

**The London School of Economics and Political
Science**

How we ascribe beliefs to others

Gary Jones

A thesis submitted to the Department of Philosophy,
Logic and Scientific Method of the London School of
Economics and Political Science for the degree of Doctor
of Philosophy, September 2008.

UMI Number: U615291

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U615291

Published by ProQuest LLC 2014. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.



ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

THESES
F
9057



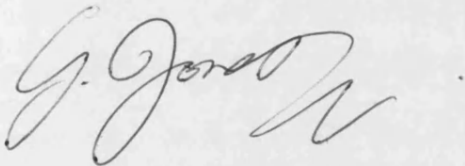
1198438

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work, other than where I have already indicated that it is the work of others (in which case the content of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of the thesis rests with the author. Quotation from it is permitted provided that full acknowledgement is made. The thesis may not be reproduced without the prior written consent of the author.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

A handwritten signature in black ink, appearing to read 'G. Jordan', followed by a small dot.

ABSTRACT

This thesis aims to provide a semantic account of belief ascriptions of the form ‘A believes that S’. It begins with a detailed discussion of Saul Kripke’s famous ‘Puzzle about Belief’, and tries to unearth a fundamental but rarely explicitly articulated assumption that gives the steps of the derivation of Kripke’s Puzzle their intuitive plausibility. The assumption can be roughly stated as follows: belief ascriptions report on some single ontologically prior mental state which grounds the truth of a true ascription. A response to one form of Kripke’s Puzzle is suggested at the end of this discussion.

Several apparently diverse kinds of response to Kripke’s Puzzle are critically evaluated and are found to be unsatisfactory. Many of the problems with these responses are traced to the fact that they are built upon the assumption that I suggest is at the heart of Kripke’s Puzzle.

The beginnings of a positive account of the semantics of belief ascriptions are given. The account gives a central role to the fact that the truth-value of a belief ascription depends on elements of the conversational setting in which it occurs. It is suggested that belief ascriptions are essentially answers to questions asked by an audience who has specific interests and makes specific assumptions about the agent and the setting of the ascription. The interests of the audience and the background assumptions that she makes are two-distinct sources of the context-sensitivity of belief ascriptions. The account makes no appeal to any kind of inner mental representations, but instead says that the truth of an ascription depends in a complex way on the agent’s dispositions and capacities to do, say, think, and feel certain things.

The thesis ends with a discussion of the *de dicto*/*de re* distinction, and suggests that the distinction does not provide the most useful way of understanding belief ascription. It is argued that truly *de re* ascriptions are probably very rare, and that this does not therefore mean that most ascriptions are *de dicto*.

Acknowledgements

I wish to thank the Arts and Humanities Research Council for their financial support from 2003-2006, and the Royal Institute of Philosophy for their Jacobsen Fellowship in 2002.

I owe a huge debt of thanks to my two supervisors, Genoveva Marti and Nancy Cartwright. It is in large part due to Genoveva that I fell in love with philosophy of language. I remember the moment: Genoveva gave the class two whole weeks to read and think about the first thirteen sentences of 'On Sense and Reference'. I think that sums up the approach that I admire so much, and try to aspire to. I have benefited greatly from Genoveva's advice and am grateful that she has been prepared to supervise from such a distance. My trips to Spain, where I stayed in her home with her family, have provided some of my fondest memories over the past few years. Nancy has been an amazing supervisor, giving me frighteningly penetrating comments on my work, and keeping me going with words of encouragement. Nancy's personal support, especially during last year, will always be appreciated.

John Worrall has also given me good practical advice for which I am very grateful.

I have benefited from, and enjoyed enormously, conversations with Helen Steward about causation, explanation and belief. I feel sure that Dr. Steward is *right* about many of these issues, and it seems to be almost impossible to be right about these things!

I would like to thank my friends at LSE: Leah McClimans, Sabina Leonelli, and Steffi Ortmann. Individually and collectively they have had a huge influence on how I think about philosophy. But more importantly they are really good friends.

I would also like to thank my friends Alan and Russell who have supported me in many ways over the past few years, as well as Alex Papadopoulos, Alastair Barr, Laurent Lachal and Rachael Botley.

Most of all, I would like to thank my parents Julie and Teg, and my sister Becca, for all their moral, emotional and practical support. They have never made me feel I'm not doing the right thing. That has meant everything to me.

Contents

Chapter One: Kripke's Puzzle about Belief

1.1 Introduction	13
1.2 How Kripke derives the puzzle	14
1.3 Irrelevancy of the Disquotatation Principle	21
1.4 The structure of Kripke's argument	27
1.5 Exposing a possible flaw in Kripke's reasoning	29
1.6 Foregrounding the basic assumption	32
1.7 Desiderata for a resolution of the puzzle	35
1.8 Suggesting a new response to the puzzle	36
1.9 Other types of response	41
1.10 Conclusion	45

Chapter Two: A Grammatical Interlude

2.1 Introduction	46
2.2 Zeno Vendler's classification of verbs	47
2.3 Nominal forms	49
2.4 A qualification about 'facts'	55
2.5 Conclusion	55

Chapter Three: Evaluating Other Theories

3.0 Introduction	57
3.1 Frege on simple sentences	58
3.2.1 Semantics of belief ascriptions and substitutivity	64
3.2.2 Separating issues about simple sentences from issues about 'thought'	70
3.2.3 Separating issues about mind from issues about ascriptions	73
3.3 What Frege might have said in response to	

Kripke's puzzle	79
3.4.1 Nathan Salmon's theory of belief ascriptions	85
3.4.2 Criticisms of Salmon's theory	93
3.5.1 A more radical no-content theory:	
Eaker and Almog	99
3.5.2 Application of this theory to Kripke's puzzle	101
3.5.3 Criticisms of Eaker and Almog's theory	107
3.6.1 Richard's contextualist account	113
3.6.2 Criticisms of Richard's theory	117
3.7 Conclusion	125

Chapter Four: Proposing an Account of Belief Ascriptions

4.0 Introduction	127
4.1 The central desideratum	128
4.2 Other desiderata	129
4.3 Giving the truth-condition	131
4.4 How to think about belief ascription	133
4.4.1 Proper questions	135
4.4.2 Contexts	136
4.5 The truth-condition stated	136
4.6 Metaphysical background	137
4.6.1 Dispositions and ascriptions	142
4.6.2 Dispositions and ascriptions	144
4.7 The examples	146
4.8 Defending and clarifying the meaning of the truth-condition	158
4.9 How my account differs from Richard's	161
4.10 De dicto and de re	163
4.11 Conclusion	167
Conclusion	169
References	171

Introduction

The object of study of this dissertation is the class of what I call ‘belief ascriptions’ or ‘belief reports’ which are utterances of sentences of the form ‘A believes that S’, where A denotes an agent and S is to be replaced by a sentence of English. There is a vast and still expanding literature on this subject. Much of the literature has developed as a response to what appear to be philosophical puzzles arising from reflection on our belief reporting practices and the apparent truth-values that our belief reports have.

