning of both of the sentences that she suspects she might hold true. She can use them in the current situation and in other contexts, she has always used both sentences consistently as interpreted by others and she can relate both to the rest of her language. She can thus accurately state the truth conditions of both through disquotation. She just does not know which sentence she holds true since there is evidence that suggests both. She now has to try to figure out which case is supported by the largest amount of evidence and accordingly self-attribute a belief in the service of an explanation of her speech and behaviour.

University of Cape Town

4. CONCLUSION

In section 1, I described the meaning asymmetry interpretation of Davidson's reconciliation between radical interpretation and the asymmetry thesis. In section 2, I argued that radical interpretation really does involve a meaning asymmetry between a speaker and her radical interpreters and, if we accept the connection between meaning and belief (which both Davidson and I do), then the meaning asymmetry can explain the general form of the attitude asymmetries. In section 3, I argued that the meaning asymmetry alone fails to provide a plausible account of situations in which the attitude asymmetries are absent, since such situations have a lot to do with which sentences we hold true. In chapter 6, I will show how a sentence held-true asymmetry

relies on, and benefits from, the strengths of the meaning asymmetry pointed out in section 2 and how it can resolve the difficulties described in section 3. Before proceeding to chapter 6, however, I want to use chapter 5 to show that the problems of the meaning asymmetry interpretation are more substantial than this chapter suggests.

University of Cape Town

CHAPTER 5
THE MEANING ASYMMETRY AND LINGUISTIC CONVENTIONS

In this chapter, I will make it clear why a meaning asymmetry alone cannot give a satisfactory account of the attitude asymmetries. I intend to do this by showing that one cannot extend the same sort of meaning asymmetry that is present in the radical case to everyday cases of language use, and that the version of the meaning asymmetry that is present in practice is not appropriate to explain the attitude asymmetries between speakers and the interlocutors with whom they share linguistic conventions. Section 1 will describe the Davidsonian claims that give rise to the suspicion that there is no meaning asymmetry between a speaker and her interlocutors. Section 2 will examine four possible strategies for accommodating the meaning asymmetry in practice, together with my reasons for thinking that they all fail to explain the attitude asymmetries.[68]

1. BACKGROUND

The problem that this chapter deals with stems from a perceived tension between four Davidsonian claims. Firstly, the meaning asymmetry explains the attitude asymmetries, secondly, a meaning asymmetry between a speaker and her interpreter relies on the fact that he has to interpret her, thirdly, many everyday cases of communication between speakers in a linguistic community do not, as a matter of fact, require any one person's interpretation of another and, fourthly, the attitude asymmetries are present in everyday cases of communication between speakers in a linguistic community.[69] The apparent tension that the third claim introduces is that, if the first three claims are true, it suggests the fourth claim is not true. The first and second claims were justified in chapters 3 and 4, the fourth in chapter 2. Sections 1.2 and 1.3 below will explicate the third, while section 1.1 will further examine the relation between interpretation and the meaning asymmetry, thereby clarifying exactly what the tension is between the four claims.

University of Cape Town

1.1. THE RELATIONSHIP BETWEEN INTERPRETATION AND THE MEANING ASYMMETRY

Radical interpretation (with its necessary constraints) can explain a meaning asymmetry between a speaker and her radical interpreter, since, firstly, the fact that she applies her words consistently[70] implies that she immediately and authoritatively knows what she means and, secondly, her radical interpreter has to interpret her behaviour to know what she means.[71] Since this is the traditional interpretationists' explanation for the meaning asymmetry, they will have to use it to explain the attitude asymmetries between speakers and their interlocutors in their linguistic communities. In other words, they will have to say, firstly, that the requirements and constraints of radical interpretation can be operative in practice, even if there is no one that is actually interpreting anyone else, secondly, that the fact that it is still operative justifies a meaning asymmetry between the speaker and her interlocutors and, thirdly, that this meaning asymmetry between the speaker and her interlocutors then explains the attitude asymmetries between the speaker and her interlocutors.

This is problematic, though. In practice our interlocutors often do not, and do not have to, interpret our behaviour to know what we mean, a claim that Davidson himself subscribes to and that I will explain below. Now, if an explanation of the meaning asymmetry between speakers and their radical interpreters relies on the fact that such radical interpreters have to interpret their behaviour to know what they mean, and in practice our interlocutors often do not have to interpret our behaviour to know what we mean, then it is not clear how radical interpretation and its constraints can secure a meaning, and thereby the attitude, asymmetries between a speaker and her interlocutors. The fact that the constraints of radical interpretation are operative in the practical case can prove that the speaker is applying her words consistently, and it can show why there would be a meaning asymmetry between the speaker and anyone who would have to interpret her behaviour to know what she means. But it cannot, without sufficient argumentation, be claimed that such a meaning asymmetry (and thereby the attitude asymmetries) can be extended to hold between the speaker and those who do not have to interpret her. In other words, if the meaning asymmetry between a speaker and her radical interpreter holds in virtue of his need to interpret her, and our

University of Cape Town

interlocutors often do not have to interpret us, then what can account for the meaning asymmetry (and thereby the attitude asymmetries) in practice?

1.2. INTERPRETATION IN THE CONTEXT OF LINGUISTIC CONVENTIONS

Before proceeding to Davidson's view of the relation between interpretation and conventions, I want to provide some reasons for thinking that we do not, as a matter of fact, interpret a speaker's behaviour to know what she means. That is, everyday situations where language communities share conventions of how to use and interpret statements do not involve, or require, the use of interpretation. When I speak, my interlocutors do not interpret to check whether they have the meaning of my phrases right, or that they have their meaning at all. They wouldn't be able to do this since they almost always lack the type of information required to do so. For example, they understand my speech easily when they first meet me, without having had access to any information about my behaviour and speech in the context of the specific environmental events we discuss, and a lot of interactions with my interlocutors occur far from the environmental conditions under discussion. They will present specific objects and events to me and prompt me to respond only when we are not communicating successfully. If they thought that I did not use the same conventions as them, they would have asked me for verbal and behavioural evidence aimed at environmental objects/events much more often than they actually do. Even in cases where we are not communicating successfully, they are more likely to relate the phrase that they are trying to understand to more basic phrases that they assume I share. The environmental objects and events that people talk about are often not easily reached when we talk about them, so they do rely on the conventional use of other phrases to make sense of the new expressions I produce.

We can treat each other as above precisely because we recognise that other English speakers speak roughly in the same way we do since they speak the same language as us with the same rules we observe.

University of Cape Town

1.3. DAVIDSON'S VIEW OF LINGUISTIC CONVENTIONS

Davidson has often been accused of denying that we use linguistic conventions when we communicate. A more detailed discussion of his view of conventions will help to show that this is actually not what he says.

One of Davidson's remarks, which has since become one of his most frequently quoted and misunderstood claims, is the following: "there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed".[72,73] Davidson identifies the "standard view" of language as a system of convention-governed meanings shared by a linguistic community. On this view, we acquire an enormous collection of words and phrases together with rules for their application to objects and events and rules for relating them to each other. These rules are shared by everyone in our linguistic community. They are applied whenever situations arise in which they dictate how we should speak in order to be understood.[74] According to this "standard view" such linguistic conventions are necessary for successful communication and, if there are such things as individual idiolects, they are deviations from the norms of our linguistic community.

Davidson observed two phenomena that led him to believe that conventions could not be necessary for successful communication. The first, and most frequently cited, is the occurrence of strange speech acts that we manage to understand, even though they cannot be included in the rules for our language. One example is malapropisms, or the confused use of words in which an appropriate word is replaced by one that resembles it in sound or spelling, but has an absurdly inappropriate meaning. "Lead the way and we'll precede", "The plane will be landing momentarily" or "The flood damage was so bad they had to evaporate the city". If the standard view was correct that linguistic conventions were necessary for successful communication, then we would not have been able to interpret any such sentences correctly. Neither would we have been able to make sense of mixed metaphors such as "I'll burn that bridge when I come to it", "Deaf as a doornail" and "Don't ruffle the boat", (don't rock the boat/ruffle his feathers). And communication involving spoonerisms (named after the Reverend William Archibald Spooner), in which corresponding consonants or vowels are transposed, would have had minimal success. Two of Spooner's well known slips

University of Cape Town

were to tell a student "You have tasted two worms" (wasted two terms) and to claim in a sermon that "The Lord is a shoving leopard" (loving shepherd).[75] These phenomena are not the type of things that can be built into linguistic conventions since all the possible slips that people can make cannot be anticipated and coded into rules for language use.[76] Accordingly, Davidson argues that, if we agree that such slips cannot be included in linguistic conventions, and we further appreciate that our hearers can make sense of such confused sentences when they are uttered, then we have to admit that some of our successful communication takes place in the absence of conventions. And if the absence of conventions in such cases does not prevent us from communicating successfully, then linguistic conventions cannot be a necessary condition for communicative success.

Davidson's second observation, made in anticipation of being criticised for drawing his conclusion from such a small number of cases, is that two people hardly ever share all words and ideas of how exactly the use them. It is not unusual to have a perfectly good conversation with someone, even though we do not know all the words they use. When we encounter a phrase that we do not understand, we are often able to infer the meaning via, as Davidson puts it, "wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely."[77,78,79] In these cases we do not employ conventions either.

If the point of language is communication (which is a claim that Davidson accepts as true),[80] and communication can occur without conventions, then it is plausible to believe that the use of linguistic conventions is not a necessary condition for language. Thus, if the point of language is to communicate, and we quite frequently manage to communicate by creatively interpreting individual idiolects with their unique word applications (what many may want to call linguistic errors or ignorance), then our traditional concept of a language needs to be modified to apply to individual idiolects, rather than to communally shared rules for word and phrase application. Hence, Davidson's widely discussed claim with which this section began: there is no such thing as a language in the sense of a system of shared convention-governed meanings.[81]

University of Cape Town

Davidson has been accused of failing to distinguish between necessary and sufficient conditions for communication. Those who hold this view think that the most Davidson's argument shows is that linguistic conventions might not suffice for every case of linguistic communication, but that cannot be taken to mean that they're not necessary for communication.[82]

This objection fails to recognise the significance of the argument, however. Davidson does not think that the situations described above are exceptions because their meanings have to be derived from an individual idiolect, as opposed to a language based on conventional meanings. He thinks, firstly, that meaning is derived from instances of successful communication, secondly, that there are many cases of successful communication without the use of conventions and, thirdly, that, if conventional language had to be removed completely (such as when communicating with someone with whom one shares no language), successful communication will still be almost guaranteed, provided that the interpreter is making use of an appropriate method of interpretation and in possession of adequate information about the speaker. Consequently, all meaning is derived from individual idiolects and where conventional language use happens to coincide with the meanings of an individual idiolect, it is nothing but a fortunate accident. Davidson is, thus, not guilty of confusing the distinction between necessary and sufficient conditions. Conventions are neither necessary for successful communication because we would, in a case where no language is shared, be able to get along quite well without them.

Davidson has also been accused of conflating linguistic meaning and speaker meaning.[83] In the case of Richard Sheridan's Mrs. Malaprop (which is the example Davidson used in his original argument), the speaker meaning of her statement might have been "Sure, if I apprehend any thing in this world it is the use of my vernacular tongue, and a nice arrangement of epithets." That is, after all, how a creative interpreter would have understood her and it is also how she would have wanted to be understood. But the linguistic meaning of her utterance still remains, "Sure, if I reprehend any thing in this world it is the use of my oracular tongue, and a nice derangement of epitaphs." This utterance makes no sense according to the rules of the language of English, which is what enables us to say that Mrs. Malaprop's use of English is incorrect.

University of Cape Town

Davidson does not conflate speaker- and linguistic meaning, though. He is quite aware of the distinction (and this is the point where he explicitly admits that we do use linguistic conventions when communicating).[84] He just thinks that, since conventions are not necessary for communication while interpretation is, the distinction needs to be reworked to respect that meaning requires interpretation rather than the application of linguistic rules. As a result, what has traditionally been identified as philosophically unimportant and labelled speaker meaning, Davidson thinks is of great Philosophical importance since it is the source of meaning. And what has traditionally been considered to be of great Philosophical significance and called "linguistic meaning", Davidson relegates to the status of a philosophically uninteresting practice.[85] It is philosophically uninteresting because the only reason why we try to speak as others do, and why we are prepared to bring our speech in line with the standard use and closely obey linguistic conventions, is that we do not want to be seen as ignorant of the most common use of English. As Davidson says, "Using a word in a non-standard way out of ignorance may be a faux pas in the same way that using the wrong fork at a dinner party is, and it has as little to do with communication as using the wrong fork has to do with nourishing oneself, given that the word is understood and the fork works."[86]

This places us in the position to correctly understand the original Davidsonian claim that there is no such thing as a convention-governed language. If the point of language is to communicate successfully, and successful communication is possible without the use of conventions, then the phenomenon that makes communication possible is actually the creative interpretation of individual idiolects rather than a shared language. This does, however, not imply that he denies that languages like English are, in some ordinary sense, spoken.[87] As previously mentioned, he recognises that, within linguistic communities, we will speak pretty much as others do and that such conformity is encouraged. It does, after all, simplify communication greatly not to have to interpret every utterance from scratch. Phenomena like malapropisms are important, however, since the theoretical possibility of communication without shared practices demonstrate that such sharing cannot be an essential constituent in meaning and communication. Davidson is not trying to make a point about what it is that we, as a matter of fact, do. He is trying to make a philosophical (rather than an empirical)

University of Cape Town

point about the necessary conditions for communication. That is why he does not deny the existence of, say English, as a language. He has made it clear that, in practice, we employ shared linguistic rules. In his own words,

> I do not think we normally understand what others say by consciously reflecting on the question what they mean, by appealing to some theory of interpretation, or by summoning up what we take to be the relevant evidence. We do it, much of the time, effortlessly, even automatically. We can do this because we have learned to talk pretty much as others do, and this explains why we generally understand without effort much that they say.[88]

Since Davidson admits, firstly, that we mostly speak like others do and, secondly, that this frequently allows us to use conventions without interpretation, he essentially concedes that, in practice, the process of radical interpretation is not always involved in knowing what someone means. In cases where we do use conventions to know what our interlocutors mean, radical interpretation without conventions has to be possible.[89] Davidson, thus, concludes that radical interpretation is something that we often do not need to do in practice, even though we potentially could.

1.4. THE PROBLEM

Now, if the meaning asymmetry between a speaker and her interpreters requires that those interpreters radically interpret her, then we want to say that a meaning asymmetry between a speaker and her interlocutors requires that her interlocutors radically interpret her. (1.1). Radical interpretation is, however, something that we often do not do in practice (1.2-1.3). This suggests that the meaning asymmetry is something that is often not present in practice. And if the meaning asymmetry is something that is often not present in practice, then we are left without an explanation for the attitude asymmetries between speakers and their interlocutors in their linguistic communities.

In Section 2 I shall consider four possible strategies for accommodating the meaning asymmetry in practice, together with my reasons for thinking that they all fail to explain the attitude asymmetries described in Chapter 2. A successful strategy will have to be able to support two claims. It will have to show, Firstly, that the constraints of radical interpretation can continue into situations of communication where we do

University of Cape Town

not actually use interpretation and, secondly, that the theoretical constraints of radical interpretation can explain a meaning asymmetry, and thereby the attitude asymmetries, between a speaker and her interlocutors in her linguistic community. My contention in the remainder of this chapter is that, even if traditional interpretationists can show that radical interpretation and its constraints are operative during communication based on linguistic conventions (rather than on interpretation), the type of meaning asymmetry that it implies is not appropriate to explain the attitude asymmetries between speakers and their interlocutors with whom they share linguistic conventions.

2. THE MEANING ASYMMETRY IN PRACTICE

2.1. INTERLOCUTORS AS SPECTATORS

Traditional interpretationists may argue that it is true that, firstly, our interlocutors do not interpret us when they use conventions to understand what we mean, and, secondly, that radical interpretation continues even in situations where conventions are used.

For all the reasons mentioned previously, our interlocutors do not interpret us in situations where they use conventions to understand what we mean. But the use of conventions does not rule out interpretation (and thereby the meaning asymmetry) since it does not rule out the possibility that the speaker is radically interpretable. As Davidson's theory of conventions suggests, if an interlocutor decide to discard all linguistic conventions, the speaker will still be interpretable to him. Even if this feat is deemed impossible, it still does not weaken the idea that the speaker remains radically interpretable. If a genuine radical interpreter, who does not speak her language at all, were to observe a speaker, she would still be radically interpretable to him, whether she is functioning in a linguistic community or not. He will be in a position to fix what she means by radically interpreting her. Consequently, if the use of her words (and the possibility of being radically interpreted as applying them consistently) fixes their meaning, then the speaker knows what she means while her radical interpreter still has to interpret for meaning. That is, even in a situation where conventions are

University of Cape Town

used, the speaker knows, while her radical interpreter may not know, what her words mean, which is the meaning asymmetry.

Traditional interpretationists can further claim that our interlocutors in our linguistic communities are not in a position to know what we mean, so whatever they know does not play any role in what we mean and, consequently, in what we know we mean. They are thereby relegated to spectators of the notion of meaning and the process of fixing it. Since our knowledge of what we mean is grounded by our radical interpreter's interpretation of us, our interlocutors knowledge of what we mean is likely to be mistaken because they do not have anything to do with fixing what we mean. Where they are correct about what we mean, they are simply lucky that their conventions coincide with the meaning that a radical interpreter finds in our personal idiolects. After all, the ultimate evidence for what someone means and believes is derived from the radical interpretation of her speech and behaviour, so the conventions that speech communities make use of are irrelevant to meaning and belief content. The meaning asymmetry between a speaker and her radical interpreter, thus, gives rise to a large meaning asymmetry between the speaker and her interlocutors, since her interlocutors are in a poor position to know what she means.

To summarise this strategy: In a situation where linguistic conventions are used, radical interpreters can know what I mean through interpretation. The fact that I am radically interpretable shows that I know what I mean. My interlocutors cannot know what I mean without checking their conclusions with a radical interpreter (or with me, of course, assuming that I am radically interpretable).[90] This means that I know what I mean while my fellow English speakers may not know what I mean. Therefore, in a situation where linguistic conventions are used, the meaning asymmetry can explain the presence of the attitude asymmetries.