Frege was the first to put forward anything like a theory of belief report sentences, which fitted into his overall semantic project, and sought to explain the apparent failure of substitutivity of co-referring terms in the embedded sentence of the belief report sentence. There has been much criticism of the details of Frege’s theory, which despite its theoretical elegance seems to face many difficulties. Although Frege’s theory has, in many of its details, been generally rejected, there are certain assumptions that Frege makes that are still retained. Some of these assumptions are amongst those that this dissertation aims to cast doubt on. I shall discuss Frege’s theory in Chapter 3, and try to cast doubt on some conceptual connections that are generally taken for granted by Frege and also by later writers. In particular, I will use my discussion of Frege to try to clarify the relations between belief reports and questions to do with our epistemic attitudes to sentences, and to separate out belief conceived of as a topic in philosophy of mind from questions about belief ascriptions. I believe that these distinctions are often conflated, and that they should be considered separately.

Much of the more recent literature on the subject has emerged as a response to a famous puzzle developed by Saul Kripke in his landmark ‘A Puzzle about Belief’ (1979). I, therefore, give Kripke’s puzzle a very central role in this dissertation. Almost every theory of belief reports that has emerged in recent years gives a central role to Kripke’s puzzle. Philosophers have often

used Kripke's puzzle as a test case for their theory; the theory can only be acceptable if it provides a good solution to Kripke's Puzzle. There is a great diversity in types of responses, and many very different perspectives on the puzzle are taken. Despite this diversity, I think that some assumptions are held in common by at least the vast majority of responses to the puzzle. In Chapter One I try to make it plausible that one particular, generally only implicitly held, assumption is at the heart of Kripke's puzzle and is maintained by most theorists which discuss it. The assumption can be roughly stated as follows (although what the assumption actually amounts to will probably only become clear throughout the whole discussion of the dissertation): the truth of belief reports depend on the obtaining of a single cognitively specifiable fact about the agent, a fact which is ontologically prior to the belief report and serves to ground the truth of the report. In traditional theories, this fact is a purely context-independent fact about the agent. In more recent theories, which tend to accept that belief ascriptions depend for their truth on elements of the context of utterance, the cognitively specifiable fact does not by itself determine the truth-value of the ascription, but nevertheless there is thought to be some single cognitive fact which underlies the truth of the report.

I see this assumption as pervasive and as being the source of many difficulties. The central aims of this dissertation are:

- 1) to suggest that such an assumption is indeed at the heart of Kripke's puzzle in the sense of providing the best explanation of the form of argument that Kripke uses to derive his puzzle about belief,
- 2) show that the assumption is pervasive and is the source of many different problems in a diverse range of different theories
- 3) provide at least the beginnings of my own account of belief ascriptions which abandons the central assumption in question.

I aim to provide a detailed analysis of Kripke's puzzle in Chapter One. The actual puzzle is relatively well-known and tends to be stated very quickly in

most writings which aim to offer a response to it, since familiarity with the puzzle is generally assumed. I intend to move slowly through a discussion of the derivation of the puzzle because I want to bring to the surface the assumption that I think underlies it, an assumption which is not generally articulated. I want also to bring out some features of the structure of Kripke's argument and to try to suggest what assumptions Kripke was making but not articulating and explain why, given these assumptions, he had to argue in a very particular way. The chapter discusses in detail some of the principles that Kripke saw our belief reporting practice as embodying and tries to cast doubt on both the truth and the relevance of such principles. By showing that the principles are not quite as central as is generally thought, I try to deflect the source of the puzzle onto other assumptions. I discuss the distinction between describing and deriving the puzzle in a metalinguistic mode and a material mode, which corresponds to the difference between seeing the puzzle as a puzzle about *belief-reporting* and seeing it as a puzzle about *belief*. I suggest that if we focus on belief ascriptions rather than belief, the puzzle is less easily derivable. The idea that belief reporting is parasitic upon belief is closely related to the central assumption that I discussed above. Chapter Two will try to offer some support to the idea that we should focus on belief reports of the form 'A believes that S' rather than assume that we have some prior and independent notion of belief which serves to ground those reports.

Chapter One suggests that Kripke's puzzle is essentially to be seen as an apparent tension in ascribing contradictory beliefs to an agent who is presumed to be rational.

Having isolated what I think are the key features in the derivation of Kripke's Puzzle, I suggest my own response to the puzzle which involves denying the argument that leads to the attribution of contradictory beliefs. I also discuss the notion of rationality and suggest that it is not to be understood in a way that depends on a prior grasp of what it is that an agent believes. The sketch of a solution suggested in the last section of this chapter will

hopefully be seen as a natural corollary to the more complete view of belief ascriptions that I develop in Chapter Four.

Chapter Two is a short interlude in which I discuss the ‘logical grammar’ of belief and look at various locutions involving ‘belief’ and its associated nominals and ‘believes that’. I use this section to suggest that ‘believes that’ is the basic expression in this family of related locutions and that sometimes philosophers use other locutions which may distort how we think about belief and belief ascriptions. I urge the methodological point that in trying to understand the phenomenon of belief we should focus on belief ascriptions of the form ‘A believes that S’. I finish this chapter with a short comment on the relation between the semantics and metaphysics of belief.

Chapter Three provides a critical evaluation of several apparently quite diverse theories of belief ascription, especially as they treat Kripke’s Puzzle. I begin this chapter with a discussion of Frege in order to allow me to make some conceptual distinctions that I think are necessary for a proper discussion of the subject. I then discuss what Frege could have said in response to Kripke’s Puzzle and suggest that while Frege’s theory has the resources to block the derivation of Kripke’s Puzzle it is in the end far too restrictive and makes the successful reporting of others’ beliefs a complete mystery. Next I discuss Nathan Salmon’s (1986) theory of belief ascriptions which is standardly seen as being in sharp opposition to Frege’s theory. This theory will be seen as having some very counter-intuitive consequences regarding the predicted intuitive truth-values of our ordinary belief ascriptions. I will argue that such counter-intuitive results are not to be explained away as not semantically relevant. In the next section I discuss a theory which involves a radical rejection of a central assumption held in common by almost all other contemporary theories, namely the view that belief reports relate agents to a content. However, I will argue that there are considerable similarities between this view and Salmon’s, and so the theories have many of the same counter-intuitive results. The theory also has other problems of its own, again having to do with the fact that the central assumption is still retained even by this no-content theory. Finally, this

chapter discusses Mark Richard's theory of belief ascriptions, which provides an account which makes belief ascriptions context-sensitive. According to this theory, even a belief ascription which contains no obviously indexical elements, other than tense, may be true with respect to one context and false with respect to another context, at a given time. One of the main objections to the other theories considered is avoided then. Despite having this considerable virtue, Richard's theory will be seen as having serious difficulties of its own, several of these stemming from the fact that, in spite of its context-sensitivity, the theory still retains the fundamental assumption that I wish to criticise. I will explicitly contrast Richard's theory with my own context-sensitive account in the final chapter, where I will try to explain how the context-sensitivity has a very different source in my account than it does in Richard's.

Chapter Four sets out the beginnings of a positive account of the semantics of belief ascriptions. It begins by providing some desiderata for a positive account, and by giving a description of the general kind of semantic project that I see myself as engaged in. The next section provides what I consider to be a useful way of thinking of belief ascriptions; I suggest that belief ascriptions should be seen, not as general bare descriptions of an agent's inner mental state, but essentially as an answer to a question that arises in a certain context for a speaker with a specific interest and background knowledge. It is only in relation to such a question that a belief ascription can properly be assessed for truth or falsity. I use such a consideration to justify a rejection of Kripke's puzzle in one of its forms.

I provide a brief description of the metaphysical presuppositions of my semantic account of belief ascriptions, which involves a discussion of an agent's behavioural, cognitive and other dispositions and capacities. I explain the relevance of such dispositions to the semantics of belief ascriptions. I explain what I consider to be the two separate sources of context-sensitivity of belief ascriptions. One has to do with the focus of interest of the speaker who poses the question to which the belief ascription is an answer; the other has to do with background assumptions that the

speaker asking the question has. I provide a statement of the truth-condition for a belief ascription of the form: 'A believes that S' is true in a context c iff

The theory is applied to several examples which I discuss in detail. Each example is designed to bring out a different fact about the theory of belief ascriptions. The examples are designed to show, among other things, how varying the interest of the audience to the report is semantically relevant; how varying the background knowledge of the audience is semantically relevant; and explaining the (mistaken) appearance of the context-*insensitivity* of belief ascriptions. In the final section I discuss the so-called *de dicto/ de re* distinction, suggest that it is generally misconceived and that the way it is generally conceived suggests a false dichotomy which obscures how belief ascriptions really work. I suggest that *de re* reports are not as common as generally thought, and deal with several arguments which suggests that they are an important form of report. I explain why this does not mean that all belief ascriptions are therefore *de dicto*. I think that very few are *de dicto*, at least as this is traditionally understood.

Chapter One: Kripke's Puzzle About Belief

1.1

Introduction

This chapter examines a well-known 'puzzle' with which, it has been said, any satisfactory account of belief and belief reporting must come to grips. It is also impossible to ignore this puzzle since it looms so large in the literature on the semantics of belief reporting; it often plays the role of a test case for one's theory of belief and belief reporting and it is often seen to support one theory rather than another if it can be shown that one theory deals with the puzzle better than another. My own interest in the so-called 'puzzle about belief' resides largely in the fact that I think by examining the derivation of the puzzle we are able to see what are the fundamental assumptions that lie behind it, assumptions which tend to be shared also by the rather diverse responses to the puzzle that have appeared in the literature. One such assumption is that *belief ascriptions report on some single psychologically specifiable fact about an agent which is ontologically independent of the report and independent of other such psychological facts about the agent* and this is the central thesis that this dissertation aims to cast doubt on. This chapter aims, in Sections 1.2-1.4 to examine the derivation of Kripke's puzzle and show what is relevant and what is not strictly relevant to the puzzle. Section 1.5 tries to expose a possible flaw in Kripke's reasoning and 1.6 suggests what picture of the relation between belief and belief ascription is behind this reasoning. Sections 1.7 and 1.8, and 1.9, suggest a response to Kripke's puzzle which abandons the assumption that was identified as lying behind the puzzle in Section 1.6. I hope, then, to make some positive remarks in this chapter as how we are to respond to the puzzle in a natural way. And I will offer some considerations to support the kind of solution that I suggest. I will, in Section 1.10, briefly show how it fits into what has come to be seen as the logical space for possible responses to the puzzle and how it contrasts with other popular responses to the puzzle.

1.2 How Kripke Derives the Puzzle

What exactly is the puzzle and what should its import be taken to be? These questions are pertinent because the puzzle occurs in a paper, Kripke's 'A Puzzle About Belief', which has a philosophical agenda of its own and many philosophers have responded to the puzzle as a defence of this agenda, specifically the defence of the claim that belief contexts are transparent with respect to proper names. Kripke himself says in the paper, however, that "as a philosophical puzzle it stands on its own". And it is from this perspective that I want to consider the puzzle. I first want to understand what the puzzle is in its own terms. This is what this first section is about.

Schematically, the puzzle comes about like this. Kripke provides us with a scenario which involves a description of an agent, his physical environment and his linguistic behaviour, that is to say his utterances and his assents to certain sentences. The description is, in itself, taken to be unproblematic in the sense that it is wholly consistent, not too improbable and not too hard to imagine. We are then invited to consider what we might say about what this agent believes. Or rather, we are first invited to agree that our ordinary belief reporting practices ordinarily embody certain principles. Then arguments are constructed to show that if we apply such principles to the agent in question then we are inevitably led to say things about what the agent believes which are either intuitively false or, even worse, paradoxical, perhaps even flatly contradictory.

The scenario is this. Pierre is a normal French speaker who lives in France and who speaks French, but not a word of any other language, including English. He has heard of London, but only by hearing of it under its French name 'Londres'. On the basis of what he has heard of it, he is inclined to say, in French, "Londres est jolie". We are to suppose that Pierre satisfies all the criteria (whatever they may be) we ordinarily use to judge that a Frenchman uses 'Londres' to talk about London and uses 'est jolie' to attribute prettiness to a thing. Later Pierre

moves to London, to an unattractive part of the city where his neighbours are fairly uneducated, do not speak a word of French, and rarely leave their part of the city. He learns English directly by mixing with his monolingual neighbours, and learns to call the city he now lives in 'London'. He does not make the connection between 'London' and 'Londres'; that is, he does not come to realise that the names are codesignative.

Based on what he hears from his neighbours and what he sees around him, he comes to be inclined to assent to the English sentence 'London is not pretty', but has no inclination to assent to the sentence 'London is pretty'. He has no inclination to withdraw his assent from the French sentence 'Londres est jolie', having not made the connection between the two names. Pierre satisfies all the criteria that we would ordinarily use to judge that an Englishman uses 'London' to refer to London and uses 'is pretty' to attribute prettiness to a thing.

I have presented the scenario in a way which leaves out some things which we could still, apparently fairly unproblematically say about the scenario, things which involve statements about what Pierre believes. I do this because I want to capture all the salient facts that go into describing Pierre's scenario which remains as close as possible to all that Kripke includes, except that I want a *pure* description in the sense of one that does not risk prejudging anything about what we might say that Pierre believes. For Kripke himself slips in this: "On the basis of what he has heard of London, he is inclined to think that it is pretty". We lose nothing important by omitting this at this point.