The attitude asymmetries explained by this version of the meaning asymmetry will look as follows:

University of Cape Town

The authority asymmetry

1. The speaker knows what she means because she is radically interpretable and the regular application of her sentences fixes their meaning.
2. Her interlocutors do not know what she means because whatever knowledge their linguistic conventions give them about her does not settle what she means (unless there is a convergence between such claims and those of the radical interpreter).
3. Thus, the speaker knows, while her interlocutors do not know, what she believes.

The evidence asymmetry

1. The speaker knows what she means without using her own behaviour and speech as evidence because the regular application of her sentences fixes their meaning.
2. Her interlocutors have to use her behaviour and speech to know what she means since their conventions do not settle what she means (unless there is a convergence between such claims and those of the radical interpreter).
3. Thus, the speaker knows, while her interlocutors cannot know, what she believes without employing her behaviour and speech as evidence.

The transparency asymmetry

1. The speaker knows the meaning of the sentences that she holds true because their regular application fixes their meaning.
2. Her interlocutors cannot know the meaning of the sentences that she holds true without using her behaviour and speech as evidence since their conventions do not settle what she means (unless there is a convergence between such claims and those of the radical interpreter).
3. Thus, the speaker self-attributes beliefs that are consistent with what she holds true about the world while her interlocutors have to attribute beliefs to her in the service of an explanation of her behaviour and speech.

University of Cape Town

This strategy gives the best account of the meaning asymmetry and, thereby, of the attitude asymmetries in practice. It does, however, rely on the likelihood that the people with whom we share linguistic conventions are often mistaken about what we mean since they are not employing an appropriate method of radical interpretation to interpret us. This is, of course, true, but it cannot be employed to argue that the conclusions that our interlocutors draw about what we mean are of no significant theoretical use at all. Even if Davidson's suggestion is accepted that such conclusions cannot constitute our linguistic and propositional content in the way that an omniscient radical interpreter's can, the reality remains that they are, more often than not, right about what we mean, without any interpretation at all.[91] The project of sorting people into linguistic communities based on information obtained via radical interpretation dictates this. If a speaker frequently deviates from the linguistic conventions of her community, then Davidson's principle of charity requires that we assume that she is speaking a different language, which an appropriate method of radical interpretation will assign to her. In such cases her community will also treat her as someone who is speaking, either a completely different language, or a different English dialect (with some of its own conventions). If I share a language with others whose radically interpreted evidence places them in the same speech community as me, then my everyday speech is likely to be very similar to the speech from which the radical interpreter constructs his scheme about me. And people are able to unreflectively understand that, exactly because it is likely to be fairly similar to the raw evidence that they provide their radical interpreters with. And then it is not clear why, even if not perfectly, others would not be in a great position to know my language. If our linguistic community is correct about what we mean a lot of the time, then we cannot merely dismiss their conclusions about us as insignificant or irrelevant to the matter at hand. If our speech community can more often than not reach the same conclusion about what we mean through conventions while the radical interpreter has to interpret, our speech community more often than not appears to have an immediate route to our meaning which even the radical interpreter lacks.

Traditional interpretationists may continue to claim that the frequency with which our interlocutors manage to get the meaning of our statements right is unimportant, since

University of Cape Town

their judgements are made to be correct by the judgements of a radical interpreter. Whatever conclusion they draw is reached through luck, rather than insight about the speaker's speech and behaviour. Hence, the fact that a radical interpreter can judge that she is, like everyone around her, speaking English does not imply that her interlocutors have a way of knowing what she means.

This underestimates the strength of my argument, however. It is not simply the case that, since our interlocutors happen to be right about what we mean a lot of the time, they have to know what we mean. The judgements that our interlocutors make about what we mean is almost guaranteed to correspond with the judgements reached by an interpreter since we deliberately try to speak very much like the other people in our linguistic community do. It might be right that we speak very much like those in our linguistic community because we do not want to be seen as ignorant or unintelligent, but the fact remains that we deliberately alter our speech to correspond with theirs. The longer the time that someone spends in a specific speech community, the more likely it becomes that there will be very few differences between her speech and the speech of her interlocutors. We do not want others to think that we are inarticulate or ignorant of the common use of English and we are eager to make ourselves understood. So we calculatingly learn to speak like they do. If this is true, then it is difficult to know why traditional interpretationists claim that radical interpretation can ground a large meaning asymmetry between ourselves and our linguistic communities.

University of Cape Town

To recap: in a situation where linguistic conventions are used, our radical interpreters will construct very similar individual idiolects for us since we often speak in the same way (even if it is by accident rather than convention). My interlocutors will recognise that I speak the same language as them since our personal idiolects are very similar. The similarity of our personal idiolects ensures that we will agree on many conventions to conform to in order to simplify communication. Linguistic conventions will often lead my interlocutors directly to conclusions that correspond with those my radical interpreter reaches via interpretation. So, a lot of the time, my interlocutors will be in as good a position as myself to know what I mean. Therefore, in situations where conventions are used, this first approach towards the meaning

asymmetry can explain the attitude asymmetries, but only at the expense of a plausible account of interpretation and conventions.

2.2. INTERLOCUTORS AS WELL INFORMED CONVENTION USERS

A second strategy that traditional interpretationists might use is to construct a situation very much like the one above, but to integrate my comments. Once again, it is true, firstly, that our interlocutors do not interpret us when they use conventions to understand what we mean, and, secondly, that radical interpretation continues even in situations where conventions are used.

Here our interlocutors use conventions, instead of an appropriate method of radical interpretation, to understand us. Their conclusions can, again, not constitute our linguistic and propositional content in the way that an omniscient radical interpreter's can. They do, however, happen to be right about what we mean more often than not, without engaging in any interpretation at all.[92] As explained above, the project of sorting people into linguistic communities based on information obtained via radical interpretation implies this and the fact of deliberately trying to speak like others in our linguistic communities do strengthens it. Our interlocutors are, accordingly, in a very good position to know our language.

This does not imply that there is no meaning asymmetry between speakers and their interlocutors, though. In this situation the speaker knows what she means since she is speaking a language that is interpretable to her radical interpreter. Her interlocutors are mostly able to know what she means because she is using the same conventions that they are. There thus appears not to be a meaning asymmetry, but this appearance is misleading since those conventions may at any time become incapable of explaining her speech. If she, for example, deliberately or accidentally stops using them or if she uses a convention or sentence that her interlocutors are not familiar with, they may at any point become wrong about what she means. In the radical case, the meaning asymmetry rests on the fact that the radical interpreter may be wrong about what she means since he has to continue interpreting her. Even if he is an omniscient interpreter of her, he has to continue interpreting her. And then his conclusions about what she means may at any point be wrong since his interpretive

University of Cape Town

scheme may at any point become inappropriate to make sense of her; that is, his past successful interpretations cannot guarantee that he will continue to be right in the future. The fact that he has to continue to collect evidence and interpret it in the light of everything else he knows about her, leaves room that his conclusions may be mistaken. Similarly, in practice, our interlocutors may at any point reach mistaken conclusions about what we mean since they rely on conventions that may at any point become inappropriate to make sense of us.

Therefore, in situations where linguistic conventions are used, a speaker knows, while her interlocutors may not know, what she means.

The attitude asymmetries explained by this version of the meaning asymmetry will look as follows:

The authority asymmetry

1. The speaker knows what she means because she is radically interpretable and the regular application of her sentences fixes their meaning.
2. Her interlocutors are mostly right, but may be wrong, about what she means since their conventions may at any point become unable to make sense of her.
3. Thus, the speaker knows, while her interlocutors may not know, what she believes.

The evidence asymmetry

1. The speaker knows what she means without using her own behaviour and speech as evidence since the regular application of her sentences fixes their meaning.
2. Her interlocutors mostly do not have to, but at any point may have to, use her behaviour and speech as evidence to know what she means.
3. Thus, the speaker knows what she believes without using her own behaviour and speech as evidence while her interlocutors may have to use such evidence to know what she believes.

University of Cape Town

The transparency asymmetry

1. The speaker knows the meaning of the sentences that she holds true since their regular application fixes their meaning.
2. Her interlocutors mostly do not have to, but at any point may have to, use her behaviour and speech as evidence to know the meaning of the sentences that she holds true.
3. Thus, the speaker self-attributes beliefs that are consistent with what she holds true about the world while her interlocutors at any point may have to attribute beliefs to her in the service of an explanation of her speech and behaviour.

OBJECTIONS

The first strategy can justify plausible versions of the attitude asymmetries but contains some problematic assumptions about interpretation and conventions. This strategy, on the other hand, seems to justify the existence of a meaning asymmetry in practice without giving an inappropriate picture of interpretation and conventions, but the attitude asymmetries that it is able to explain are too weak.

The authority asymmetry that it is able to explain is similar to the one described in chapter 2 and is thus as it should be. In case of the evidence asymmetry, however, saying that our interlocutors may have to (but mostly do not have to) use a speaker's evidence to know what she believes is just not strong enough. The original evidence asymmetry holds that we assume that our interlocutors usually employ such evidence in order to know what we believe. Similarly, in the case of the transparency asymmetry, it is not sufficient to say that our interlocutors may have to (but mostly do not have to) attribute beliefs to us in the service of an explanation of our speech and behaviour. The original transparency asymmetry holds that we assume that our interlocutors usually attribute beliefs in the service of an explanation of our speech and behaviour.

This inability to explain the proper evidence and transparency asymmetries renders this strategy hopeless.[93]

University of Cape Town

2.3. INTERLOCUTORS AS INTERPRETERS

Another possible traditional interpretationist strategy is that, while conversing with those in our linguistic communities, the meaning asymmetry (and consequently the attitude asymmetries) continue since our interlocutors always take an interpretive stance towards us. On this view, the meaning asymmetry is grounded, not by the possibility of being radically interpretable outside our linguistic communities, but by being interpretable within our linguistic communities.[94]

This response does not contradict Davidson's view that we often understand what our interlocutors mean since "we have learned to speak pretty much like they do". But even in cases where we do use conventions, we are still using interpretation since we always have to check that the conventions that we use make accurate sense of what the speaker says and does. We never simply use conventions to understand a speaker. We use conventions together with an interpretation of whether the conventions in specific cases are appropriate. Thus, even cases where we use linguistic conventions involve interpretation in the form of checking that our conventions make sense of what a speaker says and does. And if there is an omniscient interpreter with whom the speaker shares linguistic conventions, then such an interpreter will function very similarly to a radical interpreter. The difference is that he will use conventions together with a judgement of which sentence the speaker holds true and what her behaviour and environment suggest she means.

A possible justification for this idea is that the second strategy above assumes that I can know that I speak the same language as someone else. But traditional interpretationists may ask how I can know whether I share a language with the speaker or not? We cannot say that I can know this when I hear familiar sounds arranged in familiar ways. Hearing familiar sounds arranged in familiar ways, which has up to now been treated as an unacceptable appeal to phenomenology anyway, cannot give me the knowledge that I share a language with someone else. I will have to judge that those familiar sounds are applied to the same sort of things in the environment that I apply them to, otherwise I will know only that others make the same sort of sounds as I do. And knowing what a speaker applies her words to is something that can only be achieved through radical interpretation; that is, through the

University of Cape Town

employment of her behaviour, speech and environment in order to interpret the phrases that she seems to hold true. Thus, traditional interpretationists may claim that the use of conventions alone cannot give me the knowledge that I am speaking the same language as my interlocutors. I have to perform some interpretation to place them in the same speech community as myself. And I have to continue to interpret them if I want to know that they continue to speak the same language as I do. Thus, even when we do use conventions to understand what our interlocutors mean, we are still taking an interpretive stance towards them. We are not only ready to interpret at the smallest suspicion of unusual or unpredictable speech, but we also always judge whether to apply the conventions by recalling their other behaviour and speech.[95]

If a speaker, for example, claims that it is raining, her interlocutor will use the conventions of the English language to assign meaning to her utterance, which (as I argued above) seems to give him a good chance of being right about what she means. But he assigns the meaning to her utterance, not by consulting a list of conventions only, but by recalling any of her other behaviour and speech that may or may not contradict the conventional meaning, considering whether the conventional meaning is appropriate in their current conversation or environment, working out what else she could have meant if not, and so forth. The justification for the meaning asymmetry is, thus, the same as in the radical case. Here it is just an interlocutor that is judging her against the linguistic conventions that he is applying to her.

The attitude asymmetries explained by this version of the meaning asymmetry will look as follows:

The authority asymmetry

1. The speaker knows what she means since her interlocutors interpret her and the regular application of her sentences fixes their meaning.
2. Her interlocutors may be wrong about what she means since their interpretations may at any point be wrong.
3. Thus, the speaker knows, while her interlocutors may not know, what she believes.

University of Cape Town

The evidence asymmetry

1. The speaker knows what she means without using her own behaviour and speech as evidence since the regular application of her sentences fixes their meaning.
2. Her interlocutors have to use her behaviour and speech as evidence to know what she means since they are her radical interpreters.
3. Thus, the speaker knows what she believes without using her own speech and behaviour as evidence while her interlocutors have to employ such evidence to know what she believes.

The transparency asymmetry

1. The speaker knows the meaning of the sentences that she holds true since their regular application fixes their meaning.
2. Her interlocutors may not know the meaning of the sentences that she holds true since they have to interpret for meaning.
3. Thus, the speaker self-attributes beliefs that are consistent with what she holds true about the world while her interlocutors attribute beliefs to her in the service of an explanation of her behaviour and speech.

University of Cape Town

OBJECTIONS

Like the approach in Section 2.1, this strategy is able to explain the sort of attitude asymmetries that we want, but in this instance it cannot account for some of the observations made in Section 1.2. It may be true that we often keep in mind a speaker's behaviour in the context of her environment when she speaks. But this does not imply that we are actually interpreting. And there seem to be good reasons for claiming that actual interpretation of a speaker's speech against both conventions and all other information cannot be the usual course of events.

Firstly, in the case of a new acquaintance, we have access only to what she says. We know nothing about her behaviour and speech in the context of the environmental events and objects that she is talking about. There is almost nothing to interpret

against our conventions, even if we do happen to adopt an interpretive stance towards her.[96] In the radical case, the meaning asymmetry rests on the idea that, since a radical interpreter has to interpret the speaker, he may be wrong about what she means. In the case of new acquaintances, since her interlocutors attribute conventional meaning without access to evidence of what she says under which environmental conditions, the meaning asymmetry will have to be lost.

Secondly, on the other extreme, in the case of a very close relation, we may be in possession of a lot of information about her, but we will also be aware of a lot more linguistic conventions that we share. Once again, there will be almost nothing to interpret. Even if we agree with such traditional interpretationists that we do use some interpretation to verify that a speaker is speaking the same language as us, there appears to be little to justify the claim that we go on interpreting a speaker to continue verifying that she is speaking the same language as us. This is precisely the type of objection that persuaded Davidson himself that we did not always use interpretation when understanding the speech of someone in our linguistic community. The more confident we become that we speak the same language with the same conventions, the less evidence about a speaker's behaviour and environment we use. That's exactly why Davidson is not able to claim that we use radical interpretation in practice. He has to talk about the more modest "creative interpretation". Instead of being able to claim that we use the behaviour of a speaker in the context of our environment as evidence for our interpretation, he has to talk about ""wit, luck, and wisdom from a private vocabulary and grammar, knowledge of the ways people get their point across, and rules of thumb for figuring out what deviations from the dictionary are most likely."[97]The knowledge that we share a lot of conventions simplifies communication, since it allows us to use evidence only in cases where the speaker produces a phrase that we do not understand at all, or a surprising phrase (like a malapropism or spoonerism). And even in such cases the evidence does not concern the speaker's behaviour and speech in the context of the environmental objects/events that she is talking about, but instead likely deviations from conventions that we are certain she shares with us. It, thus, seems as if we use a lot less evidence in interactions with close relations/friends with whom we share a language with its conventions, and even this smaller amount of evidence is used only in cases where speakers produce surprising speech.

University of Cape Town

The two preceding paragraphs show that, if the interpretation of other evidence against our conventions is meant to ground the meaning asymmetry, then this strategy cannot explain the general form of the meaning asymmetry (and consequently the attitude asymmetries). It implies that, at least in situations with new acquaintances and close relations, the interpretation of gathered evidence against our conventions is not the usual course of events. If the meaning asymmetry is meant to spring from our interlocutors' interpretations of us, then at least in these situations we cannot assume that the speaker usually knows, while her interlocutors may not know, what she means (and believes).

2.4. INTERLOCUTORS AS POTENTIAL INTERPRETERS

The last strategy open to a traditional interpretationist is to use our interlocutors as potential, as opposed to actual, radical interpreters. All the responses above attempt to maintain that there is some actual meaning asymmetry between the speaker and her interlocutors based on some actual interpretation that they either do or should do. But why do we have to assume that her interlocutors actually have to interpret her? Just so long as she is radically interpretable to her interlocutors, even if they do not, in fact, interpret her, then the meaning asymmetry between her and her interlocutors can be retained.

This parallels the argument for the meaning asymmetry in the radical case. The idea of a fully informed radical interpreter is meant to be an abstract one. It does not describe some actual individual. So, whether there is a fully informed radical interpreter interpreting or not, there is always a meaning asymmetry present. Any situation where someone speaks is a situation of radical interpretation since, to qualify as speaking a language and entertaining propositional content, a speaker has to be interpretable to a fully informed radical interpreter. Whether anyone is actually interpreting or not is not important. Whether someone will in principle be able to interpret is what grounds the meaning asymmetry. In other words, even in situations where no one is actually interpreting there is a meaning asymmetry since the fact that there is no actual interpretation does not suddenly render the speaker uninterpretable.

University of Cape Town

Analogously, during conversations between a speaker and her interlocutor there is no one that is interpreting since the communication will be based mainly on their shared linguistic conventions. But the fact that there is no one that is actually interpreting does not suddenly render the speaker uninterpretable. And if the speaker remains interpretable, then they are communicating in a situation of radical interpretation coupled with its meaning asymmetry

The attitude asymmetries explained by this version of the meaning asymmetry will look as follows:

The authority asymmetry

1. The speaker knows what she means since she is radically interpretable and the regular application of her sentences fixes their meaning.
2. Her interlocutors may not know what she means since, if they had interpreted her, they may not have known what she meant.
3. Thus, the speaker knows, while her interlocutors may not know, what she believes.