Kripke then adds on top of this *pure* description, the stipulation that Pierre is presumed to be rational. This stipulation also appears to be unproblematic in that its truth is consistent with the pure description given so far. Rationality is, importantly, not defined here, but there is supposed to be a clear intuitive sense in which we would be happy to say that Pierre is rational, based on what we have so far. (The relevant concept of rationality will be discussed later, in Section 1.9)

As I understand it, Pierre's presumed intuitive rationality forms a central element in the sense of paradox which Kripke wants to derive.

The other central element in the sense of puzzlement has to do with what we say that Pierre believes, specifically involving certain belief ascriptions of the form 'A believes that S'. Kripke argues that there are some very plausible and intuitive principles that our ordinary practice of belief reporting includes that lead us to say something apparently in tension with the assumption of Pierre's rationality. The principles in question are the Translation Principle (henceforth 'TP') and the Disquotation Principle (henceforth 'DP') which can be stated as follows:

(TP) If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language).

(DP) If a normal English speaker, on reflection, sincerely assents to 'p', then she believes that p.

In (DP), 'p' is to be replaced inside and outside the quotation marks by any appropriate standard English sentence.

Both of the above principles must, Kripke insists, be qualified so as to be strictly true, but he insists that taken in their obvious intent, they seem to express almost self-evident truths. The argument taking us from the facts involved in the pure description of the scenario I presented above to apparently puzzling belief reports (puzzling, that is, when held in conjunction with the assumption of Pierre's rationality) then goes as follows:

Pierre assents first to 'Londres est jolie'. He counts as a normal French speaker, and we can assume that his assent to this sentence is reflective and sincere. We can apply a French version of (DP), for if there is such a true principle for English, there must surely be one for French speakers as well. Applying the French analogue of (DP) we are entitled to infer the truth of the sentence 'Pierre

croit que Londres est jolie'. We then use (TP) to infer the truth of 'Pierre believes that London is pretty'. We have an argument then that takes us from the pure description of the scenario, using only very plausible principles taken to be embodied in our practice of belief reporting, to the conclusion that

(1) Pierre believes that London is pretty.

Pierre later assents to 'London is not pretty'. It is assumed that he counts as a normal speaker of English; at least, it is stipulated that he satisfies all the normal criteria for counting a native English speaker to count as a normal speaker of English. We can assume that his assent to the sentence 'London is not pretty' is sincere and reflective, and so we can apply (DP) to derive

(2) Pierre believes that London is not pretty.

The puzzle stems from the fact that we are to suppose that (1) and (2) are true together. This is important, for the same puzzle cannot be derived if we allowed that (1) and (2) were both true, but not true together, if, for instance, they were true at different times. For in that case we could simply say that Pierre had changed his mind.

Kripke makes a superficially plausible case that (1) and (2) are indeed true together. He argues that it is unquestionable that Pierre once believed that London is pretty; he differed not at all from his countrymen in any relevant observable way and so we would have exactly the same grounds to say of him that he believes that London is pretty; nor, Kripke adds, is there any plausibility in the idea that we should retroactively judge Pierre never have to believed that London is pretty, for do so would endanger our belief ascriptions to all monolingual Frenchmen. Kripke also argues that Pierre still believes that London is pretty since he has not changed his mind, since he still forcefully assents to 'Londres est jolie'. (2) cannot be denied, Kripke insists, just because (1) is true. He says that we cannot deny him his later belief just because of his French past. In support of this, Kripke says that if an electric shock had wiped out all French memories he would be indistinguishable from his English

neighbours in all relevant observable ways, including his verbal and non-verbal behaviour. He adds that it surely has no plausibility to assume that an electric shock, which destroys part of his brain, gives Pierre new beliefs. His French past is therefore irrelevant, says Kripke. We cannot deny Pierre both his beliefs for this combines the difficulties of denying (1) and denying (2).

It seems that we are forced to conclude then that (1) and (2) are true together. If they are true together then that undoubtedly entails that (3) (below) is true, since this simply involves an application of the law of conjunction to (1) and (2). It might also be supposed that (4) (below) is true, although this does not logically follow, unless we assume that belief factors out of a conjunction, but this is generally assumed and nothing important rests on the distinction between (3) and (4) in the sense that the same problems arise whether we take (3) or (4) as the ascription we are warranted in asserting.

- (3) Pierre believes that London is pretty and Pierre believes that London is not pretty.
- (4) Pierre believes that London is pretty and that London is not pretty.

Given the truth of (3) or (4), it seems natural to say that Pierre has contradictory beliefs. In fact, Kripke does not distinguish between saying that (3) is true and saying that Pierre has contradictory beliefs. And I suspect that this is why he thinks that we have a puzzle about *belief*. This is important because sometimes it is said that we do not, strictly speaking, have here a puzzle about belief, but rather a puzzle about belief *ascriptions*. From a formal point of view, this is a distinction that certainly can be made and as we shall see later (in Section 1.6) this alleged distinction has been put to philosophical work in several types of response to the puzzle. My first substantial point, however, which I shall defend later, is that I think that the distinction is unnatural and positively misleading. For, what could be more natural than saying that having a belief report of the form exhibited in (3) is true of an agent is just the same as saying that an agent

has contradictory beliefs. I think that it does not matter whether we see the puzzle as a puzzle about belief or as a puzzle about our talk about belief.¹

The puzzle can then be put in the following way. We seem to have compelling grounds to say that Pierre has contradictory beliefs and we seem to be unprepared to say that Pierre is not rational. Yet there seems to be a conceptual connection between rationality and belief which does not allow that a rational agent can have contradictory beliefs. Kripke says:

“Pierre is a leading philosopher and logician. He would never let contradictory beliefs pass. And surely anyone, leading logician or no, is in principle in a position to notice contradictory beliefs if he has them. Precisely for this reason we regard those who contradict themselves as subject to greater censure than those who merely have false beliefs. But it is clear that Pierre is in no position to see, by logic alone, that at least one of his beliefs must be false. He lacks information not logical acumen. He cannot be convicted of inconsistency: to do so is incorrect.”

Kripke does not put forward a resolution of this puzzle. He simply raises the puzzle and forestalls some possible natural responses. In particular, he makes a great deal out of the fact that the situation can be completely described in a consistent way and that there, as he says, “in this sense no paradox”. This way, however, evades the question of whether Pierre believes that London is pretty. In fact, Kripke says that the puzzle is to answer the question ‘Does Pierre or does he not believe that London is pretty?’

But, nothing, I think, could be more important at this stage, than the observation, that this puzzle is, strictly speaking, a distinct puzzle from the puzzle that seemed to emerge earlier and what characterised (two paragraphs above) as a tension between two very plausible theses.

The puzzle put in the form of the tension of two compelling conflicting claims presupposes that we have an answer to the question that the second characterisation of the puzzle says we cannot answer!

¹ The view that I want to defend downplays the distinction and suggests that to understand our ordinary notion of beliefs we look to our belief ascriptions, rather than assume we have a clear notion of what beliefs are and then build our theory of ascriptions on top of that.

To clarify, I think that there are two distinct puzzles posed by Kripke. One is how to reconcile the intuitive truth of ascriptions which seem to ascribe contradictory beliefs to Pierre with his intuitive rationality. This puzzle is dealt with in this chapter. Another puzzle stems from the fact that we apparently cannot directly answer the simple question ‘Does Pierre, or does he not, believe that London is pretty?’. This puzzle is dealt with in Chapter Four.

I want to focus on the former version of the puzzle, firstly as it has a structure which facilitates a clear discussion, but more importantly, because the second version of the puzzle has the appearance of a trick question. I mean this seriously and literally. I think that holding on to this insight is the key to a natural resolution of the alleged paradox. In itself, the fact that we cannot answer a question of a certain specific grammatical form (in this case a question of the form ‘*Does A or does he not believe that S?*’) with an answer of a specific grammatical form (in this case an answer of the form ‘*A believes that S*’) in a non-misleading way, is not paradoxical. It is not even an unusual situation to find oneself in. (I give a more detailed response in Chapter Four).

I have described the derivation of Kripke’s puzzle in a rather conventional way; I have said that it relies on an argument which involves principles that Kripke thinks our ordinary belief reporting practice embody. But now I want to look more closely at the structure of the argument and try to see what is really absolutely essential to the derivation of the paradox, as I think this is the way to see what the most basic assumptions presupposed in the derivation of the puzzle are. I want to make it seem plausible that the argument that Kripke actually uses, understood in the way I have characterised it so far, is too specific for us to see what the assumptions are that we must ultimately reject if we are to have a natural resolution of the puzzle. I see this strategy as basically a reapplication of a strategy that Kripke uses himself in the paper. For Kripke argues that trying to show that (TP) is false is not the right way to tackle the puzzle about belief, for (TP) is not actually essential in the argument at all, even though some philosophers have continued to blame the paradox on (TP), for example, D.

Over (1983). He does this by showing that a strictly analogous puzzle can be derived from a description of a scenario which only involves one language being spoken so that a principle can have no application here. I want to reapply this strategy by showing that a strictly analogous puzzle can be derived even without the use of either (DP) or (TP). If this can be done successfully then it would of course show that a fully satisfactory solution to the puzzle cannot be one which works merely by showing that (DP) is false. (Let me make it clear that I do think, nonetheless, that (DP) is in fact false and I will present reasons for thinking this in some detail later on because seeing why (DP) is false can further illuminate what is being presupposed by Kripke at the deepest level.)

1.3

The Irrelevancy of the Disquotation Principle

I believe that we often do ascribe beliefs without relying, even tacitly, on anything like a disquotation principle, and, moreover, that an analogous puzzle can be created using an argument based on such beliefs. I shall now describe a scenario in which we would feel justified in ascribing a certain belief to Pierre where Pierre does not provide us with an utterance which allows us to disquote him.

Suppose that Pierre is at home watching television and sees a documentary featuring London and that he seems very impressed with what he sees on television (images of London's lovely green parks, the Thames, St. Paul's Cathedral.) He says: 'Quelle belle ville!'. If we know French we would surely feel justified in saying, on the basis of his utterance: 'Pierre believes that London is beautiful'. I take it that someone who objected to such an attribution, saying, perhaps, that we should instead say that 'Pierre believes that the city he sees on the television in front of him is beautiful' thinking that this is a more accurate description of Pierre's belief than the ascription 'Pierre believes that London is beautiful' has some theoretical considerations in mind.² I wish to

² A Fregean, for instance, might be tempted to make such a claim if he thought that belief reports must relate agents to thought-contents which contain descriptive senses and that such senses must appear in the embedded clause of the report if it is to be strictly true.

emphasise however, that it would be absolutely natural to say that he believes that London is beautiful and I take it as a datum that we would say such a thing without taking ourselves to be saying anything inaccurate, non-literal or elliptical.

It seems, furthermore, that we could make such an attribution on the basis of other purely non-verbal behaviour. For Pierre could have given some other sign to show that he believed the same thing. Facial expressions and other bodily movements can often guide us to what a person thinks. Further support for the claim that we often ascribe beliefs without relying on a principle of disquotation comes from the observation that we ascribe beliefs to those who have not spoken on topic, perhaps because they can't. We are often willing to say of a dog who has been chasing squirrels across the lawn and then suddenly stops and barks furiously up a tree, for example, that it believes that the squirrel ran up the tree, or of a small child who we see has got separated from her father in a supermarket that she believes that she has been abandoned. It is expected that one might object to taking these kinds of belief ascriptions seriously and that they may be theoretical reasons for doing so. On such a view, the ascriptions are to be treated as useful, but non-literal ways of talking about beliefs. It is of historical interest to note here that Frank Ramsey allowed that we might ascribe beliefs to non-verbal animals and that we might be speaking non-metaphorically in doing so, but that we must then admit that the term 'belief' is semantically ambiguous and that in his philosophical work on belief he was only interested in beliefs of linguistically sophisticated animals. Both types of responses bring in theoretical considerations to partition off the belief reports which should inform a proper philosophical study of belief reports. It would be too quick to say that this is methodologically improper, but I suggest that we gain real insight by taking the data seriously and only trying to develop our theory of belief and belief reporting later.³ The data in this case are that we often issue belief reports about animals and non-verbal humans.

³ I do not think that we want an a priori argument for this methodological claim here. Rather, the rightness of the claim can only be judged in light of the success or otherwise of a project, such as mine, which employs this methodology.

If we take seriously such belief attributions not based on verbal behaviour, there may still be a temptation to say here that such behaviour would necessarily underdetermine exactly what belief we are entitled to ascribe. I suggest that it is some such consideration which lies behind Kripke's use of the disquotation principle, because if we are using the principle then that would seem to fully determine the belief held. For, on this picture, if the agent says S then we say that she believes that S. On this traditional view, the content that she asserts is taken to be the very same thing that the belief report is taken to relate her to. This is why the question of underdetermination cannot arise here. I want to draw attention to the way in which such an objection as the underdetermination objection only seems to even make sense on the assumption that there is a fact of the matter about what content an agent is psychologically related to and that the belief sentence aims to specify that content.

It might also be thought that such ascriptions based solely on non-verbal behaviour are necessarily provisional in the sense that further evidence could lead us to revise our view that the original ascription was true. But, while this may be true, it is also true that *any* belief ascription is provisional in this sense, even one based on assertion of or assent to a particular sentence. But this point is, I think, often obscured by the way in which (DP) is often qualified. For the antecedent of the conditional principle says: 'If a competent, sincere and reflective speaker...' And this can make it seem, as it has seemed to many, virtually a truism. But, it is important to note that in any actual application of the principle, we do not, and cannot, know that the agent is fully competent, is trying to speak the truth and is taking a certain minimal expected amount of care over what she says. The assumption that these conditions are fully met in any real case is always defeasible. The difference between belief reports based on utterances and belief reports based on only non-verbal behaviour with respect to whether or not they are revisable in light of further evidence should not be over-emphasised. My point is that in both kinds of cases we are entitled to have confidence in the ascription at least until we have reason to attribute another belief which is a contradictory of the belief originally ascribed, or is, say, a better explanation of the agent's behaviour. So, in this respect, the situation with

respect to the two kinds of reports is parallel. The difference is, at most, a matter of degree.