The evidence asymmetry

1. The speaker knows what she means without using her own speech and behaviour as evidence since the regular application of her sentences fixes their meaning.
2. Her interlocutors have to use her speech and behaviour to know what she means since, if they had interpreted her, then they would have had to use it.
3. Thus, the speaker knows what she believes without using her own speech and behaviour as evidence while her interlocutors have to use such evidence to know what she believes.

The transparency asymmetry

1. The speaker knows the meaning of the sentences that she holds true since their regular application fixes their meaning.

University of Cape Town

2. Her interlocutors have to use her behaviour and speech to know the meaning of the sentences that she holds true since, if they had interpreted her, they would have had to use it.
3. Thus, the speaker self-attributes beliefs that are consistent with what she holds true about the world while her interlocutors attribute beliefs to her in the service of an explanation of her speech and behaviour.

OBJECTIONS

This strategy is of no use in explaining the attitude asymmetries between a speaker and her interlocutors since it is merely a restatement of Davidson's original arguments for radical interpretation and for the meaning asymmetry as a conceptual picture of the mental. Chapter 4 argues that the meaning asymmetry coupled with radical interpretation, when understood as a conceptual model of the mental, is able to explain why, in theory, the attitude asymmetries are always present. What this chapter is attempting to accomplish is to find a way in which this conceptual picture of the mental can be extended to explain what we, as a matter of fact, do in practice.

The point in this case is precisely that her interlocutors are not mythical characters called radical interpreters who serve to illustrate the nature of the mental. They are people with whom we share a language with its conventions who mostly do not have to interpret us to know what we mean. Casting them in the role of potential radical interpreters, like this strategy does, still does not tell us what exactly is going on in a situation where we are conversing with our interlocutors. We already know that a radical interpreter has to find the speaker's speech interpretable for the meaning asymmetry to continue. But that will be a meaning asymmetry between a speaker and a radical interpreter. Throughout this chapter our strategies to translate the meaning asymmetry into practice have failed since we have been unable to find the role that our interlocutors play in interpretation and in the meaning asymmetry. We have discovered that they are not interpreters since they do not observe our speech and behaviour in the context of our environment to judge what we mean. And regarding them as potential (as opposed to actual) radical interpreters leaves the question unanswered as to what they actually are and how our actual relations with them can be explained. The strategy of rejecting the role of our interlocutors in favour of

University of Cape Town

turning back to the conceptual picture of the mental can be read as a reluctant admission that the conceptual picture cannot be translated into practice.[98]

3. CONCLUSION

A meaning asymmetry appropriate to the task of explaining the attitude asymmetries can, therefore, not be accommodated in situations where we use linguistic conventions, which is what we do almost all of the time. Even if we agree with traditional interpretationists that a meaning asymmetry is present in the radical case, and that the meaning asymmetry in the radical case can explain why the attitude asymmetries (understood as general assumptions) are always present, it is all of very little use if the meaning asymmetry cannot fit into what we actually do in our linguistic communities. Such traditional interpretationists follow Davidson by admitting that we use conventions in practice, but they simply do not take such cases of convention use seriously enough. They may be right that we can, in principle, stop using conventions and that successful communication will continue. As argued throughout, it does seem as if it is right that Davidson's system of radical interpretation can give a plausible account of the nature of the mental. But in practice we do use conventions and we will not stop using them, and it is at this level that the meaning asymmetry interpretation collapses. We need a theory that either explains the attitude asymmetries between ourselves and our interlocutors during interactions with them within our linguistic communities, or one that can at least be extended to explain them. The meaning asymmetry alone is clearly not intended to explain them directly, since it is embedded in a system of radical interpretation. It also seems impossible to extend it to accommodate conventions, since it cannot be separated from a system of radical interpretation. We, as a result, have a theoretically coherent picture of the mental that cannot be developed to be anything more than just that.

University of Cape Town

CHAPTER 6
THE SENTENCE HELD-TRUE ASYMMETRY TO THE RESCUE

In this chapter, I propose to explain a strategy whereby Davidson's system of radical interpretation can accommodate the three attitude asymmetries. In section 1, I intend to explicate my solution by applying it to the asymmetries, by justifying each individual claim that gives rise to it and by explaining its advantages over the meaning asymmetry interpretation. In section 2, 3 and 4, I will defend it from some possible objections that can arise. The previous chapter left us with two dilemmas: first, how exactly can radical interpretation be made compatible with what we do within our linguistic communities and, second, how can we construct an explanation for the attitude asymmetries that applies both in the radical and non-radical cases, given that the two cases are so different. In this chapter both these problems will be solved and it will become clear both how radical interpretation relates to what we do in our linguistic communities and how its accommodation of the attitude asymmetries can be translated into practice.

1. THE SENTENCE HELD TRUE ASYMMETRY

1.1 THE SENTENCE HELD TRUE ASYMMETRY AS AN EXPLANATION OF THE ATTITUDE ASYMMETRIES

Given that our interest is primarily in an explanation for the attitude asymmetries in practice, I will begin by applying my solution to interactions between speakers and their interlocutors within their linguistic communities. In sections 1.2 and 1.3, I will show how it is compatible with radical interpretation.

Davidson's system of radical interpretation can accommodate all three of the attitude asymmetries once we understand that the speaker usually knows, while her interlocutors may not know, which sentences she holds true. Traditional interpretationists treat the held-true sentence symmetrically by assuming that the speaker and her interpreter both know which sentence she holds true, while he, unlike her, then has to interpret to know its meaning. I propose that the meaning asymmetry,

University of Cape Town

in fact, is a sentence held-true asymmetry and that we should treat both knowledge of the sentence that is held true and knowledge of its meaning asymmetrically.

The justification for the sentence held true asymmetry is, accordingly, parasitic upon the meaning asymmetry and will be defended in detail below. For now, it will suffice to say that the fact that the speaker is interpretable as speaking a language means that she is applying her sentences consistently. The fact that she is applying her sentences consistently means that she knows what she means, for all the reasons given in chapter 4. The fact that she usually knows what she means means that she usually knows the truth conditions of her sentences (or what environmental conditions she usually applies her sentences in the presence of). Since she knows which truth conditions her sentences usually have, as soon as she is in a situation where such truth conditions obtain, she usually knows which sentence she holds true.

It is important to note that I am not claiming that the speaker has to wait until those truth conditions obtain before she knows which sentence she holds true in them. The fact that she knows her language means that she always knows the truth conditions of her sentences and that she, as a result, knows which truth conditions have to obtain for her to hold a specific sentence true. Since I am, however, comparing and contrasting her position with that of her interpreter/interlocutor, it is helpful to apply the sentence held-true asymmetry by placing them in a situation where the truth conditions of one of her sentences obtain. My argument for the sentence held-true asymmetry does, however, hold even in situations where the truth conditions of that sentence do not obtain.

If the fact of speaking a language can give the speaker the knowledge of which sentences she usually holds true in the presence of which conditions, then she does not have to wait until she speaks or behaves to know which sentences she holds true. On this strategy the speaker usually knows what she believes, usually does not need to use her own behaviour and speech to know what she believes and usually self-attributes beliefs that are consistent with her view of the world because the fact of speaking a language gives her the knowledge of which sentences she holds true. In practice, her interlocutors may be wrong about what she believes since they may not know what she holds true. They have to use her behaviour and speech to know what

University of Cape Town

she believes since they have to wait for her to utter a sentence and display behaviour that indicates that she holds it true. Similarly, they have to attribute beliefs to her in the service of an explanation of her speech and behaviour since they have to wait for her to utter a sentence and display behaviour that indicates that she holds it true. Her radical interpreter, similarly, has to wait for her to utter the sentences that she holds true. The only difference between the radical and non-radical cases is that her radical interpreter has to interpret for meaning as well, unlike her interlocutor in the first statement of each of the following schematic outlines below).

The attitude asymmetries explained by the sentence held-true asymmetry will look as follows

The authority asymmetry in the non-radical case

1. Both the speaker and her interlocutor usually know what she means by the sentences that she holds true.
2. The speaker usually knows which sentences she holds true when encountering situations in which she holds them true since she is generally interpretable and thus knows the meaning of her sentences.
3. Her interlocutors may not know which sentences she holds true when encountering situations in which she holds them true unless she has uttered them.
4. A belief is attributed by giving a meaningful sentence that a speaker holds true.
5. The speaker usually knows, while her interlocutors may not know, what she believes.

The evidence asymmetry in the non-radical case

1. Both the speaker and her interlocutor usually know what she means by the sentences that she holds true.
2. The speaker usually does not have to use her own behaviour as evidence to know which sentences she holds true when encountering situations in which

University of Cape Town

she holds them true since she is generally interpretable and thus knows the meaning of her sentences.

3. Her interlocutors usually have to use her behaviour as evidence to know which sentences she holds true when encountering situations in which she holds them true since they have to either count on her to inform them or prompt her appropriately to utter them.
4. A belief is attributed by giving a meaningful sentence that a speaker holds true.
5. The speaker usually knows what she believes without using her own speech and behaviour as evidence while her interlocutors have to use such evidence to know what she believes.

The transparency asymmetry in the non-radical case

1. Both the speaker and her interlocutor usually know what she means by the sentences that she holds true.
2. The speaker usually does not have to use her own speech and behaviour as evidence to know which sentences she holds true when encountering situations in which she holds them true since she is generally interpretable and thus knows the meaning of her sentences.
3. Her interlocutors usually have to use her speech and behaviour as evidence to know which sentences she holds true when encountering situations in which she holds them true since they have to either count on her to inform them or prompt her appropriately to utter them.
4. A belief is attributed by giving a meaningful sentence that a speaker holds true.
5. The speaker usually self-attributes beliefs that are consistent with what she holds true while her interlocutors attribute beliefs to her in the service of an explanation of her speech and behaviour.

University of Cape Town

1.2. THE JUSTIFICATION FOR THE SENTENCE HELD TRUE ASYMMETRY

I have defended some of the individual claims that contribute to this strategy previously, so I will repeat some of them only briefly. In this section, I will assume that we are potentially interpretable to a radical interpreter and justify the sentence held-true asymmetry accordingly. In section 1.3, I will explain how our use of linguistic conventions fits in with this picture. Moreover, in the chapters in which these arguments first appeared, I followed Davidson and talked about "the environmental conditions in which a speaker holds a sentence true", whereas in this chapter I talk about "the situations in which the speaker holds a sentence true" or "situations in which the truth conditions of her sentences obtain". The former applies more accurately to the simple perceptual beliefs that Davidson himself illustrated his system with, while the latter describes more accurately the complex beliefs (with their more complex truth conditions) that I will use in this chapter to illustrate my solution. (More about this distinction below.) The arguments below hold just as much for the latter as for the former, so for now I will continue to use simple perceptual beliefs to illustrate.

Firstly, the fact that the speaker is interpretable as speaking a language means that she is applying her sentences consistently. If she is not interpretable to an omniscient radical interpreter under favourable circumstances, then she is clearly not applying her sentences consistently and thus not speaking a language.

Secondly, the fact that the speaker applies her sentences consistently means that she usually knows what she means. As seen in chapter 4, if she applies her sentences consistently, she can use disquotation to state what she means. She can relate her sentences to each other both when applying them to her environment and when attributing meaning and belief to others. She is interpretable to others as knowing what she means since everything she does say, and can possibly say, will indicate both that she knows what she means, and exactly what she knows she means.

University of Cape Town

Thirdly, the fact that the speaker usually knows what she means means that she usually knows the truth conditions of her sentences or what environmental conditions she usually applies them in.

The first justification for this third claim is the definition of meaning. If we agree with Davidson that the speaker knows what she means, and meaning just is the truth conditions of a sentence, then the speaker knows the truth conditions of her sentences. From her own perspective, the truth conditions of a sentence that she holds true are given by the sentence that she holds true in the presence of certain environmental conditions. The sentence that she holds true is that "It is raining", and its truth conditions are given by the sentence that she holds true under those environmental conditions which happens to be that "It is raining". She gives the truth conditions of the sentence that she holds true as, "My sentence 'It is raining' is true if and only if it is raining". The sentence on the left-hand-side of the biconditional is the sentence that she holds true, and the sentence on the right-hand-side is the one that she usually holds true in the same environmental conditions that she holds the sentence on the left-hand-side true in. Thus, if she knows the meaning of one of her sentences, then, by definition, she knows its truth conditions. If she knows its truth conditions, then, by definition, she knows which environmental conditions she usually applies it in. Otherwise her disquotation will not work.

The second justification for this third claim relies on all the justifications thus far, since it requires the speaker to be interpretable as applying her sentences consistently and thereby as the speaker of a language. If she speaks a language, then she is able to be an interpreter of others. In the radical case, the speaker will give the meaning (or truth conditions) of another's sentence by using a sentence in her idiolect that she usually holds true in the same environmental conditions that she thinks he is holding his true in. She hears him say, "It is raining". She judges that he is holding that sentence true in the same environmental conditions that she would have held a sentence "It is raining" true in. From her perspective, she judges which environmental conditions he is holding a sentence true in, figures out which sentence in her language she holds true in those environmental conditions and then, via her perception of her environment and her knowledge of her own language, she produces the sentence that gives the meaning of his sentence. She, in other words, knows which sentences she

University of Cape Town

typically holds true under which environmental conditions, since that is the only way in which she can work out what he means (by mapping his language onto her own).[99] In the non-radical case, she also knows which environmental conditions she holds which sentences true in, since she employs the same conventions as her interlocutor and she understands the sentences that he holds true.

Fourthly, the previous point related to knowledge of meaning in specific cases, rather than to knowledge about what is usually the case. The fact that the speaker knows which environmental conditions she usually applies her sentences in the presence of means that, once she encounters such conditions, she usually knows which sentence she holds true.

She is interpretable to others as knowing which sentence she holds true since everything she does say, and can possibly say, will indicate that she knows which sentence she holds true, and except for cases where she is uncertain about her environment, there is nothing that she can say that will indicate that she does not know which sentence she holds true. There is no evidence on the basis of which interpreters can withhold an attribution of the knowledge of which sentence she holds true once she encounters the environmental conditions that she usually uses it in the presence of. Furthermore, as seen above, when encountering environmental conditions in which her interpreter or interlocutor utters a sentence that he holds true, she knows which sentence she would have held true in those same conditions. She thus cannot encounter any specific set of environmental conditions without knowing exactly which sentence she would have held true in them.

Moreover, if we want radical interpretation to be more than a collection of sounds that an interpreter manages to find meaningful, then we are required to assume that the speaker knows which sentence she holds true when encountering the environmental conditions that she usually applies it in. Davidson himself encourages us to think of speakers as speaking because they want to be understood, and as knowing that they have to apply their sentences consistently in order to be understood. Speakers thus apply their sentences consistently with the intention of providing clues about their meaning to their interpreters. A theory that allows for speakers to be in environmental circumstances without knowing which sentence they hold true will have difficulty

University of Cape Town

accounting for the sense in which language use is intentional. On such a theory the consistency of language application will be a matter of something in the environment causing speakers to produce a specific sequence of sounds. It will allow the speaker to become a spectator of the way in which the environment causes her body to move.

Theories that want to make the speaker a participant in her own linguistic responses to her environment have to allow that she applies her sentences consistently to events that she thinks are apparent to her interpreters with the intention of being understood. And if the speaker is applying her sentences consistently (as judged by an omniscient radical interpreter), then it means that she is applying them consistently to events that she thinks are apparent to her interpreters with the intention of making herself interpretable. If she applies them to something she thinks is apparent to her interpreters, then she apparently knows which environmental conditions she is in. And if she intentionally applies her sentences consistently in those same environmental conditions, then she apparently knows which sentence she usually holds true in them and, consequently, which sentence she on particular occasions holds true in them. We, accordingly, need to attribute to her the knowledge of which sentence she holds true in the environmental conditions she finds herself in.

Someone may object that the fact that such intentions and knowledge have to be attributed to her by an interpreter does not give us a good reason to conclude that she genuinely knows which environmental conditions she is in and which sentence she holds true. We want radical interpretation to be more than a collection of sounds that an interpreter manages to find meaningful, so we attribute things to the speaker that will prevent this from being the case. But this does not mean that the speaker genuinely speaks intentionally and that she knows which sentence she holds true in the environmental conditions that she is in.

This is not the case, however, since we will attribute to her whatever the relevant information about her suggests she is doing. To return to the measurement analogy, we treat our attributions of thirty centimetres or twelve inches as picking out a property of the ruler. Length is not something that it does not genuinely have but that we attribute to it because we have something to gain from thinking of a ruler as something with length. Whether we attribute to it the number thirty or the number

University of Cape Town

twelve depends on whether we employ the centimetres or inches measurement scheme, like the precise nature of the environmental conditions in which the speaker holds a sentence true depends on whether I am employing my language or on whether John is using his language as interpretive scheme to make sense of what she says. But, since there is ample evidence that she knows which sentence she holds true once she encounters the environmental conditions that she usually applies it in, knowledge of which sentence she holds true is something that she genuinely has, even though the precise nature of the environmental conditions that she knows she holds a sentence true in will depend on the specific interpretive scheme used to interpret her. She utters the same sentence in the same environmental conditions almost all of the time, she can relate that sentence to other environmental conditions and the sentences that she holds true in those, she can describe or point out the environmental conditions in which she does not hold it true, she can (and is required to) use her understanding of which sentences she holds true in which environmental conditions to attribute meaning and belief content to others, etc. There is as much evidence to suggest that she knows which sentence she holds true when encountering a set of environmental conditions as there is to claim that she knows what she means by her sentences. A speaker, thus, knows which sentence she holds true when encountering the environmental conditions that she holds it true in.

Lastly, the sentence held true asymmetry holds in both the radical and the practical cases. Both a radical interpreter and an interlocutor will be in some doubt about exactly which environmental conditions the speaker is going to respond to before she responds, and this is where the sentence held true asymmetry lies. The only difference between the radical and practical cases is that, in practice, we assume that our interlocutors know what our sentences mean, but since they may not know to which environmental conditions we are going to respond before we say or do something, they have to wait for us to tell them which sentences we hold true.