I would like to say then that our intuitions about what could be the case about an agent's beliefs does not depend upon our assuming a principle like (DP). Even if my arguments supporting this claim have been accepted, it must be recognised that I have still not shown that Kripke's puzzle can be derived without such a principle. For it is conceivable that while we often make attributions without relying on (DP), there is something special about the kinds of scenarios that are involved in generating the paradoxes that means that in these cases we must rely on (DP).

I think that there is something very importantly correct behind this doubt, and I shall return to this shortly. But it does seem that we can generate a puzzle without using (DP). For imagine a story such as the one Kripke tells about Pierre which is different in that Pierre does not or is not heard to make any proclamations about London, but that he exhibits non-verbal behaviour which allows us to feel justified in saying, when thinking about the first half of his story, 'Pierre believes that London is pretty' and, when thinking about the second half of the story, allows us to feel justified in saying 'Pierre believes that London is not pretty'. Can we imagine such a story? If we can, then we can simply say that (DP) is not essential to the derivation of the puzzle, and that we must look elsewhere for the source of the trouble. If we cannot, then this must be because the difference in defeasibility of the two types of evidence for the reports (one an assertion that appears to be sincere and based on competence and sufficient reflectiveness, the other based on purely non-verbal behaviour) does play a significant role. It would be as if the evidence which is an assertion is not outweighed by other considerations deriving from our overall view of Pierre's situation, whereas the evidence for a belief attribution which is solely non-verbal behaviour is outweighed by such considerations. I see no reason to privilege verbal behaviour to this extent. (Discussion in Chapter Four will show the limited relevance of verbal behaviour to belief ascriptions).

I think that someone who accepts the reasoning that Kripke employs taking us from a pure description of Pierre's scenario to a paradox would be very likely to accept similar reasoning that takes us from a description of the scenario to a paradox but which does not involve the use of (DP). If this is correct, then in this sense (DP) is inessential to Kripke's puzzle.⁴ Before I say what I think really is essential to the derivation of the puzzle, however, I want to be more clear about what the status of (DP) really is. This is the only way to see what the underlying structure of Kripke's puzzle really is. Kripke and others seem to see (DP) as expressing some kind of conceptual truth, provided that it is suitably qualified. It is not claimed that (DP) provides a reductive analysis of belief in terms of assent to a sentence, since it seems clear that the qualifications contained in the antecedent of the conditional principle themselves involve the concept of belief. Nevertheless, it has been assumed that belief and assent are conceptually connected in a way that satisfies the principle. It is natural to think of the principle as expressing a conceptual truth. Consider the qualifications employed in the antecedent of the principle; the speaker must be linguistically competent, sincere and reflective. Being linguistically competent means having enough true beliefs about the English language to be able to express one's beliefs in English. Sincerity entails the desire to speak only in accordance with one's beliefs. Reflectiveness requires that the agent makes sure that what she says does in fact express her beliefs. Given these qualifications, then, there seems to be no room for a counter-example to the principle.

I think that this is why Kripke and others regard (DP) as a self-evident truth. But if we think that this is a principle that we actually employ, then it cannot be an analytic truth, for as I have said earlier we could never know that the conditions in the antecedent are fulfilled in any real case. Taken as an analytic truth, we should challenge whether it should be taken to be applicable in any real case. Taken as a principle that we actually employ to arrive at what an agent believes, we should understand the principle as embodying the idea that there is an

⁴ Prominent examples of philosophers who seem to think that the paradox can arise with the use of (DP) are Keith Donnellan (1979) whose paper 'Belief and the Identity of Reference' seems to argue that a paradox can arise without (DP), and Genevieve Marti and Joseph Almog, both of whom have verbally expressed such a view to me.

evidential relation between assent and belief, rather than the tight *conceptual* connection that is widely believed to hold.

When we think of (DP) as a principle of evidence that we actually use, then we are more likely to think of assent to a sentence as defeasible evidence. Since the principle can easily seem to be self-evident, I think it is useful to have in mind an example where assent to a sentence is at best a privileged guide to what an agent believes, but certainly not an automatically overriding form of evidence. A man who works in a residential home for the old, when being asked about why he chose his career, says 'Old people deserve respect.' Suppose that in saying that the man seems sincere, and that we generally think of him as being sincere; that he seems to be reflective and that he seems to use the word 'deserves' as all other competent English speakers. Suppose that his behaviour seems to us very disrespectful so that we might say that his actual treatment of the elderly gives the lie to what he says. What are we to make of this? We might very naturally refuse to say that he believes that the elderly deserve respect, for if he really thought that then he just couldn't treat them the way he did. Are we then forced to say that he is either being insincere, unreflective or linguistically incompetent?

My intuition is that we are not forced into saying any one of these. One could say one of these things, or one could talk of self-deception, false consciousness. I do not know what we should say in such a case; whether we must choose one of the above options or whether these options do not exhaust the possibilities. Nor am I sure that an analytic philosopher bent on defending (DP) could not show either that one of the conditions met in the antecedent was not in fact met or that further restrictive conditions could be included in the antecedent so that it was made immune to counter-example. The main point of my example is to encourage us to think that in realistic cases, the most natural thing to say is that behavioural evidence is not superfluous to our judgements about what an agent actually believes, and that assent to a sentence is only one form of evidence that counts towards our judgements.

If neither (TP) nor (DP) is strictly essential to the derivation of the puzzle, then the question is; what *is* at the core of the puzzle? This is the question that I now try to answer.

1.4

The structure of Kripke's argument

First I want to draw attention to the fact we need some *argument* involving some specific assumptions about belief reporting in order to get as far as making the ascription : 'Pierre believes that London is pretty and that London is not pretty'. Many thought-experiments in philosophy of mind or language having to do with belief ascriptions simply start with a stipulation about what an agent believes. In discussions of substitutivity, for instance, it is simply assumed that an agent can believe, for instance, that Hesperus is shiny and Phosphorous is shiny. Intuitions about the possibility of an agent having such beliefs do not seem to depend on any specific assumptions about an agent's assertions or other behaviour. It is supposed that an ascription like 'John believes Phosphorous is shiny and that Hesperus is not shiny' is not considered, pre-theoretically puzzling. It only seems to become problematic when we make theoretical assumptions which seem to commit us to a principle of substitutivity, for then we are led to say that if this last ascription is true then so must this one be true: 'John believes that Phosphorous is shiny and that Phosphorous is not shiny' and this ascription does seem puzzling. My point here is that this ascription and the structurally similar one 'Pierre believes that London is pretty and that London is not pretty' seems immediately puzzling. This form of ascription just does not seem to be one we can make sense of straight off. This is why Kripke needs an *argument* to derive the claim that Pierre believes that London is pretty and that London is not pretty'.

The question then becomes: What is the argument that would establish such a conclusion if it does not necessarily involve either (DP) or (TP)? I shall try now to say what I think the underlying argument structure is, before defending this construal and, then saying what picture of belief and belief reporting lies behind it.

An argument like Kripke's occurs in two parts, and this is important⁵. First, we consider an agent and we are encouraged to think that we have sufficient grounds for attributing to him the belief that London is pretty. The grounds here (which may include sincere assertions, but need not) are considered sufficient to establish that the agent has that belief. Importantly, the grounds have only to do with what we learn about the agent in the first part of the story. Next, we consider the agent with respect to another context, and we are encouraged to think that we again have sufficient grounds for ascribing to him a certain belief, namely, the belief that London is not pretty. We are invited to consider various facts about the agent, but we are required, it seems to me, not to consider the earlier facts which were taken to provide the grounds for the claim that he believes that London is pretty. In particular, we are not to consider the fact that we earlier decided to say of him that he believes that London is pretty. The grounds underlying the first ascription and the grounds underlying the second ascription are taken to be *independent* of each other. The argument requires that they are independent in the sense that each of the set of facts constituting the grounds for each ascription is simply irrelevant when it comes to deciding the truth of the other ascription. This independence is required for each of the grounds to be sufficient grounds for each ascription. The attribution which says 'Pierre believes that London is pretty and that London is not pretty' does not seem to be derivable if we let each of the grounds interfere with the other. In other words, once we are put in a position in which all the facts are considered together, neither of the ascriptions seem particularly plausible. But if they are both implausible considered individually then we lose our reason for saying that the conjunctive report is true. For, given the observation made in the previous paragraph that the conjunctive report is *prima facie* puzzling, we would have to derive the conjunctive report from the two other reports.

This, I think, is the basic structure of the argument for the conclusion that 'Pierre believes that London is pretty and that London is not pretty' is true. It

⁵ Keith Donnellan in his paper 'Belief and the Identity of Reference' (1979) is, as far as I know, the only philosopher who has drawn attention to this very important structural aspect of Kripke's derivation of the puzzle. I believe that this structure is crucial, as I explain.

does seem absolutely essential to the puzzle that we are to consider the scenario in two parts which we the reporters must keep apart when considering our intuitions about what to say about Pierre's beliefs. And it seems essential that the factors we consider are sufficient to determine what Pierre believes, otherwise our acceptance of each of the individual reports would not hold. It is this requirement of sufficiency, I suggest, that lies behind Kripke's insistence on using (DP) and his insistence in claiming that (DP) is conceptually true. But it is not (DP) per se that is essential to the derivation, but the notion of sufficient grounds. I claimed that the idea of sufficiency relevant here entails that, for both of the reports, considered individually, the truth of the other report and the allegedly sufficient grounds for it must be independent. I take it that this idea of independence lies behind what Kripke is arguing for when he argues 1) that we cannot retroactively judge that Pierre never believed that London is pretty because of 'later vicissitudes' (thus showing the independence of the first report from the grounds for the second) and 2) that we cannot withhold our attribution of Pierre's belief that London is not pretty on the basis of facts deriving from his earlier and quite separate life in France because Pierre is indistinguishable from his English neighbours (thus showing that the later facts are independent from the grounds for the first ascription.)

In summary, the paradox arises because we have two sets of facts both of which when we consider them each by itself seem warrant to a certain conclusion about what Pierre believes. It is further made to seem that each set of facts exhausts all the relevant information about Pierre's beliefs, so that each set facts is independent from the other. The conclusions which seem to be independently warranted, however, seem highly problematic when considered together when we hold the additional claim that Pierre is rational.

1.5

Exposing A Possible Flaw in Kripke's Reasoning

I now want to focus more closely on the structure of Kripke's argument and to try to find a fallacy in the argument. It seems to me that there is a fallacy in the argument, but that it is almost invisible from a certain perspective, specifically

from a perspective which involves talking about Pierre's beliefs rather than talking about which ascriptions are true of him.

Let us focus on Kripke's argument for the claim that we are bound to say that Pierre believes that London is pretty. The argument can be reconstructed as follows:

1. We would say, truly, given the information about Pierre's life in France, and given no other information, 'Pierre believes that London is pretty'.
2. We should not retroactively judge that Pierre never believed this.
3. It would be wrong to say that Pierre has changed his mind.

Therefore,

4. We must now say that 'Pierre believes that London is pretty' is true.

I think that this is a fair reconstruction of the reasoning that Kripke employs. Moreover, stating the premises and conclusion in a metalinguistic mode is designed to bring out the fallacy in the argument. Stating the argument in the metalinguistic mode also has the advantage that the premises are all plausible even once the whole argument of Kripke's paper has been given. (If the argument had instead been from the premises 1') Pierre believed that London is pretty and 3') Pierre has not changed his mind to the conclusion 4') Pierre now believes that London is pretty, then the argument is less plausible as the first premise would be rather certain, given the paradox eventually arrived at). The premises all appear, then, to be true. Premises 1 and 3 certainly seem beyond doubt, although 2 may give us some pause. But, Kripke does say in support of this claim that if we were to allow such retroactive judgements then all our belief attribution would, as long as the future is uncertain, be put in danger. This in itself is not a convincing reason to ban retroactive judgements, since surely all our empirical claims are defeasible in light of what we may learn in the future. So I assume that Kripke must mean that only certain kinds of future findings can be considered relevant for the truth of our belief attributions and that these kinds do not include the kinds of findings that we have presented to us in the description of Pierre's situation. The ban on retroactive judgements about belief attributions is to be understood in a sense in which it would follow from the

claim that the information we are given about Pierre in the first part of the story exhausts what is relevant for the truth of the ascription 'Pierre believes that London is pretty'.

I assume that the premises can be understood, then, in a way which makes them all true. But is the argument valid? It seems to me that it is not, for it seems clear to me that the conclusion is false. At least, it seems clear to me that we would not find it at all natural to say, when we have the whole story before us, that Pierre believes that London is pretty. One might think that it would be better to be more tentative here, and instead of saying that it is false, to say that it feels like a strange, misleading thing to say. It is, of course, not unusual in philosophy of language to find philosophers saying things which fall short of attributing a truth-value to a particular statement. But I see no good reason not to say that 'Pierre believes that London is pretty' is not true, no good reason, that is to say, in the absence of an argument such as the one considered here. I will now focus on why I think that this argument is invalid.