First, in the radical case, our interpreters do not know what our sentences mean, which gives them the additional task of examining the environmental conditions in which we utter the sentence that we hold true in order to know what we mean by it. But they have to wait until we tell them which sentence we hold true before they can interpret its meaning. Even if such radical interpreters already have a lot of

University of Cape Town

information about us and they know what we typically mean by which sentences, they still have to wait for us to utter the sentences that we hold true. They, thus, have to derive the sentences that we hold true from our behaviour while, for all the reasons given above, we can know which sentences we hold true even before we have actually spoken.

Second, in the non-radical case, the conventions employed by a speaker's interlocutors are likely to tell them what she means. Since meaning just is the truth conditions of a sentence, it may be tempting to object that such interlocutors have a good way of knowing which sentences she holds true in which environmental conditions, and that they, accordingly, have to be in a good position to know which sentences she holds true when they encounter the environmental conditions in which she holds them true.

The big mistake that such an objection will make is to assume that a speaker's interlocutors are in as good a position as herself to know when the conditions under which she holds a sentence true are present. It is true that their linguistic conventions give her interlocutors a good way of knowing which sentences she will hold true under which conditions, but it does not follow from this that they will know which conditions she is encountering before she utters the sentence that she holds true in them. The objection assumes that the truth conditions of a speaker's sentence are given by a sentence mined from their linguistic conventions that her interlocutor would have used in their current environment only, while the truth conditions of her sentence are actually given by a sentence that he would have used in their current environment in addition to sentences that describe numerous past environmental perceptions and experiences.[100]

Through their conventions her interlocutor knows, for example, that she is likely to hold true the sentence "Dogs are delightful pets" when encountering situations in which he would have used that same sentence: a woman playing with, or cuddling, a dog, for example, following from a background of positive experiences with dogs of his own and of others. In theory he can, thus, describe the type of environmental conditions that she will hold the sentence true in, together with numerous past perceptions and experiences of her environment that need to be in place for her

University of Cape Town

current environmental conditions to prompt her to hold it true. He knows that she may not hold true that sentence when encountering a woman hugging or playing with her dog alone, since he knows that the idea of dogs being delightful pets is informed by numerous other environmental conditions that she may have encountered in the past. If she owns a cat for whom she feels affection or if a dog has bitten her in the past, the environmental conditions that he thinks will prompt her to utter one sentence may be so informed by past environmental events that she utters something completely different. If he had been an omniscient radical interpreter, which no actual person can ever be, then he would have had this information and he would have known in the light of which past events to understand their current environment. But since our interlocutors do not have such information, they cannot understand which truth conditions we judge to obtain at any specific time, even though they know what type of environmental conditions and past experiences we typically hold which sentences true in.

In practice we realise that truth conditions go far beyond simple perceptions of a speaker's current environment; hence, my reluctance in this chapter to talk about sentences that are held true in specific environmental conditions. We know that, even though we can describe the type of complex truth conditions that a speaker's sentence should have in order to mean what our conventions tell us it does, we can never tell in advance whether such complex truth conditions obtain.

This is an impressive advantage of a sentence held-true asymmetry as an explanation for the attitude asymmetries. It can give a more nuanced account of the different categories of beliefs that we attribute. The attribution of straightforward perceptual beliefs such as "The sun is shining" clearly pose less of a problem for our interlocutors than beliefs that value events/people/experiences and those that require an assessment of environmental conditions that goes beyond simple perceptions. If the speaker's interlocutor encounters the environmental conditions in which he usually says, "This is a dog", he will effortlessly attribute to the speaker the belief that this is a dog. He is very likely to be right about what she believes, even without employing her behaviour or speech as evidence. If the speaker argues that she does not really believe this, we will usually think that she actually does believe it, whether she self-attributes it or not, or we will find some attentional or perceptual irregularity to

University of Cape Town

explain either why she actually does believe it, or why she does not believe it, even though she should. In cases of pure perceptual beliefs, even though we treat each other as if the attitude asymmetries are still present, it is intuitively plausible that first-person authority is fairly weak and that third-persons can make attributions without employing too much evidence.

Beliefs regarding the owner, the behaviour, the breed and the appeal of the dog are more difficult to attribute correctly without first observing what the speaker says and does. The speaker may say, "The dog is agitated", when he is doing what her interlocutor takes to be running back and forth in the corridor. He can then attribute this belief to her upon hearing which sentence she holds true. When he sees the dog running back and forth in the corridor, he usually holds true the sentence, "The dog is exhilarated because he anticipates a walk". Their conventions give them the knowledge of the meaning of both these sentences, but since the contexts in which they interpret these specific environmental conditions differ, she knows, while he may not know, which sentence she holds true.

Their shared linguistic conventions do, in other words, give him the knowledge that, by the sentence "The dog is agitated", she means that the dog is doing something to which he would have responded that "the dog is agitated" if he had certain background conditions in place. But whether she actually has such background conditions in place is something that there conventions cannot settle for him before she speaks. He, in other words, knows the type of truth conditions that her sentence "the dog is agitated" has to have, but since he lacks a lot of information about her past perceptions and experiences of her environment, he cannot know whether their current environmental conditions are conditions where those truth conditions obtain, unless she tells him.

Thus, the speaker knows, while her interlocutors may not know, which sentences she holds true before she speaks since she knows, while they may not know, how her background affects her interpretation of her current environmental conditions. And since she knows, while he may not know, how her background informs her understanding of her current environmental conditions, she knows, while he may not know, which truth conditions obtain and, accordingly, which sentence she holds true.

University of Cape Town

1.3. THE ADVANTAGES OF THE SENTENCE HELD TRUE ASYMMETRY

In the previous chapter, the strategy in 2.2 suggested itself as the best attempt at accommodating a meaning asymmetry in linguistic communities with their shared conventions. It suggests that the speaker knows what she means since she is speaking a language that is potentially interpretable to a radical interpreter. Her interlocutors are mostly able to know what she means because she is using the same conventions that they are. Those conventions, however, may at any time become incapable of explaining her speech, which will prompt her interlocutors to creatively interpret her. If she, for example, employs sentences that her interlocutor is not familiar with or she makes a surprising utterance like a mixed metaphor or a spoonerism, they may become wrong about what she means because they then, like a radical interpreter, have to interpret her.[101]

This strategy is preferable to the others in chapter 5 since it does not rely on problematic assumptions about interpretation and conventions like the strategies in sections 2.1 and 2.3 do, and it does not simply disregard the problem like the strategy in section 2.4 does. Its only shortcoming is that the evidence and transparency asymmetries that it manages to explain are not strong enough. It quite correctly accepts that linguistic conventions can often give us access to what people mean without using any interpretation, but then it does not have a way of preventing the further claim that, since our interlocutors have a route that often leads them to what we mean without the need for interpretation, they consequently have a route that often leads them to what we believe without the need for interpretation. Then there is simply no way of holding onto sufficiently robust evidence and transparency asymmetries since both these asymmetries require our interlocutors' general use of behavioural and verbal evidence.

This is why I propose that the attitude asymmetries are explained by a sentence held true asymmetry instead of by a meaning asymmetry. Proponents of the meaning asymmetry as an explanation for the attitude asymmetries will have to admit that a speaker's interlocutors will only use evidence to know what she means in cases where they misunderstand each other or where they communicate poorly. And then they lose

University of Cape Town

the general form of the evidence and transparency asymmetries. My strategy of admitting of such a meaning asymmetry in practice (for the reasons given in 2.2 in chapter 5), but of explaining the attitude asymmetries via a held-true asymmetry improves on this since the speaker's interlocutors always have to wait until she speaks or behaves to know to which environmental conditions she is responding. They will probably be right about which environmental conditions she identifies, but they have to wait for her response to judge whether they are right. Since they are likely to know what she means without having to make use of any evidence, once she has responded to her environment, her interlocutors will know which sentence she holds true and what she means by it. They will thus know what she believes. Their knowledge of their environment alone is, however, not sufficient to provide them with the knowledge of which sentences she holds true because they cannot know exactly which part of their shared environment she is going to respond to. Her actual response to her environment is the crucial step without which they cannot know anything about her beliefs. And since her response takes the form of speech and non-verbal behaviour, a held-true asymmetry can show why her interlocutors employ her speech and behaviour to attribute beliefs to her.

This strategy is also preferable to strategies like the one in section 2.4 in chapter 5 since it does not rely exclusively on the theoretical "radical case" and simply disregard what we as a matter of fact do in practice. It is, of course, vital that it is compatible with the radical case, which I can easily show it to be.

When a speaker speaks, her interlocutors use their shared linguistic conventions to know what she means. But whether they are right about what she means can only be established by an omniscient radical interpreter that is using an appropriate method of interpretation. Since such a radical interpreter will employ information of what sentence she holds true under which environmental conditions to know what she means, whatever environmental conditions she responds to, and whatever sentence she holds true in those environmental conditions, are issues that a radical interpreter has to settle as well. During communication between the speaker and her interlocutor, in other words, she knows which sentences she holds true since she knows what she means and thereby to which environmental conditions she is responding. But whether she is correct about what she means, and whether she is correct about which

University of Cape Town

environmental conditions she is responding to, are questions that have to be settled by a radical interpreter. This strategy can, thus, explain what it is that we do in practice. But it remains true to Davidson's criteria for something to qualify as being mental since it is consistent with the conceptual possibility of being interpretable to an omniscient radical interpreter. If an environmental condition, a sentence that is held true or a speaker's meaning is not in principle interpretable to an omniscient radical interpreter, then it just cannot be an environmental condition, a sentence that is held true or a speaker's meaning and we or our interlocutors are in error when self-attributing or attributing it.

The relationship between a case of radical interpretation and a case of communication that involves the use of shared linguistic conventions is thus that the former provides criteria for testing the latter. A system of radical interpretation is a test that we should use to confirm that something is genuinely mental. If, for example, there is something that we suspect might be a speaker's (authoritative) knowledge of one of her beliefs, we should put it through the test of radical interpretation and the held true asymmetry. If it can be thus explained, then it can be knowledge of her belief since it can qualify as something mental. If it cannot, then it cannot be knowledge of her belief since it cannot qualify as something mental. That is, if we know that, in principle, an omniscient radical interpreter would have been able to interpret the speaker as someone who is applying her sentences consistently, would have attributed to her a sentence that she holds true under environmental conditions that are accessible to him, would have attributed to her the knowledge of which sentence she holds true when encountering which environmental conditions, and so forth, then the event that we suspect to be knowledge of her belief really is that. If, on the other hand, we know that, in principle, an omniscient radical interpreter would not have been able to, for example, identify the environmental conditions that she is responding to, then the event that we suspect might be knowledge of her belief cannot be that, since something can qualify as the meaning of an utterance only if it can be attributed based on publicly available observations of behaviour and environment. Radical interpretation, thus, informs us about the nature of mental events by setting the limits on how they are to be attributed, and it gives us a way of testing whether our attributions (and self-attributions) based on linguistic conventions conform to the rigorous standards of objectivity that the mental requires.

University of Cape Town

A possible objection is that this strategy portrays what we do in practice as a pseudo linguistic and mentalistic performance. We think that we understand what others mean, we can have a guess at what they believe and we think that a speaker usually knows what she believes, but it is quite possible that we are wrong most of the time, or even all of the time. There is, in practice, no way of verifying the conclusions that we draw about the meaning and beliefs of anyone, not even of ourselves. Only an omniscient radical interpreter, which is not someone that can ever exist, can ever truly know anything about what we mean and believe.

This problem is not as grave as it may seem, however. Firstly, as argued in the previous chapter, in a situation where linguistic conventions are used, the radical interpreters of English speakers will construct very similar individual idiolects for them since they often speak in the same way. Linguistic conventions will, thus, mostly lead our interlocutors directly to conclusions that correspond with those our radical interpreters reach via interpretation. What we do in practice is, therefore, unlikely to lead us to conclusions that are wrong most or all of the time.

Secondly, we should keep in mind that Davidson's commitment to the holistic nature of meaning and belief entails that our strategies for making sense of others should rationalise, not particular isolated bits of their behaviour and speech, but should unearth and explain overall patterns of their behaviour and speech. Our linguistic conventions effectively constitute a whole language.[102] Our use of conventions, and our wish to obey such conventions as closely as possible, thereby allow us to detect the patterns in the speech of others. Moreover, we are willing temporarily to abandon our linguistic conventions when a speaker's speech can be rendered more coherent by creatively interpreting her. We also try to attribute a coherent set of beliefs evidenced by the numerous questions that we tend to ask when our use of linguistic conventions and our understanding of our shared environment suggest that the speaker holds incompatible beliefs. In practice we clearly attempt to make sense of overall patterns of a speaker's speech and behaviour, even though we do use different procedures than a radical interpreter would. And since the methods that we do use is more often than not able to make systematic sense of a large amount of the overall speech and behaviour of those with whom we communicate, we are clearly doing a lot right.

University of Cape Town

These two points, of course, do not guarantee that, in practice, we are always right about what a speaker means and believes. But the fact that our interlocutors' understanding of us will mostly coincide with our radical interpreters' interpretation of us (since we all do speak in roughly the same way), together with our strategies for trying to attribute coherent sets of meanings and beliefs, do suggest that our practices place us on approximately the same course as a radical interpreter's interpretations would. Our practices should, as a result, not be understood as having little to do with what people mean and believe.

Lastly, in chapter 2, I specified that one requirement is that we are looking for a unified account of the three attitude asymmetries. We do not want an account that can quite accidentally allow for all three in a haphazard way. We want a theory that allows for a systematic explanation of the three asymmetries. Traditional interpretationists cannot provide us with such a unified account. Even though they can show that all three asymmetries, in principle, can be accounted for by the meaning asymmetry, it cannot be extended to explain what actually happens between speakers and their interlocutors. They, thus, will have to rely on something different to explain the attitude asymmetries that we assume exist between speakers and those with whom they share linguistic conventions. A truly unified account is not one that differs depending on whether it is meant to explain the attitude asymmetries in principle or the attitude asymmetries as they actually occur between speakers and their interlocutors.

University of Cape Town

My strategy, on the contrary, utilises a sentence held true asymmetry to account for the attitude asymmetries. It can explain both what we in fact do in practice and what we would, in principle, have to do to confirm that our practices give an accurate picture of the mental. That is, since it makes use of a principle that can be extended to interactions within our linguistic communities, it can explain both the attitude asymmetries between us and our interlocutors and the criteria for something to qualify as an attitude asymmetry.

Traditional interpretationists' mistake is to place the explanatory power in the meaning asymmetry while, even in the radical case, it should be in the sentence held

true asymmetry. They are right that there is a meaning asymmetry between a speaker and her radical interpreter and, as seen in section 2.2 in chapter 5, that there is a meaning asymmetry between a speaker and those with whom she shares linguistic conventions. But since the meaning asymmetry in practice does not require that our interlocutors always use evidence to interpret us, it just cannot sustain all three the attitude asymmetries and should therefore not be used to explain them.

1.4. DAVIDSON CLARIFIED OR CORRECTED

Up to now I have deliberately steered clear of placing Davidson's own reconciliation strategy within one of these two interpretations of it. As mentioned in the schematic outline of the project at the end of chapter 3, he aimed his reconciliation at accounting for the attitude asymmetries in a situation of radical interpretation only, which is probably why he did not deem it necessary to spell out his approach in more detail. He could have prevented opening himself up to a meaning asymmetry interpretation if he had applied his reconciliation strategy more broadly to explain the type of situations that we want theories of self-knowledge to explain.

As mentioned in chapters 3 and 4, Davidson did admit to the existence of a sentence held-true asymmetry. He just did not think that it alone could explain the attitude asymmetries since, without the interpretation of meaning, the sentence held-true asymmetry is just as mysterious as the attitude asymmetries.[103] The first-person's knowledge of what she means can be built into the fact that she is speaking a language, which, according to both of us, is the only publicly accessible information that immediate and authoritative knowledge of meaning can be derived from. Davidson, like me, thus treated the meaning asymmetry as primary.

The difference between us is that nothing in Davidson's writings indicates that he deemed a sentence held-true asymmetry to be important, while I think that the fact that the sentence held-true asymmetry has to be derived from the meaning asymmetry does not render it unimportant.

It is not clear from Davidson's writings whether he understood that his reasons for defending a speaker's knowledge of what she means could also serve to defend a

University of Cape Town

speaker's knowledge of which sentences she held true. It is, accordingly, difficult to know whether my strategy clarifies or improves on Davidson's. My critics might claim that, for all the reasons I have given, Davidson meant the sentence held-true asymmetry all along. In his 1984 paper, "first-person authority", he certainly minimized the importance of a sentence held-true asymmetry for the reasons given above, and in his 1993 responses to critics who challenged the ability of the meaning asymmetry to explain the attitude asymmetries, he never mentioned the possibility of a sentence held-true asymmetry that is necessarily entailed by it. But in his 1991 paper, "the three varieties of knowledge", he argued, as I did in chapter 3 and above, that a speaker of a language can be an interpreter of others only if she knows the content of her own language and beliefs. This suggests that he was aware of the idea of a speaker's knowledge of which sentences she holds true in which environmental conditions. But he advanced this argument in the epistemological context of arguing that no one form of knowledge can be reduced to any other, as opposed to in an argument for self-knowledge or first-person authority. Whether he was aware that this reasoning could enhance his argument for first-person authority by implying a sentence held-true asymmetry remains a mystery.

Thus, since there is nothing in his writings that suggests that he did mean the sentence held-true asymmetry all along, and since there is nothing that indicates that he did not, I cannot take a definite stand on whether my account clarifies what he was trying to say right from the start, or whether it develops what he did say.

University of Cape Town

2. OBJECTIONS TO A SENTENCE HELD-TRUE ASYMMETRY IN THE RADICAL CASE

I have just offered an alternative way of explaining all three the attitude asymmetries in both the radical and non-radical cases based on the notion of a sentence held-true asymmetry. In order to both clarify and defend this claim, I will consider a number of possible objections to it. I will divide them into three groups. In the first section I will consider objections to the claim that the sentence held-true asymmetry can explain the asymmetries in the radical case. In the next I will consider objections that focus on the non-radical case. In the third section, I will consider objections to it that arise from situations in which the attitude asymmetries are not present.

2.1. THE SENTENCE HELD-TRUE ASYMMETRY AND THE DANGER OF CIRCULARITY

Some theorists may want to argue that the above account is circular since each stage requires another stage in a circle to explain it. There is no way into the circle to justify non-circularly one of its stages. According to this objection, my approach proceeds as follows:

The speaker knows what she believes because, when encountering a set of environmental conditions, she knows which sentence she typically holds true in it. She knows which sentence she holds true in it because she knows which environmental conditions she typically holds which sentences true in. She knows which environmental conditions she typically holds which sentences true in because she knows what she means by the sentences that she holds true. She knows what she means by the sentences that she holds true because she knows which sentences she typically holds true in which environmental conditions. She knows which sentences she typically holds true in which environmental conditions because, when encountering a specific set of environmental conditions, she knows which sentence she typically holds true in it.

University of Cape Town

Firstly, this objection separates my argument into three stages: the speaker knowing the meaning of the sentences that she holds true, the speaker knowing the environmental conditions in which she typically holds sentences true in and the identification of which sentence she holds true when she encounters a specific set of environmental conditions. I, and I am sure Davidson will concur, propose that these are not three separate stages that are meant to explain one another, but three different ways of expressing the same point. As explained earlier, meaning just is the truth conditions of a sentence. Its truth conditions are given by the sentence that she holds true in the presence of certain environmental conditions. If she applies her sentences consistently and the sentence on the right-hand-side of her biconditional is, accordingly, the sentence that she typically holds true in the same environmental conditions than the sentence on the left, then she knows which environmental conditions she typically holds which sentences true in and, when encountering a

specific set of environmental conditions, she knows which sentence she holds true in it. Again, if she knows the meaning of one of her sentences, then, by definition, she knows its truth conditions. If she knows its truth conditions, then, by definition, she knows which environmental conditions she usually holds it true in. The three points are not meant to explain each other; they are one and the same thing.

Furthermore, I do not leave the account without some outside anchor since we can only say that a speaker speaks a language if she is interpretable. Whether she is right about what she means by the sentences that she holds true is, as in Davidson's original system, still up to her interpreter. The way into the circle is thus still via the interpretation of meaning, but since everyone agrees that the speaker does not have to interpret her own speech to know what she means, the circle from her perspective does not need such an opening. She knows what she means, knows which sentences she typically holds true in which environmental conditions and knows which sentence she holds true when encountering a specific set of environmental conditions only if she is interpretable. It is exactly the same type of transcendental argument from interpretability that Davidson employed to explain his reconciliation strategy and that I defended in chapter 4.

2.2. FIRST-PERSON AUTHORITY AND PUBLICLY ACCESSIBLE INFORMATION

Here is another objection to my strategy. In the radical case a sentence held-true asymmetry, even if we agree that it can explain the evidence and transparency asymmetries, just cannot be large enough to justify the authority asymmetry. Information of which sentences a speaker holds true is non-intentional information that is, by definition, publicly observable and accessible to all. There is no justification for claiming that one person's access to publicly observable information is superior to that of another. The fact that it is publicly accessible makes everybody access it with the same likelihood of being right. A speaker's omniscient radical interpreter is just as likely to know which sentences she holds true as she is because, if he is an omniscient radical interpreter, he knows all the sentences that she holds true, both those that she has actually uttered and those that she would have uttered had she been appropriately prompted under favourable circumstances. The information that

University of Cape Town

her consistent use of her language gives her about which sentences she holds true cannot give her any knowledge beyond what such omniscient interpreters can learn from the information of which sentences she has actually uttered and of which sentences she would have uttered if prompted appropriately. Even if we agree that the speaker's knowledge does not actually stem from her own behaviour, if she cannot know anything about which sentences she holds true that an omniscient interpreter cannot know, then we cannot say that she knows, while her omniscient radical interpreters may not know, which sentences she holds true.

This objection simply re-states the original problem regarding the authority that a speaker has over her beliefs and over the meaning of her sentences. There is no element of the meaning of a speaker's sentences that cannot be known by an omniscient radical interpreter and, consequently, there is no element of the content of a speaker's beliefs that cannot be known by such an interpreter. Davidson and I, nonetheless, claim that a speaker is in a position of authority with regard to the meaning of her sentences and the contents of her beliefs since an interpreter always has to continue interpreting her to know what she means. And the accuracy of past interpretations of what she means and believes cannot guarantee the accuracy of future interpretations of what she means and believes.

The same holds for the sentences that she holds true. If he is an omniscient radical interpreter at time T1, he will know all the sentences that she holds true at time T1. But even though he is an omniscient radical interpreter at time T1, he has to continue observing her speech, non-verbal behaviour and environment to enable him to remain an omniscient radical interpreter into time T2. And if he has to continue to observe her and prompt her appropriately under favourable circumstances to know which sentences she holds true, he may fail to notice some sentences that she does hold true. And then she knows, while he may not know, which sentences she holds true, not because the sentences that she holds true are facts that only she has access to, but because he has, for some reason or another, failed to notice facts that he could have accessed.

The objection conceives of the notion of an omniscient radical interpreter as someone who knows all the sentences that a speaker has ever uttered and would have uttered

University of Cape Town

and will still utter in the whole of her life if prompted appropriately under ideal circumstances. But this removes such an interpreter from the context of radical interpretation. He has to actually prompt her and observe her to obtain the information that allows him to become an omniscient interpreter. And during this process of prompting and observing he may prompt her in a manner that is inappropriate to elicit the right type of information, prompt her in circumstances that are not favourable for acquiring honest or correct responses or he may fail to notice some of the sentences that she holds true. And then she knows, while he does not know, which sentences she holds true. We can make no sense of the idea of an interpreter that is born with, that the speaker psychically supplies with, or that otherwise acquires information of all the sentences that a speaker holds true. Such mechanisms will make it impossible for him to interpret such information because, in order to make sense of what she says, he needs to know what she holds true under environmental conditions that he needs to perceive for himself. Knowing what she holds true under environmental conditions that are apparent to him is the only way in which he can map her language onto his own. The concept of an omniscient radical interpreter, thus, poses no threat to the sentence held-true asymmetry because, so long as we understand that an omniscient radical interpreter has to continue prompting and observing a speaker to know which sentences she holds true, the possibility remains that she knows, while he does not know, which sentences she holds true when encountering the environmental conditions in which she holds them true.

University of Cape Town

3. OBJECTIONS TO A SENTENCE HELD-TRUE ASYMMETRY IN THE NON-RADICAL CASE

3.1. THE SPEAKER'S KNOWLEDGE OF UNUTTERED SENTENCES

One can put forward the following objection to a sentence held-true asymmetry in both the radical and non-radical cases. A sentence held-true asymmetry unwarrantedly relies on the speaker's continuous use of a language to give her the knowledge of which sentences she holds true. It is acceptable to use her past use of a language to claim that she knows that she holds true sentences that she has actually uttered. But, since we cannot assume that she will continue to speak a language, I cannot claim that she knows that she holds true sentences that she has never uttered. If the speaker has,

for example, uttered the sentence "It is raining" in the presence of something her interpreters or interlocutors judged to be rain, then it is reasonable to claim that, when next she encounters rain, she knows that she holds true the sentence "It is raining". If she has never uttered the sentence, "Wolverine has metal claws that shoot out from the back of his hands" while watching what her interpreters or interlocutors take to be an X-men movie, I cannot claim that she knows that she holds this sentence true when encountering the X-men movie, since she has never been interpreted as holding true this sentence consistently in these type of environmental conditions. If she has never uttered it, then she has never been judged to use the sentence consistently, and then I cannot claim that her consistent sentence application gives her the knowledge of which sentence she holds true when encountering that specific set of environmental conditions.

This objection assumes that every single of the speaker's sentences has to be actually interpreted for us to know that she is still speaking a language and that she is thereby still applying most of her sentences consistently. It, thus, assumes that the meanings of different sentences are attributed one by one on their use since each sentence has a meaning that is isolated from whatever other sentences in a language mean. This objection, in other words, overlooks the holistic character of meaning and belief content. Part of what Davidson's theory of meaning tries to do is to spell out the meanings of sentences holistically through the relationships among sentences within the structure of the language as a whole. That is why the only way of getting to the meanings of a speaker's sentences is to map it onto another language so that their meanings can be holistically established by relating them to sentences, and the relationships between those sentences, in that other language. So, even if the speaker has never uttered the sentence, "Wolverine has metal claws that shoot out from the back of his hands" while watching what her interpreters or interlocutors take to be an X-men movie, she is likely to have uttered numerous sentences about metal claws, shooting and backs of hands quite consistently in the presence of what such interpreters or interlocutors took to be metal claws, shootings and backs of hands. There is, thus, more than a sufficient amount of information to claim that, even if she has never been judged to apply a specific sentence consistently, due to its links with other sentences in her language, and due to her knowledge of such related sentences in her language, she knows which sentence she holds true when encountering

University of Cape Town

environmental conditions that are similar to the ones she usually holds the related sentences true in.

We do not need an interpreter or interlocutor to judge that a speaker is applying every one of her sentences consistently before we can rightly label her a speaker of a language. If she is mostly interpretable, even if there are occasional lapses in interpretability, she is speaking a language and she still generally knows what she means and which sentences she holds true. If she remains generally interpretable when she does speak, we can thus say that she knows which sentences she holds true in which environmental conditions, even in cases where she has not actually uttered them.

3.2. THE ACTIONS OF SPEAKERS AND THEIR INTERLOCUTORS

The next possible complaint is that my account still relies on the idea of a radical interpreter. In section 2.4 in chapter 5, I objected to the notion of an interlocutor as a potential radical interpreter by arguing that such a strategy still does not explain what exactly the role of interlocutors really is. I argued that treating them as potential radical interpreters cannot explain the attitude asymmetries between them and those with whom they communicate since it does not explain what they actually do in practice. Now, however, I claim that it is acceptable to use a similar strategy by defining the role of the speaker in a way that does not capture what she actually does in practice. This account needs a radical interpreter to ensure that the speaker continues to speak a language, since we need the fact that she is speaking a language to base her knowledge of which sentences she holds true on. An omniscient radical interpreter, as we have seen, has access both to what the speaker actually says, and to what the speaker will say if he prompts her appropriately under favourable circumstances. But in practice the speaker is not prompted by an omniscient radical interpreter and she does not utter sentences that she would have uttered if appropriately prompted. And without uttering sentences that she will have uttered when appropriately prompted by an omniscient radical interpreter, we cannot say that she continues to be interpretable (and speak a language). We can only say that she continues to be interpretable if judged so by such an omniscient radical interpreter. If we cannot judge that she is speaking a language and applying her sentences

University of Cape Town

consistently from whatever it is that she as a matter of fact does in practice, then we do not have a way of claiming that the consistent use of her sentences gives her the knowledge of which sentence she holds true when encountering the situation in which she holds it true.

As mentioned earlier, my account does not require the interpretation of a radical interpreter, and it does not need the speaker to do anything that she actually does not do. In practice, the speaker's interlocutors are sufficiently well qualified to judge whether she is still interpretable or not.

Firstly, as argued both in chapter 6 and above, if the speaker and her interlocutors are both members of the English speech community, they both recognise themselves as such since they speak very similarly and they use conventions based on their already similar idiolects to further simplify communication, then their radical interpreters will construct very similar interpretive schemes for them; interpretive schemes that will lead to approximately the same conclusions about meaning than their conventions do.

Secondly, even though we mostly understand others without examining their speech and behaviour in the context of their environment, the interpretive constraints that we bear in mind when making sense of those with whom we communicate are similar to those that a radical interpreter will have to obey. The conventions that her interlocutors use are, like the interpretive scheme of a radical interpreter, aimed at explaining patterns in her speech, as opposed to making sense of discrete bits of speech and behaviour in isolation. Moreover, on particular occasions where their conventions cannot make sense of what she says, her interlocutor still seems to follow the right type of constraints to make accurate sense of her. He will rely on some of their other shared conventions (that are likely to be accurate), likely deviations from such conventions, their shared environment and her behaviour, all employed with the dual aim of attributing to her perceptual beliefs that correspond with theirs and a coherent set of meanings and belief contents.

Thirdly, our interlocutors are in a good position to notice if we stop applying our sentences consistently and even then, their use of interpretation can still lead them to an accurate judgment of our interpretability. The conventions that our interlocutors

University of Cape Town

use to make sense of us make holistic sense of us. The meaning of one sentence with, for example, the phrase "It is raining" in it is systematically connected with the meaning of sentences like "It is raining hard" or "It is not raining". "It is raining hard", in turn, is systematically connected with sentences like "The powerful blow knocked the boxer down" while "It is not raining" systematically resembles "John is not a teacher" which is linked with "teaching is a noble profession" and so forth. All our sentences rely for their meaning on numerous other sentences, and our shared conventions prompt us to speak and understand in a way that respects such connections. If a speaker stops applying her sentences consistently, it is consequently likely to become apparent to her interlocutors almost immediately.

The random application of just a few of her sentences will render their conventions of little use in making sense of her since such conventions are geared towards attributing such interconnected meanings. This will compel them to abandon their shared conventions in the search for a system that can assign meaning to her speech that is systematically connected. And since they are likely to fall back on what Davidson calls radical interpretation, they will be reasonably likely to stumble onto an interpretive scheme that can make sense of her, in case she actually is applying her sentences consistently. If their project of radical interpretation still fails to make sense of her, then there is a reasonably good chance that she actually has become uninterpretable. Only an omniscient radical interpreter can, of course, make an accurate judgment of whether she really is interpretable or not, but since the conventions that her interlocutors use are likely to detect it if she does become uninterpretable, and since they can then resort to the best interpretive scheme that they can find, they are likely to be fairly good judges of her interpretability.

My explanation, thus, does not need the speaker to do anything other than what she in fact does in practice. What she does, and what her interlocutors do, will, much more often than not, tell us whether she is applying her sentences consistently or not (and thereby whether she is speaking a language or not). This account, thus, can explain why the speaker knows, while her interlocutors may not know, which sentences she holds true without having to assume that they act in a way that they actually do not.

University of Cape Town

4. OBJECTIONS TO THE SENTENCE HELD-TRUE ASYMMETRY IN THE ABSENCE OF THE ATTITUDE ASYMMETRIES

The following objections involve situations in which the speaker is uncertain about what she believes or where she has to employ her own behaviour and speech as evidence to know what she believes and self-attribute a belief in the service of an explanation of her own behaviour. According to the sentence held-true asymmetry, it will have to be because she is uncertain about which conditions she is in and, accordingly, uncertain about which sentence she holds true. When a speaker's interlocutor, thus, is right about what she believes while she is wrong or uncertain, it will have to be because he correctly identifies the current conditions as conditions that she holds it true in while she is wrong or uncertain. To illustrate how the sentence held-true asymmetry will explain situations in which we realise that the attitude asymmetries are absent, let us return to the same example employed in chapters 2 and 4. A speaker says that she believes that her husband is not a murderer while someone with whom she communicates maintains that she actually believes that her husband is a murderer, since a lot of her behaviour indicates that she fears him. As in chapter 4, I shall give a schematic outline of the authority asymmetry only, since it will be evident how it explains the other two.

University of Cape Town

The authority asymmetry

1. Both the speaker and her interlocutor know the meaning (or truth conditions) of both her sentences "My husband is a murderer" and "My husband is not a murderer".
2. The speaker utters the sentence "My husband is not a murderer" under conditions that would usually have prompted her to say, "My husband is a murderer".
3. Since the speaker is applying the sentence inconsistently, she has misidentified the conditions under which she is uttering it and has thus lost her position of authority over her language and belief contents.
4. Both the speaker and her interlocutor have to employ her speech and behaviour as evidence to know what environmental conditions she finds herself in.

5. Her interlocutor is, thus, just as likely as herself to know which sentence she holds true because he is just as likely as her to know which conditions she finds herself in.

Here meaning is the objective element that fixes whether she is uttering the right or wrong sentence as the one that she holds true in her current environmental conditions. This is true in both the radical and non-radical cases. In the radical case, her radical interpreter has access to his perception of their environment, her speech and behaviour and whatever he has already attributed to her. He knows that she usually holds true the sentence "My husband is not a murderer" in environmental conditions that he holds true the sentence "My 'spouse' is not a murderer". Now he realises that she is uttering the sentence "My husband is not a murderer" in environmental conditions that he would have held true the sentence "My 'spouse' is a murderer". He will attempt to re-interpret her and assign a new meaning to that sentence, but that will cause too many other parts of her speech not to be interpretable, since it will then become impossible to map many of her other sentences onto his language. He will, thus, use the coherence constraint and simply assume that, on this specific occasion, she is getting the meaning of the sentence that she is uttering wrong. In other words, she is uttering it as a sentence that she holds true in conditions that she does not usually hold it true in. And the only reasonable explanation for this is that she has misidentified the conditions in which she is speaking. In the non-radical case, her interlocutor, via their shared linguistic conventions, knows which sentences she typically holds true under which conditions, namely, the same conditions that he does hold, or would have held, them true in. If she utters the sentence "My husband is not a murderer" under conditions that he is certain are different from the ones she should be applying them in, he is unlikely to try to re-interpret her language, since he assumes that she is using the same linguistic conventions that he is when she speaks. He will immediately assume that she is mistaken about which sentence she holds true, and the only reasonable explanation for this is that she has misidentified the environmental conditions in which she is speaking.

University of Cape Town

4.1. A MEANING OR A SENTENCE HELD-TRUE ASYMMETRY

Now traditional interpretationists may argue that, since I cannot show how the sentence held-true asymmetry deals with situations where we realise that an attitude asymmetry is not present without using meaning as the mechanism whereby the sentence held-true asymmetry explains such situations, then I am simply returning to the meaning asymmetry to explain such situations. Then my strategy is not a unified explanation of the attitude asymmetries since it uses the sentence held-true asymmetry to explain the general forms of the attitude asymmetries and the meaning asymmetry to explain situations in which we realise that they are absent.

I, however, am not doing this. My claim is that the speaker's misidentification of the conditions in which she applies a sentence introduces an inconsistency in her language use and makes her get the meaning of her sentence in the current situation wrong. But the reason for this, as it has been throughout this chapter, is that she is uncertain about the conditions in her environment or the background conditions in the context of which she understands her current environmental conditions. I am, thus, still focusing on how we understand our environment and which sentence we, accordingly, hold true. If, as I have done, one starts with the speaker's identification of (or her interlocutors' inability to identify) the conditions that are currently in place, and move to the sentence that she holds true in those conditions, then one is in a position to accommodate what it is that we do in practice. But it, quite correctly, continues to rely firmly on Davidson's approach to meaning, firstly, since meaning is the only way of talking about what we hold true in our environment, secondly, since meaning is the only objective way of fixing what we do hold true and, thirdly, since the attribution of meaning is the only way of establishing that the speaker is speaking a language and thereby of guarantying her knowledge of what she means, her knowledge of which sentences she holds true under which conditions and her knowledge of which conditions she is currently encountering. The presence of meaning is one of the strengths of my account, not one of its weaknesses, since it shows how a sentence held-true asymmetry can be established objectively.

University of Cape Town

4.2. THE SENTENCE HELD-TRUE ASYMMETRY AND THE ABSENCE OF THE AUTHORITY ASYMMETRY

My critics may object that a sentence held-true asymmetry cannot give a satisfactory account of situations where we assume that the authority asymmetry is absent. A sentence held-true asymmetry may explain the absence of the evidence and transparency asymmetries, since it can explain why a speaker may have to use her speech and behaviour as evidence to self-attribute a belief. But it cannot explain the absence of the authority asymmetry. If the account holds that the speaker is in a better position to identify the truth conditions that hold at any specific time since they are strongly informed by past perceptions, then it implies that it is highly improbable that her interlocutors are right, while she is wrong, about which conditions she is responding to.

My critics may even ask how the speaker can ever be wrong about the conditions that she is speaking in. Since her interlocutors, given the same background conditions as hers, would have used the same sentence as the one she now uses, it appears as if whatever she utters as a sentence that she holds true is right. If, for example, her interlocutor trusted and loved his spouse, if he was committed to his spouse, if he dreaded the end of his marriage, if he was ashamed of being married to a murderer, and so forth, then he would have uttered the same sentence as she utters as the one he held true. There is, thus, no reason for thinking that the speaker has uttered the wrong sentence as the one that she holds true, considering that the background conditions in the context of which she understands her current environment effectively validate whatever she says as something that she holds true.

Firstly, from her interlocutor's perspective, understanding her reasons for holding true a sentence is different from actually endorsing it as something that he would have held true, given the same background conditions. He will understand that he may have held true the same sentence if he had been in exactly the same position as her, but he will know that, given the presence of some powerful emotions and the obvious nature of their current environment, he would have been wrong to do so. He, thus, will conclude that she is, likewise, wrong about what she holds true as a result of the way in which some powerful emotions have caused her to misidentify her current

University of Cape Town

environmental conditions (and thereby misidentify her current situation as one that fulfils the truth conditions that this sentence usually has). Afterwards, the speaker is also likely to acknowledge that she did not identify the nature of her current environmental conditions in the same way she usually would have identified them in the absence of such powerful emotions, (and that she thereby misidentified her current conditions as the truth conditions of that sentence).

In situations like the Tex-bar case in chapter 2, where the speaker does not realise that she believes that a Tex-bar tastes better than other chocolate bars, there are obviously no strong emotions that interfere with her ability to identify the conditions in which she speaks. But, as suggested in chapter 2, this is not the type of case in which her interlocutor will confidently claim that he is definitely right about what she believes. He will suggest to her that she might believe that Tex-bars taste better than other chocolate bars, since the last fifteen chocolate bars that she bought included ten Tex-bars. But then he will wait for her to tell him whether that behaviour genuinely indicates that she believes what he thinks she does. In this situation she is uttering the sentence, "No one chocolate bar tastes better than the others" in a situation where he, firstly, thinks that most of her other behaviour indicates that she holds true the sentence, "Tex-bars taste better than other chocolate bars" and, secondly, thinks that such behaviour in his own case would have indicated a preference for Tex-bars. They will consequently enter into a discussion about her past and present experiences of Tex-bar eating and buying since she clearly seems unsure about what she holds true. But even in this discussion he will still regard her as an authority on what she holds true, since he regards her as an authority on what exactly conditions she is speaking in.

It, lastly, should be noted that my account does not imply that her interlocutors are mostly wrong about the conditions to which she is responding. It only implies that they may be wrong since they do not have all the information about her past perceptions that the speaker has. They, in other words, can use their mutual environment, their knowledge about her history and their linguistic conventions to guess which sentence she holds true, but they have to wait until she speaks and behaves to know what she believes, since they have to wait until she speaks and

University of Cape Town

behaves to know to which aspect of their environment she is responding and how her past perceptions have influenced the nature of the conditions that she is identifying.

5. CONCLUSION

The traditional interpretationists, who argue that Davidson proposed a meaning asymmetry alone to account for the attitude asymmetries, wind up with an implausible explanation of those asymmetries because the meaning asymmetry cannot be translated into practice. They, however, completely overlook the fact that their reasons for defending a speaker's knowledge of what she means can also serve to defend a speaker's knowledge of which sentences she holds true. In sections 1.1 and 1.2 I showed how the fact that all language speakers know what they mean, which is what Davidson uses to defend his original strategy, can sustain a sentence held true asymmetry. A sentence held-true asymmetry can explain the attitude asymmetries in both the radical and the non-radical cases, and nun of the objections considered in sections 2, 3 and 4 can cast doubt on its ability to do so.

In sections 1.2 and 1.3, it also became apparent how Davidson's system of radical interpretation relates to what we in fact do when communicating with those in our linguistic communities. It effectively provides theoretical criteria for establishing the nature of the mental by ensuring that meaning and belief contents are the type of things that are socially constituted. It provides criteria for testing the accuracy of the attributions that we make within our linguistic communities by ensuring that nothing that cannot potentially be attributed based on publicly accessible evidence can qualify as something genuinely mental.

We, therefore, now know why we treat a speaker as if she is usually right about what she believes, as if she does not usually have to use her own behaviour and speech as evidence to know what she believes and why she does not usually self-attribute beliefs in the service of an explanation of her behaviour and speech. The relationship between the radical and the non-radical cases also show us when we are, and when we are not, correct when doing so.

University of Cape Town

CHAPTER 7
CONCLUSION

In this chapter, I intend to evaluate the strength of my conclusion by relating it, firstly, back to the methodology in chapter 2, secondly, to the generic interpretationist account sketched in chapter 1 and, thirdly, to other states of mind.

In chapter 2, I argued that the best approach for describing the asymmetries that theories of beliefs should aim to account for, was to describe some of our common everyday epistemic and conversational practices. My reluctance to derive them from theories of self-knowledge stemmed from my concern that it would be difficult to convince a theorist of the desirability of accounting for asymmetries derived from theories of self-knowledge that she did not subscribe to. I, as a result, motivated that theories of beliefs, and in this case particularly interpretationism, were more desirable if they could account for our different treatments of, as opposed to some actual differences between, the first- and third-person perspectives of beliefs.

My critics may, however, think that this approach is too weak since it may allow theories that cannot truly account for the first-person's perspective of beliefs to qualify as desirable, simply because they can accommodate our practices. Such critics may remain unconvinced by my methodology in chapter 2, and may continue to claim that the philosophically interesting question is not whether our theories can accommodate our practices, but whether our theories can accommodate what genuinely is the case. They may object that being able to accommodate our practices is a minimum requirement, but that theories that can accommodate what genuinely is the case are more desirable than those that can only accommodate our practices. In chapter 4, I quoted Sarah Sawyer, who expressed the concern as follows: "According to Davidson, if we do not assume a subject knows her thoughts, then we cannot begin the process of radical interpretation. This explains why the presumption is needed, but it is not clear that this in itself provides a justification for it."[104]

I do not want to take issue with how we can possibly know what is genuinely the case, which is precisely the type of question that I was attempting to avoid when I

University of Cape Town

proposed that we should derive the asymmetries from our practices. But now that we have reached the conclusion of my investigation into the role of the first-person perspective of beliefs in Donald Davidson's version of interpretationism, I want to suggest that it shows that radical interpretationism can accommodate a stronger version of the asymmetry thesis than the one described in chapter 2.

Sawyer's objection, that is quite common, stems from understanding the transcendental argument from interpretability in isolation from, firstly, Davidson's other interpretationist commitments and, secondly, from the other two arguments for Davidson's reconciliation strategy.

Firstly, the other interpretationist commitments. Interpretationism, as we have seen, maintains that beliefs are properties that people have, in the same way as length is a property that a ruler has. But a belief that P, since its content has to be constituted socially, cannot be a belief that P if it is not so describable by someone with whom we communicate. Similarly, a length of thirty centimetres, since its numerical value needs to be determined in relation to other numerical values, cannot be a length of thirty centimetres without being so describable by someone who is making use of a specific type of measurement scheme. A belief is not something that a person does not really have, and that we only attribute to her in order to make sense of her. It is something that she really has, but its precise description depends on which interpretive scheme is interpreting her; like length is something that the ruler has, but its precise description depends on which measurement scheme is used to measure it. This can be contrasted with instrumentalism, according to which we credit our interlocutors with whatever will assist us in making sense of them. Beliefs, on this theory, are not inner states or, in fact, properties that our interlocutors in any sense have. They are useful tools that we use to keep track of what they say and do, and their reality stretches no further than that.

For the same reasons as above, Davidson can maintain that a speaker's knowledge of her beliefs is not just something that we attribute to her in order to render her interpretable, but is, in fact, something that she really has (only if it is describable as such by an omniscient radical interpreter, of course). He will have to go about this in the following way: the speaker, since everything she can and does say indicates that

University of Cape Town

she knows what she means, does know what she means (see chapter 4). The speaker, since everything she can and does say indicates that she knows which sentences she holds true, does know which sentences she holds true (see chapter 6). The speaker, since everything she can and does say indicates that she knows what she means without using her behaviour as evidence, does know what she means without using her behaviour as evidence (see chapter 4). The speaker, since everything she can and does say indicates that she knows which sentences she holds true without using her behaviour as evidence, does know which sentences she holds true without using her behaviour as evidence (see chapter 6). It is not just something that we have to assume in order to be able to interpret her, like beliefs are not just things that we have to presume she has in order to interpret her. It is something that she really has but that needs to be interpretable, like beliefs are things that she really has but that need to be interpretable. In other words, all the arguments in this thesis suggest, not only that we have to treat her as if she has knowledge of her meaning and beliefs in order to successfully interpret her, but that we have to treat her as if she has knowledge of her meaning and beliefs because all the evidence that we have suggests that she has such knowledge.

Secondly, the contribution that the other two arguments in Davidson's reconciliation strategy make. The argument from disquotation actually allows the speaker to state the truth conditions, not only because she is interpretable as such, but because she is an interpreter of others. If she did not know which sentences she held true in which environmental conditions, then she would not have been able to map the languages of others onto her own. The facts that she can state the truth conditions of her sentences and that she has to know her own language in order to interpret others (together with interpretability, of course), show how interpretationism can account easily for the view that the speaker as a matter of fact knows what she means, and that she knows this without evidence. If we take seriously the connection between meaning and belief, and we take seriously the fact that authoritative and immediate knowledge of meaning necessarily implies authoritative and immediate knowledge of which sentences we hold true, then Davidson's radical interpretationism can explain the view that a speaker, as a matter of fact, has authoritative and immediate knowledge of her beliefs that are transparent to the world. Davidson's interpretationism can, as a result, satisfy both philosophers that subscribe to my methodology in chapter 2, and

University of Cape Town

those that insist that authority, immediacy and transparency are more than just practices.

The second question I want to address is the significance of my conclusion for other interpretationist accounts. In chapter 1, I described some claims that all interpretationists endorse. Beliefs qua beliefs are not inner states. I.e., claims about what a subject believes are not claims about her internal physical (or any other) states. They are properties or statuses that individuals have in virtue of being interpretable as such by others. A belief that P is a property that we have to be able to attribute to a person, like a length of thirty centimetres or twelve inches are properties that we have to be able to attribute to a ruler. Beliefs just are properties that rationalise other propositional attitudes, behaviour and speech. There is no element of beliefs that fall outside a project of interpretation that rationalises our attitudes, behaviour and speech. For something to qualify as a belief, it, thus, has to be attributable by a third-person in a context of this type of rationalisation or interpretation. What a fully informed interpreter, who employs a proper system of interpretation, cannot find out about a belief, is, thus, not a belief at all.

Since all interpretationist models define beliefs as rationalising our behaviour and speech, they can all potentially account for a meaning asymmetry via a transcendental argument from interpretability. The speaker is interpretable as knowing what she means because she is applying her sentences consistently. But by now we know that the meaning asymmetry alone is not capable of accounting for the asymmetry thesis. The three primary claims that enable Davidson's radical interpretation to account for it is, firstly, his insistence on starting with a sentence that the speaker holds true, secondly, his definition of the meaning of a sentence as the truth conditions of that sentence and, thirdly, his view that all speakers are interpreters of others. If meaning just is the truth conditions of a sentence, then knowing the meaning of our sentences implies that we know which sentence we hold true when which truth conditions obtain. Further, if we have to know the truth conditions of our sentences before we can map the languages of others onto our own, then our knowledge of which sentence we hold true before we utter it is secured. My conclusion, thus, can be generalised only to those interpretationist systems that subscribe to these three claims.

University of Cape Town

The questions related to the first-person perspective that remains for Davidson's radical interpretationism is whether my conclusion can be generalised to propositional attitudes other than beliefs and, ultimately, to sensations.

Philosophers usually treat beliefs as the most important of the propositional attitudes, since several others require or rely on them. We, for example, cannot hope for sunny weather without being able to believe that it is sunny, since the belief that it is sunny is the success condition of the hope. The same holds for desires, intentions, fears, and so forth. But the central place that Davidson gives truth in his system, by starting with a sentence that the speaker holds true, casts doubt on its ability to be extended to these other attitudes. If one, however, believes that a desire for P can be reduced to a collection of beliefs like the belief that not-P, the belief that P is good to have, the belief that being without P is disappointing, and so forth, then his system is safe from these doubts. The extension of my conclusion to other propositional attitudes, thus, depends upon what one takes such other attitudes to be.

Sensations pose a more substantial problem. If our immediate and authoritative knowledge of our beliefs rely on our immediate and authoritative knowledge of which sentences we hold true, then so should our knowledge of our sensations. Since sensations do not seem to involve the holding true of meaningful sentences as much as propositional attitudes do, Davidson himself specified that he wished his account to be extended only to propositional attitudes. This is, thus, not a concern only for my conclusion, but for interpretationism as a whole. If we want a unified account of the first-person's perspective on everything that is mental, interpretationists will have to show how their account can be extended to sensations. If they think that sensations are sufficiently different from propositional attitudes to warrant a different explanation for them, then interpretationists will still have to show what type of different explanation for them interpretationism can incorporate.

I, therefore, want to conclude that Davidson's system of radical interpretationism is capable of accounting for the differences between the first- and third-person's stances towards beliefs, but I also want to advise that a lot of work still needs to be done to vindicate it as a plausible theory of mind.

University of Cape Town

NOTES

[1] G. Ryle, *The Concept of Mind* (New York: Barnes and Noble, 1949).[2] Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press, 2001).[3] Few contemporary philosophers defend strict or absolute infallibility and omniscience, but several more modest versions of infallibility and omniscience are still supported. For a well-known version of limited infallibility, see F. Jackson, "Is There a Good Argument Against the Incorrigibility Thesis?," *Australasian Journal of Philosophy* 51 (1973). For contemporary versions of omniscience, see R. Chisholm, *The First Person* (Minneapolis: University of Minnesota Press, 1981). and even C. Peacocke, *A Study of Concepts* (Cambridge, MA: MIT Press, 1992). If infallibility and omniscience are theoretical versions of the first-person authority claim, then the fact that there are different versions of infallibility and omniscience should make my point even clearer. If we are not prepared to become involved in a lot of theorising, we should not use theory to determine the exact nature of the differences between first- and third-person beliefs.

[4] There is, of course, another strategy that can lead us to the asymmetries, namely, our phenomenological experiences of beliefs and believing. But this is by far the weakest approach. The best we can do is to rely on our verbal reports of our feelings about beliefs, but then a whole series of unanswerable questions will arise: does language accurately express feelings about thoughts and beliefs or are they more intricately connected? Do people who use the same words to refer to a feeling about a belief mean the same feeling by their words? How can disputes be resolved? And so forth. Verbal reports cannot help in an investigation of phenomenological experiences and those experiences themselves cannot be revealed in any other way.

[5] This difference in treatment can be read to correspond to the traditional theoretical notions of directness and/or immediacy. The evidence asymmetry differs from these claims in two ways. First, I am not claiming that first-person beliefs are more direct than third-person ones. I am merely claiming that they are treated as if they are not self-attributed by employing behavioural evidence. Second, where the immediacy claim proposes that first-person beliefs are not self-attributed on the basis of any evidence at all, I am refraining from taking a stand on whether self-attributions rely on evidence. I am just claiming that they are not treated as if they make use of behavioural evidence specifically.

[6] I am not necessarily unaware of the truth of the beliefs that I ascribe to others, and I am likely to understand that the beliefs of others are states that contain their grasp of the world. Still, my interest in their grasp of the world is primarily in the explanatory power it gives me.

[7] It is very much part of our ordinary discourse to remark, "even though Ben believes that coffee is healthier than tea, tea is, in fact, healthier." It is a way in which a speaker lets everyone know that Ben believes something that the speaker thinks is false. But if Ben had to report, "I believe that coffee is healthier than tea, but coffee contains more harmful ingredients than tea", we would either ask him to explain his belief, or we would simply label him irrational or deeply confused.

[8] On a theoretical level, the fact that we are treated as if we are authoritative with regard to our beliefs is usually taken to mean that we, in fact, possess such authority. In other words, we are not only treated as if we are more likely to be correct about our own beliefs during interaction with others; we are actually more likely to be correct about them. Why else would we treat people as default authorities over their own attitudes if they did not demonstrate such authority to others on a regular basis? Crispin Wright once criticised the practice of explaining away our authority over our propositional attitudes as either a linguistic convention like Wittgenstein did or as courtesy like Richard Rorty did. He believes strongly that such theories fail to justify or explain the practice of treating persons as default authorities. As he once noted, "They are no more than a mere invitation to choose to treat as primitive something which we have run into trouble trying to explain". See C. Wright, "The Wittgensteinian Legacy," in *Knowing our Own Minds*, ed. C. Wright, B. Smith, and C Macdonald (Oxford: Clarendon Press, 1998)., 45 Whether we are just treated as authorities, or whether we are, in fact, such authorities, is once again a theoretical debate which I do not need to settle here. Our practices indicate that we treat others as such authorities, and the question of whether they truly are such authorities just does not arise on the practical level.

[9] Which, of course, demonstrates a relationship between the transparency and authority asymmetries.

[10] This idea was first advanced by Wittgenstein, and later defended by Gareth Evans. See L. Wittgenstein, *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell, 1953). and G. Evans, *The Varieties of Reference* (Oxford: Oxford University Press, 1982).

[11] This picture of the existence of stable beliefs that are ready to be discovered is not opposed by interpretationists only. See, for example, Moran, *Authority and Estrangement: An Essay on Self-*

University of Cape Town

Knowledge. Moran conceives of beliefs as states that are partly shaped by our awareness and understanding of them. On this picture there is also little room for the question of how we come to know what the beliefs are that we already hold. After all, if the belief is altered by our awareness and interpretation of it, it is more shaped than discovered.

[12] The significance of univocality and the importance of accounting for the asymmetries are stressed by most theorists who have written on the subject. See, for example, R. Brandom, "Expressing and Attributing Beliefs," *Philosophy and Phenomenological Research* 54, no. 4 (1994). Richard Moran, "Interpretation Theory and the First Person," *Philosophical Quarterly* 44, no. 175 (1994). P. F. Strawson, *Individuals* (London: Methuen, 1959)., 110[13] This inference does not necessarily have to be a conscious one.

[14] This is the point where I will be accused of over-simplifying the theory theory's explanation of how we access our own beliefs, and the criticism will probably be justified. Most theory theorists deliberately attempt to steer clear of this picture because of its apparent implausibility. Most of them hold onto the claim that we access our own beliefs through some theory-mediated inference, but they all have different ideas of what the theory contains. Peter Carruthers, for example, believes that the theory whereby we attribute beliefs to others allows us, not only access to their behaviour, but also access to the environment in which they behave. Accordingly, the theory whereby we access our own beliefs has to give us this type of access to our environment and, since it gives us access to a third-person's speech, it has to allow us access to our own inner speech (or thoughts). Shaun Nichols and Stephen Stich, on the other hand, believe that this Carruthers-type of account does not rely on a theory anymore, and should, thus, be abandoned in favour of their theory that proposes a self-monitoring mechanism for detecting the contents of our belief box. They are aware of having to account for situations where the attitude asymmetries seem not to be present, which leads them to propose different mechanisms for detecting and reasoning about our beliefs. When we are wrong about our beliefs, we are not usually wrong about what we believe, but rather about how our beliefs cause our behaviour. So, when we are wrong about our beliefs, it is because we are using our theory of mind, instead of the self-monitoring mechanism. This certainly seems to be the kind of haphazard explanation that I referred to in section 4. For Carruthers' account, see P. Carruthers, "Simulation and Self-Knowledge: A Defence of Theory Theory," in *Theories of Theories of Mind*, ed. P. Carruthers and P. Smith (Cambridge, MA.: Cambridge University Press, 1996).. For Nichols and Stich's account, see S. Nichols and S. Stich, "Reading One's Own Mind," in *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds* (Oxford: Oxford University Press, 2003).. For their reasons for thinking that we usually get wrong the reasoning from our beliefs, as opposed to the beliefs themselves, see R. Nisbett and T. Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84 (1977). and R. Nisbett and L. Ross, *Human Inference: Strategies and Shortcomings of Social Judgment* (New Jersey: Prentice Hall, 1980)..

[15] Even though there are many who believe that this is not what he was trying to say.

[16] W. Sellars, "Empiricism and the Philosophy of Mind," in *Minnesota Studies in the Philosophy of Science*, ed. H. Feigl, Scriven, M. (Minneapolis: University of Minesota Press, 1956).[17] For Davidson's debate with Jerry Fodor, a contemporary theory theorist, see D. Davidson, "Psychology As Philosophy," in *Essays on Actions and Events* (London: Oxford University Press, 2001). and J. Fodor, "The Persistence of the Attitudes," in *Psychosemantics* (Cambridge: MIT Press, 1987).[18] This does not, as many may be lead to believe, mean that Davidson is an anti-realist or instrumentalist about beliefs. He thinks that there is a token (as opposed to a type/law-like) identity between beliefs so attributed and physical brain processes. Davidson talks about beliefs as mental events, since they are attributed under a specific description, but this makes them no less real than the attribution of properties like thirty centimetres or twelve inches to a ruler. Beliefs are physical brain states, in the same way as length is a property of a ruler, but the actual attribution of "a belief that the sun is shining" or of "thirty centimetres" depends upon the person's description of what he is observing from his own perspective.

[19] D. Davidson, "Thought and Talk," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1985)., 158[20] D. Davidson. "*A Coherence Theory of Truth and Knowledge*," In *Subjective, Intersubjective, Objective*,. (Oxford: Clarendon Press, 2001)., 148

[21] Ludwig Wittgenstein is commonly understood as putting forward a similar, though in some respects different, argument against the possibility of a private language. Wittgenstein's argument can be summarised as follows: if the meaning of a word did involve elements that were not publicly determinable, then it would imply a private language: a language whose meanings, or elements that make up parts of meanings, are necessarily available to only its originator; a language that is in principle indecipherable to others because the objects/events that give the words their meaning, or to

which the words refer, are necessarily available only to the speaker. There cannot be such a private language, though, because, for such a language to exist, the speaker would have to be able to correlate words with objects that are necessarily private. Davidson once noted that, "The point of meaning is synonymy - sameness of meaning, whether of different sentences for the same speaker or different speakers, or of the same sentence from speaker to speaker." The speaker can say that what he means by tree is the private mental picture he is now entertaining, and tomorrow he can say that he is entertaining the same mental picture that his word tree refers to. Thus, semantic content is bestowed on his private language by his private objects. But how can it ever be established that the mental pictures that are supposed to define "tree" are the same from one day to the next? As Wittgenstein will say, then ... "whatever is going to seem right to him is right. And that only means that here we can't talk about 'right'". This is because, in order to make a factual statement, such as "the mental picture I am now having is the same one I was having yesterday when I named it 'tree'", it has to be possible for the speaker to be wrong. Otherwise it cannot be a statement of fact. And this is the crux of Wittgenstein's argument against a private language. For a word to mean something, a correlation between the word and the object has to be established that allows for consistency in use. In a private language, however, such an initial correlation cannot be established, exactly because no method for future recognition of the object exists. If there is no second person to agree, or disagree with, about the nature of the object that defines the word, then no future consistency of meaning can be secured because then whatever seems right to the private linguist will be right. He, thus, cannot establish correlations between his words and objects that allow for future consistency in use because no strategy for future consistency will allow such a linguist to make factual statements about his private objects. Therefore, a language that is in principle private is impossible not only because it would be unintelligible to others, but because a private linguist would not be able to establish meanings for its words which will render such a language unintelligible to its creator as well. There is probably little in this argument that Davidson would disagree with. If there are differences between his overall project and that of Wittgenstein's, it is in the detail of how exactly language is essentially social. For his argument against the possibility of a private language, see Wittgenstein, *Philosophical Investigations*, sections 244-271. For a contemporary statement of the reading of Wittgenstein that Davidson claims to disagree with about the way in which language is essentially social, see S. Kripke, *Wittgenstein on Rules and Private Language* (Oxford: Blackwell, 1982).and D. Davidson, "The Second Person," in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001).[22] When Wittgenstein first proposed this argument against the possibility of a private language, one widespread objection employed private codes or diaries to show that a private language must be possible. The argument against private language does not, however, include situations in which people construct secret codes to represent environmental objects or events. Children, for example, often construct their own secret languages by making up words that refer to objects or events in their environment. My brother created fairly exotic words by substituting A for E, B for G, C for Z, D for T, etc., but he still applied the words he created to objects and events around him. So, in principle his secret code was not a private language because, by matching his utterances with the objects in his environment, others would have been able to figure out the meaning of the words in his language. Then there would have been no element of the meaning of his words that would have been available only to himself because, with some effort, cooperation and sufficient information, observers would have been able to decipher it.

[23] It should also be noted that the argument against the possibility of a private language holds, not only for what my words mean, but also for what I mean by my words. After all, what I mean by my words requires semantic content just as much as the meaning of words does. For, if what I mean by my words is not derived from some publicly available source, then what I mean by my words will have to be derived from an essentially private language also, such as from an exclusively first-person accessible picture in my mind. But, as mentioned above, there are good reasons for thinking that such a private language cannot exist. So, what I mean by my words will also have to be publicly determinable.

[24] D. Davidson, "Reality Without Reference," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1985)., 220[25] "Correct" obviously in the sense of whether it captures the truth conditions of the speaker's phrase accurately, as will be judged by an omniscient interpreter.

[26] The first question is a version of the classic problem of universal scepticism of whether our experiences genuinely connect us with an external world and the second the question of how we can know whether we are right about the nature of the world.

[27] D. Davidson. "Knowing One's Own Mind," *Proceedings and Addresses of the American Philosophical Association* 60, no. 3 (1987)., 445

[28] This is an argument that Davidson advanced against w.v.o. Quine's concept of stimulus meaning. For Davidson's argument, see D. Davidson, "Meaning, Truth and Evidence," in *Perspectives*

on Quine, ed. R. Barrett, Gibson, R. (Oxford: Blackwell, 1990). For Quine's response, see W.V.O. Quine, "Three Indeterminacies," in *Perspectives on Quine*, ed. R. Barrett and R. Gibson (Oxford: Blackwell, 1990).[29] D. Davidson, "The Emergence of Thought," in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001)., 129[30] Davidson, "The Second Person.", 119[31] D. Davidson. "*Radical Interpretation*," In *Inquiries into Truth and Interpretation*, (Oxford: Clarendon Press, 1985), 135

[32] Davidson, "Radical Interpretation.", 136

[33] Davidson, "Radical Interpretation.", 138 and D. Davidson, "Belief and the Basis of Meaning," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1985)., 154

[34] Interpretationists differ with regard to the amount of evidence that is essential to establish a speaker's interpretability. As seen here, Davidson includes not only a speaker's current behaviour and speech, but a huge amount of information about her past behaviour, speech and environmental factors.

[35] Davidson, "Knowing One's Own Mind," 441

[36] D. Davidson, "First-Person Authority," *Dialectica* 39 (1984)., 103[37] Ibid., 110[38] Ibid., 110

[39] [Davidson, "Knowing One's Own Mind," 456 and D. Davidson, "Reply to Thele," in *Reflecting Davidson* (Berlin: Walter de Gruyter, 1993)., 250[40] In Davidson's own words, "It would once more make the account circular to explain the basic asymmetry by assuming an asymmetry in the assurance you and I have that I hold the sentence I have just uttered to be a true sentence. There must be such an asymmetry, of course, but it cannot be allowed to contribute to the desired explanation." See Davidson, "First-Person Authority.", 109[41] Ibid., 111[42] That is, the speaker is not interpretable because she accidentally happens to apply her words consistently. She is interpretable because she applies her words consistently with the intention of providing clues about the meaning of her words to her interpreters. It is true that the intention has to be attributed by an interpreter. But that does not mean that the interpreter is attributing something that isn't actually present just because it will make better sense of what the speaker does. That is instrumentalism, not interpretationism. He is interpreting something that is already there and that is accessible to him.

[43] For Davidson's view on intentions in communication, see Davidson, "The Second Person."[44] Davidson, "First-Person Authority.", 110[45] One potential problem often raised to Davidson's use of Tarskian truth definitions is that the biconditional "if and only if" guarantees only that the phrase on the left will have the same truth value as that on the right, and that this makes it possible to use any phrase on the right so long as its truth value is identical to that on the left. Since Davidson insists that a theory of meaning, and therefore the construction of the biconditionals, has to conform to what people as a matter of fact do when they make sense of others, this is not a problem. In practice, we do not simply use any phrase on the right of the biconditional. We judge under what environmental conditions speakers use the phrase on the left. The way in which the biconditional is allowed to be constructed has to be empirically verifiable by what language speakers actually do. Furthermore, simply using any phrase on the right of the biconditional with an identical truth value to the phrase on the left is likely to fail to make sense of the speaker in other contexts. That is precisely the role that Davidson wants correspondence and coherence to play in his system of interpretation.

[46] This is part of Davidson's idea of triangulation. For a full defence, see D. Davidson, "Three Varieties of Knowledge," in *Subjective, Intersubjective, Objective* (Oxford: Clarendon Press, 2001)., originally published in 1991.

[47] [B. Thele, "The Explanation of First-Person Authority," in *Reflecting Davidson*, ed. R. Stoecker (Berlin: Walter de Gruyter, 1993)., 239

[48] K. Ludwig, "First-Person Knowledge and Authority," in *Language, Mind and Epistemology*, ed. G. Preyer, F. Siebelt, and A. Ulfig (Dordrecht: Kluwer Academic Publishers, 1994)., 186[49] D. Beisecker, "*Interpretation and First-Person Authority: Davidson on Self-Knowledge*" (cited April 06 2006); available from http://www.unlv.edu/faculty/beisecker/Research/FPA-SWPR.pdf.[50] The terms immediate and authoritative are heavily theory-laden, but in this context I am using them to apply to the evidence and authority asymmetries as set out in chapter 2.

[51] [Thele, "The Explanation of First-Person Authority," 243

[52] B. Smith, "On Knowing One's Own Language," in *Knowing Our Own Minds*, ed. C. Wright, B. Smith, and C. MacDonald (Oxford: Clarendon Press, 1998)., 417

[53] Ludwig, "First-Person Knowledge and Authority.", 391, seems to miss the point that it is a deflationary approach to knowing what one means when he criticises Davidson for putting forward an explanation of self-knowledge that "gets the order of the explanation backwards". Ludwig thinks that Davidson proposes that the fact that one is interpretable gives one the knowledge of what one means. He wants it to be the other way around, namely, that knowing what one means makes one interpretable.

Davidson, however, does not have to struggle with the order of the explanation, because he isn't proposing that the one causes the other.

[54] Smith, "On Knowing One's Own Language," 418

[55] Thele, "The Explanation of First-Person Authority," 243-244

[56] I deliberately refrained from employing an omniscient radical interpreter to make the same point, since I suspect that it is theorists' reluctance to accept it that leads them to objections such as these. This is precisely what Davidson used his concept of an omniscient interpreter for, though. If an interpreter had all the non-semantic information about the speaker, then such information, after interpreted for meaning, would have revealed the speaker's agnosticism.

[57] Another difficulty for Thele's objection is that, if the speaker does not know whether two words mean the same or not, then what could possibly be responsible for the fact that she applies them interchangeably to refer to the same sort of event, and especially the fact that she does so consistently. Thele, because he is responsible for the more unlikely claim, will have to accept the burden of accounting for this phenomenon.

[58] Smith, "On Knowing One's Own Language.", 417 Thele, "The Explanation of First-Person Authority.", 245[59] Ludwig, "First-Person Knowledge and Authority.", 390 Thele, "The Explanation of First-Person Authority.", 245[60] Ibid., 245[61] Davidson, "First-Person Authority.", 111

[62] S. Sawyer, "*An Externalist Account of Introspective Knowledge*", 1999 (cited April 06 2006); Available from http://www.philosophy.ku.edu/faculty/Sawyer/Ext_Int.html.

[63] E. Picardi, "First-Person Authority and Radical Interpretation," in *Reflecting Davidson*, ed. R. Stoecker (Berlin: Walter de Gruyter, 1993)., 202, for example, expresses the typical objection as follows: "I cannot be meaning to refer to anything to which I explicitly do not mean to refer". Firstly, Davidson will reject this way of stating the problem, since he rejects reference as the basis of meaning. Secondly, even if two interpreters assign different truth conditions to a speaker's utterance, it does not imply that the speaker "refers" to something she explicitly does not mean to "refer". She means whatever truth conditions her biconditional states, and there is a possibility that an interpreter can make sense of her by attributing whatever truth conditions he judges her utterance to state in his language.

[64] For Davidson's view on intentions in communication, see Davidson, "The Second Person."[65] Davidson, "Reality Without Reference."[66] The interpreter maps her language onto his own and judges that, by the sentence "My husband is not a murderer", she means what he, in his own language, means when he says "My husband is a murderer".

[67] Think in the sense of being so interpretable to an omniscient radical interpreter, of course.

[68] Barry Smith and Kirk Ludwig have both proposed something like the following: the meaning asymmetry can explain the attitude asymmetries only if there is a meaning asymmetry between a speaker and her interlocutors. Intuition tells us that there is no meaning asymmetry between her and her interlocutors. Therefore, in practice there is no meaning asymmetry. Those who put forward this objection usually claim that those with whom a speaker communicates just hear the meaning of her words automatically and unreflectively along with her speech. When she speaks, her interlocutors do not hear her words and then interpret to check whether they have their meaning right, or that they have their meaning at all. They know the meaning of her sentences straight away, and they will only hesitate and carefully interpret when there are severe difficulties communicating. See Ludwig, "First-Person Knowledge and Authority." Smith, "On Knowing One's Own Language.". This objection seems to appeal to personal reports about the experience of hearing meaning in the speech of others. Up to now I have opted to steer clear of personal reports about believing which is an approach I choose to stand by in the case of meaning as well. Most of us can probably relate to the above observations as an accurate reflection of what we experience when someone speaks. It is, however, once again the type of report that it is difficult to learn something from, since the reports are likely to vary according to different individuals' levels of self-examination, language proficiency and theories about meaning and communication. Moreover, even if it is possible to reach agreement about our experience of hearing meaning along with speech, the objection will still have to assume that it is necessarily the case that the meaning asymmetry will be apparent in our experience if it is present. Such awkward questions about phenomenology and its connection with reports about personal experience can be easily avoided by focussing on arguments for and against the presence of radical interpretation in practice. If the meaning asymmetry requires a situation of radical interpretation (which I intend to show below), then reasons for and against the occurrence of radical interpretation in practice can settle the matter of the presence of the meaning asymmetry without entangling us in the tricky web of phenomenology.

[69] Claim 3 can be understood in two ways: first, as a claim about what we as a matter of fact do in practice and, second, as a denial of the claim that the theoretical assumptions of the radical case must constrain any conception of the mental, including the practical case. The first may, as a result of

University of Cape Town

convention, be something Davidson would subscribe to; the second definitely not. In this paragraph I obviously meant the first.

[70] Together with being able to state the truth conditions accurately through disquotation and being an interpreter of others.

[71] The terms "immediate" and "authoritative" are heavily theory-laden, but I am employing them to apply to the evidence and authority asymmetries as set out in chapter 2 respectively.

[72] D. Davidson, "A Nice Derangement of Epitaphs," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore (Cambridge, MA: Blackwell, 1986)., 446[73] Davidson accuses Michael Dummett and Andreas Kemmerling, among others, of omitting the qualifying phrase "what philosophers and linguists have supposed" from their analysis of his statement. Dummett once remarked that, "Whatever force Davidson's arguments may have, they cannot sustain the bald conclusion, but cry out for some account of an indispensable concept', namely, the concept of a language M. Dummett, "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. Lepore (Oxford: Clarendon Press, 1986)., 465-466 For Davidson's response, see D. Davidson, "The Social Aspect of Language," in *The Philosophy of Michael Dummett*, ed. B. McGuinness (Dordrecht: Kluwer, 1994)., 2-3 Also see D. Davidson, "Reply to Andreas Kemmerling's the Philosophical Significance of a Shared Language," in *Reflecting Davidson*, ed. R. Stoecker (Berlin: Walter de Gruyter, 1993)., 118[74] Some philosophers think that Davidson is mistaken about the exact nature of linguistic conventions. For one example, see P. Rysiew, "Conventional Wisdom,"*Analysis* 60, no. 1 (2000).[75] In his 1775 comedy, The Rivals, Richard Sheridan (1751-1816) introduced a humorous character called Mrs. Malaprop. The self-educated Mrs. Malaprop constantly substituted similar-sounding words for the words that she actually intended to use, which often resulted in the most ridiculous sentences. Examples include: "I hope you will represent her to the captain as an object not altogether illegible." (eligible) "...she might reprehend the true meaning of what she is saying." (comprehend) "...if ever you betray what you are entrusted with... you forfeit my malevolence forever..." (benevolence) and the classic sentence which Davidson himself used to illustrate his point, "Sure, if I reprehend any thing in this world it is the use of my oracular tongue, and a nice derangement of epitaphs." (apprehend, vernacular, arrangement, epithets)

[76] It must be noted that one such slip does not change the meaning of, say epitaph, in Mrs. Malaprop's personal idiolect. Subsequent conversations with her will settle whether the word "epitaph" means "epitaph" or "epithet" in her idiolect, depending on what she applies it to consistently. This point was also made by Alexander George. He mistakenly attributed the opposite view to Davidson, though. See A. George, "Whose Language Is It Anyway? Some Notes on Idiolects,"*Philosophical Quarterly* 40, no. 160 (1990). Whether the meaning of "epitaph" changes in her personal idiolect or not, the interpreter still has to use creative interpretation to figure out what she means when she uses it.

[77] Davidson, "A Nice Derangement of Epitaphs.", 446[78] Davidson explains this process as follows: During interpretation the interpreter transforms a "prior theory" into a "passing theory. The hearer's prior theory is how he is prepared to interpret an utterance of the speaker beforehand, while his passing theory is how he actually ends up interpreting it. The speaker's prior theory is what she thinks the interpreter's theory is, while her passing theory is the theory she wants the interpreter to use. What the interpreter and the speaker should share is the passing theory. That is, he has to interpret her in the way that she intends to be interpreted. The prior theory cannot be learned in advance since we don't have the same prior theory for different people. Moreover, we have different prior theories for one person, depending on the situation in which we are interpreting her. The passing theory can also not be learned in advance (as the case of Mrs. Malaprop illustrates. It has to be created on the fly. Ibid.[79] Uwe Wirth offers an explanation mined from Charles Peirce. The surprising fact, C, is observed; But if A were true, C would be a matter of course, Hence, there is reason to suspect that A is true. U. Wirth, "Abductive Reasoning in Peirce's and Davidson's Account of Interpretation,"*Transactions of the Charles S. Peirce Society* 35, no. 1 (1999). For another view on how this might work, see, S. Yarbrough, "Passing Theories Through Topical Heuristics: Donald Davidson, Aristotle, and the Conditions of Discursive Competence,"*Philosophy and Rhetoric* 37, no. 1 (2004).[80]

Davidson, "The Social Aspect of Language.", 10[81] See D. Davidson, "Communication and Convention," in *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1985). D. Davidson, "A Nice Derangement of Epitaphs." and D. Davidson, "The Social Aspect of Language." amongst others.

[82] See J. Bennett, "Critical Notice: Davidson's Inquiries Into Truth and Interpretation,"*Mind* 94 (1985)., 603[83] D. Bar-On and M. Risjord, "Is There Such a Thing As a Language?,"*Canadian*

University of Cape Town

Journal of Philosophy 22 (1992)., 185-186 and A. Kemmerling, "The Philosophical Significance of a Shared Language," in *Reflecting Davidson*, ed. R. Stoecker (Berlin: Walter de Gruyter, 1993).[84] Davidson made clear that he was aware of this distinction. See Davidson, "A Nice Derangement of Epitaphs.", 434 and Davidson, "Reply to Andreas Kemmerling's the Philosophical Significance of a Shared Language.", 119[85] If one wants to analyse Davidson's argument, one should, thus, focus on his idea that speaker meaning is the source of meaning.

[86] Davidson, "The Social Aspect of Language.", 9[87] Davidson, "Reply to Andreas Kemmerling's the Philosophical Significance of a Shared Language.", 117 and Davidson, "The Social Aspect of Language.", 11[88] Ibid., 10[89] Again, something cannot be called the meaning of an utterance or the knowledge of the meaning of an utterance if it is not in principle interpretable.

[90] My interlocutors, of course, do not actually have to check with me or with my radical interpreter. The correctness is determined by what I or my radical interpreter would say if the interlocutor were to check.

[91] The notion of "right" appealed to here is obviously that of being similar or identical to what a radical interpreter would have attributed.

[92] The notion of "right" appealed to here is obviously that of being similar or identical to what a radical interpreter would have attributed.

[93] Firstly, if traditional interpretationists can offer an argument in support of the view that our practices are mistaken and that they should be modified, then this strategy can work. In the absence of such an argument, I will assume that this strategy is not satisfactory. Secondly, my modification of the evidence asymmetry in chapter 2 cannot help, since the strategy will then have to show why we usually employ a speaker's behaviour as evidence to know what belief to attribute to her. It is, however, the use of evidence which it cannot account for.

[94] By "outside our linguistic communities" I simply mean those with whom we share no language. Those inside our linguistic communities, I take to be those with whom we do share a language and conventions of how to use that language.

[95] This does not contradict anything that Davidson says. In this type of scenario, it is possible for their to be a radical interpreter who is using no conventions at all to interpret the speaker.

[96] Theorists may claim that this is no different from the radical case, where our radical interpreters also have to start with almost no evidence, but prompt us for evidence. The case remains, however, that our interlocutors seem to know what we mean without ever having to prompt us for verbal or behavioural cues in the presence of environmental events/objects. Radical interpreters have to prompt, our interlocutors do not have to. And radical interpreters, since they are conceptual rather than actual, have access to information about what we would be disposed to do in certain situations, our interlocutors do not.

[97] Davidson, "A Nice Derangement of Epitaphs.", 446[98] A fifth strategy that a critic offered to me is the following: (A) in both the radical and the non-radical case, there is the potential of radical interpretation. (B) The potential of radical interpretation in both cases lead to a meaning asymmetry which can explain the attitude asymmetries. (C) In the non-radical case, we use conventions instead of radical interpretation to understand what a speaker means. However, since, even in the non-radical case, we know that (A) and (b) pertain, we treat our actual conclusions as secondary to those that we would have reached through radical interpretation. According to this strategy, however, we still do not use evidence to know what our interlocutors mean, even if we know that we potentially could or should have. And then this strategy does not amount to much more than the one discussed in 2.2. The speaker knows what she means since she is radically interpretable as applying her words consistently. We may not know what she means since our conventions may fail to make sense of her. She, thus, knows what she believes while we may not. But the evidence and transparency asymmetries require that we use evidence to interpret for meaning, which this strategy cannot make room for. And if it does make room for it, then it will become the strategy in 2.3, and require us to use evidence which just is not available.

[99] This is a point that Davidson himself makes with his use of the notion of triangulation. A speaker cannot have knowledge of the minds of others without having knowledge of the world and of her own mind. Knowledge of the world is required to know the minds of others since she has to judge which environmental conditions he is holding his sentences true in. Knowledge of her own mind is required to know the minds of others since she has to assign meaning to his utterances by judging which sentences she holds true in the environmental conditions that she thinks he holds his sentence true in. Davidson, however, treats triangulation like a postscript to his system of radical interpretation; possibly because it is something that only dawned on him much later. Taking the point seriously that a

University of Cape Town

speaker has to know which sentences she holds true in which environmental conditions can solve many of the problems that his meaning asymmetry was criticised for.

[100] This also holds in a radical case where an interpreter has been interpreting a speaker for long enough to know approximately which sentences she holds true under which environmental conditions. Her past perceptions of her environment play just as much of a role in the truth conditions of her sentences in the radical case than it does in the non-radical case.

[101] This is similar to the radical case, in which the meaning asymmetry rests on the fact that the radical interpreter may be wrong about what she means since he has to continue interpreting her, even if he is a fully informed interpreter of her. And then his conclusions about what she means may at any point be wrong since his interpretive scheme may at any point become inappropriate to make sense of her.

[102] Not a language in the sense of having to map speakers' sentences onto it, though. In the radical case, the interpreter interprets the speaker by observing her speech and behaviour in the context of her environment. He hears a sentence that he assumes she holds true, he prompts her to indicate which environmental event she is applying it to (by pointing, for example), he assumes that her sentence means the same as the sentence that he would have held true when applying it in the presence of that environmental event. In the non-radical case, the speaker utters a sentence which meaning her interlocutor knows without having to consult environment or further behaviour since she uses the same conventions when speaking that he uses when understanding her. There is, thus, no mapping of one language onto another because of the assumption that the two are the same.

[103] Davidson, "First-Person Authority.", 111 [104] S. Sawyer, "*An Externalist Account of Introspective Knowledge*", 1999 (cited April 06 2006); Available from http://www.philosophy.ku.edu/faculty/Sawyer/Ext_Int.html.

University of Cape Town

References:

Bar-On, D., and M. Risjord. "Is There Such a Thing As a Language?" *Canadian Journal of Philosophy* 22 (1992): 163-90.

Beisecker, D. *Interpretation and First-Person Authority: Davidson on Self-Knowledge* , [cited April 06 2006]. Available from http://www.unlv.edu/faculty/beisecker/Research/FPA-SWPR.pdf.

Bennett, J. "Critical Notice: Davidson's Inquiries Into Truth and Interpretation." *Mind* 94 (1985): 601-26.

Brandom, R. "Expressing and Attributing Beliefs." *Philosophy and Phenomenological Research* 54, no. 4 (1994): 905-12.

Carruthers, P. "Simulation and Self-Knowledge: A Defence of Theory Theory." In *Theories of Theories of Mind*, edited by P. Carruthers and P. Smith, 22-38. Cambridge, MA.: Cambridge University Press, 1996.

Chisholm, R. *The First Person*. Minneapolis: University of Minnesota Press, 1981.

Davidson, D. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E. LePore. Cambridge, MA: Blackwell, 1986.

———"A Coherence Theory of Truth and Knowledge." In *Subjective, Intersubjective, Objective*, 137-54. Oxford: Clarendon Press, 2001.

———"Belief and the Basis of Meaning." In *Inquiries into Truth and Interpretation*, 141-55. Oxford: Clarendon Press, 1985.

———"Communication and Convention." In *Inquiries into Truth and Interpretation*, 265-80. Oxford: Clarendon Press, 1985.

———"First-Person Authority." *Dialectica* 39 (1984): 101-11.

———"Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60, no. 3 (1987): 440-59.

———"Meaning, Truth and Evidence." In *Perspectives on Quine*, edited by R. Barrett, Gibson, R., 68-79. Oxford: Blackwell, 1990.

———"Psychology As Philosophy." In *Essays on Actions and Events*, 229-41. London: Oxford University Press, 2001.

———"Radical Interpretation." In *Inquiries into Truth and Interpretation*, 125-39. Oxford: Clarendon Press, 1985.

———"Reality Without Reference." In *Inquiries into Truth and Interpretation*, 215-26. Oxford: Clarendon Press, 1985.

University of Cape Town

———"Reply to Andreas Kemmerling's the Philosophical Significance of a Shared Language." In *Reflecting Davidson*, edited by R. Stoecker. Berlin: Walter de Gruyter, 1993.

———"Reply to Thele." In *Reflecting Davidson*, 248-51. Berlin: Walter de Gruyter, 1993.

———"The Emergence of Thought." In *Subjective, Intersubjective, Objective*, 123-35. Oxford: Clarendon Press, 2001.

———"The Second Person." In *Subjective, Intersubjective, Objective*, 107-23. Oxford: Clarendon Press, 2001.

———"The Social Aspect of Language." In *The Philosophy of Michael Dummett*, edited by B. McGuinness. Dordrecht: Kluwer, 1994.

———"Thought and Talk." In *Inquiries into Truth and Interpretation*, 155-71. Oxford: Clarendon Press, 1985.

———"Three Varieties of Knowledge." In *Subjective, Intersubjective, Objective*, 205-21. Oxford: Clarendon Press, 2001.

Dummett, M. "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by E. Lepore, 459-76. Oxford: Clarendon Press, 1986.

Evans, G. *The Varieties of Reference*. Oxford: Oxford University Press, 1982.

Fodor, J. "The Persistence of the Attitudes." In *Psychosemantics*, 1-27. Cambridge: MIT Press, 1987.

George, A. "Whose Language Is It Anyway? Some Notes on Idiolects." *Philosophical Quarterly* 40, no. 160 (1990): 275-98.

Jackson, F. "Is There a Good Argument Against the Incorrigibility Thesis?" *Australasian Journal of Philosophy* 51 (1973): 51-62.

Kemmerling, A. "The Philosophical Significance of a Shared Language." In *Reflecting Davidson*, edited by R. Stoecker. Berlin: Walter de Gruyter, 1993.

Kripke, S. *Wittgenstein on Rules and Private Language*. Oxford: Blackwell, 1982.

Ludwig, K. "First-Person Knowledge and Authority." In *Language, Mind and Epistemology*, edited by G. Preyer, F. Siebelt and A. Ulfig, 366-98. Dordrecht: Kluwer Academic Publishers, 1994.

Moran, Richard. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press, 2001.

———"Interpretation Theory and the First Person." *Philosophical Quarterly* 44, no. 175 (1994): 154-73.

University of Cape Town

Nichols, S., and S. Stich. "Reading One's Own Mind." In *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, 150-200. Oxford: Oxford University Press, 2003.

Nisbett, R., and L. Ross. *Human Inference: Strategies and Shortcomings of Social Judgment*. New Jersey: Prentice Hall, 1980.

Nisbett, R., and T. Wilson. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (1977): 231-59.

Peacocke, C. *A Study of Concepts*. Cambridge, MA: MIT Press, 1992.

Picardi, E. "First-Person Authority and Radical Interpretation." In *Reflecting Davidson*, edited by R. Stoecker, 196-209. Berlin: Walter de Gruyter, 1993.

Quine, W.V.O. "Three Indeterminacies." In *Perspectives on Quine*, edited by R. Barrett and R. Gibson, 1-16. Oxford: Blackwell, 1990.

Ryle, G. *The Concept of Mind*. New York: Barnes and Noble, 1949.

Rysiew, P. "Conventional Wisdom." *Analysis* 60, no. 1 (2000): 74-83.

Sawyer, S. *An Externalist Account of Introspective Knowledge* , 1999 [cited April 06 2006]. Available from http://www.philosophy.ku.edu/faculty/Sawyer/Ext_Int.html.

Sellars, W. "Empiricism and the Philosophy of Mind." In *Minnesota Studies in the Philosophy of Science*, edited by H. Feigl, Scriven, M. Minneapolis: University of Minesota Press, 1956.

Smith, B. "On Knowing One's Own Language." In *Knowing Our Own Minds*, edited by C. Wright, B. Smith and C. MacDonald. Oxford: Clarendon Press, 1998.

Strawson, P. F. *Individuals*. London: Methuen, 1959.

Thele, B. "The Explanation of First-Person Authority." In *Reflecting Davidson*, edited by R. Stoecker, 213-47. Berlin: Walter de Gruyter, 1993.

Wirth, U. "Abductive Reasoning in Peirce's and Davidson's Account of Interpretation." *Transactions of the Charles S. Peirce Society* 35, no. 1 (1999): 115-27.

Wittgenstein, L. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Blackwell, 1953.

Wright, C. "The Wittgensteinian Legacy." In *Knowing our Own Minds*, edited by C. Wright, B. Smith and C Macdonald. Oxford: Clarendon Press, 1998.

Yarbrough, S. "Passing Theories Through Topical Heuristics: Donald Davidson, Aristotle, and the Conditions of Discursive Competence." *Philosophy and Rhetoric* 37, no. 1 (2004): 72-91.

University of Cape Town