It obviously does not follow from the fact that it would be correct to utter a sentence in a certain context that it would be correct to utter that sentence in a different context, even if it were also true that the state of affairs allegedly described the first utterance had not changed between the two contexts of utterance. To make such an inference would obviously be fallacious in the case of utterances involving the explicit use of indexicals such as in the following argument:

I believe it's going to rain (spoken by Gary)
 Gary has not changed his mind about this
 Therefore,
 I believe it's going to rain (spoken by Leah)

This might seem an irrelevant counter-example involving explicit indexicals as it does, although it should be noted that there is however a sizeable amount of recent work on the semantics of belief reports according to which belief

ascriptions involve tacit elements of indexicality⁶. I wish to dissociate myself from such accounts however and I will explain why in the second part of this chapter. For now I wish to point out that Kripke's argument as I have reconstructed it above is fallacious and suggest that this is because the context of utterance of a belief ascription can be relevant to the truth-value of a report even when the report does not exhibit explicit elements of indexicality. In fact, I do not think that the contextual-sensitivity of belief reports has anything to do with indexicality as this notion is usually understood. (My account of the contextual sensitivity of ascriptions will be explained in chapter Four).

Anyone who believes that the argument I reconstructed above is fallacious for something like the reason I have indicated can be said to hold a 'contextualist' thesis about belief sentences.⁷ 'Contextualism' can be defined thus:

It is possible that a belief-sentence containing no obvious indexical expressions is true of an agent relative to one context, and simultaneously false of the agent with respect to another context.

Contextualism is certainly an interesting thesis and does not seem to be among the possibilities that Kripke considers when addressing the question of how the puzzle might be solved. I now want to explain why contextualism did not, I think, occur to Kripke.

1.6

Foregrounding the most basic assumption

I think the most plausible explanation is that Kripke presupposes a certain analysis of belief reports sentences as well as a certain relationship between a belief report and belief itself, (not that this assumption entails, by itself, that contextualism is false, since as I will mention below, contextualism can be incorporated into theories which maintain the basic assumption). Although he

⁶ Crimmins (1992) and Crimmins and Perry (1989) are prominent examples of a so-called hidden-indexical theory, as is Schiffer (1977). I do not discuss these theories in the dissertation, but note that they each embody the central assumption that I want to cast doubt on, and do so in a particularly tendentious way.

⁷ Joseph Moore (1999) has used the label 'contextualist' to describe such views, and my definition of it is derived from his.

does not explicitly distinguish between the psychological phenomenon of believing and the semantic issue of belief ascriptions, it seems to me that he does presuppose that belief ascriptions are true (when true) because they explicitly relate an agent to a representation which the agent is as a matter of (psychological) fact related to. My reason for saying that Kripke presupposes these two things is that this seems to be the best explanation of why he employs the form of argumentation that he does employ. Specifically, consider the point in the argument where Kripke gives us a reason for thinking that the conjunctive report 'Pierre believes that London is pretty and that London is not pretty' is true. He cannot simply ask us to consult our intuitions about what we would say using such a form of sentence, given all the information about Pierre. For then we would certainly not want to issue or endorse the conjunctive report. He is instead forced to proceed indirectly. He says he has established that Pierre believed that London is pretty and that Pierre has not changed his mind. He then concludes that Pierre still believes that London is pretty. Now, in this formulation, the reasoning seems very plausible. Notice, however, that if we put the same argument in the metalinguistic mode, it immediately seems less plausible. If we move from (1) 'Pierre believes that London is pretty' was true, together with (2) 'Pierre has not changed his mind' is true, to (3) 'Pierre believes that London is pretty' is now true, it seems to me that we now feel less sure that we have arrived on firm ground. Why should this be the case?

I think that the plausibility of the first argument derives from our tacit assumption that we report on a fact about Pierre which is (ontologically) prior to our report, namely, the fact that Pierre stands in a certain relation to a particular proposition, specifically the proposition expressed by 'London is pretty'. Pierre has not changed his mind. This is assumed to mean that he has not ceased standing in that same relation (of acceptance) to that same proposition. We feel bound then to say that he still stands in that relation to that proposition. And this fact (again thought of as a real fact about Pierre which is independent of what we would naturally want to say about him using belief sentences of a certain form) is thought to ground our saying that Pierre still believes that London is pretty. The same picture seems to be at work when Kripke tries to establish that Pierre has later come to believe that London is not pretty. Pierre is then thought

of as standing in a real psychological relation to the proposition expressed by 'London is not pretty' which is the contradictory of the one expressed by 'London is pretty'. If we think of these facts as prior to the reports, then it can seem that the facts are independent of each other. This is why it seems metaphysically possible that Pierre can have contradictory beliefs.

By contrast, the argument formulated in the metalinguistic mode allows us to focus our attention on what we would naturally want to say to describe Pierre's situation and this allows our intuition that belief reports are context-sensitive to come to the surface. Also, from this metalinguistic perspective, it seems that the reports are not likely to be independent of each other. While the two reports 'Pierre believes that London is pretty' and 'Pierre believes that London is not pretty' seem natural when considered individually, their conjunction seems anything but natural. It seems, to naïve intuition, to be the kind of thing we would never say.

Nevertheless, the presuppositions in question (that belief reports report on a prior fact about what proposition an agent is related to in the relevant way) do not, by themselves, entail that contextualism is false. The only way that contextualism could be true, however, given these assumptions, is if the overall semantic contribution of an utterance of the subsentence S, in the context of the report, itself differed from one context to another. Kripke, of course, did not think that this was the case, thinking that 'London is pretty' always expresses a univocal content to which the belief report reports Pierre as being related to in the relevant way. And this is the very assumption that recent work in the semantics of belief reports has been aimed at criticising. (I will critically examine one such theory in Chapter Three). The theories which do this are rather diverse from a purely semantical perspective, but they all have in common that they try to account for contextualism with respect to belief reports by showing how certain features of the context of utterance of the belief report contribute to what is said in uttering it on a given occasion by altering the entity that the agent is said to be related to by the relevant acceptance relation. What is left in place are the two assumptions that I mentioned above, namely, that belief reports report on a prior psychological fact and that this prior fact consists of an

agent standing in a certain relation to a representational entity. I do not want to criticise these views here. I merely want to suggest that this may be an alternative way to account for contextualism. The remainder of this chapter tries to suggest how an alternative picture might address the puzzle about rationality and belief reports.

1.7

Desiderata for a Resolution of the Puzzle

I am now in a position put forward what I consider to be the main desiderata for a resolution to the version of Kripke's puzzle that I have focused on in this chapter (that is, concerning the relation between Pierre's rationality and the argument which leads to his being attributed contradictory beliefs. The puzzle should be addressed in its both of its two guises, that is to say, both when presented as a puzzle about *belief* and when seen as a puzzle about belief *ascriptions*. I want an account which can accommodate, first, the felt incompatibility of an agent's rationality with her having contradictory beliefs, and secondly, the felt incompatibility of saying of an agent presumed to be rational that she believes 'that S and that not-S'. The intuition that a rational agent cannot have contradictory beliefs is to be taken seriously and assumed to be true (so that there is no question of a revisionary proposal according to which a rational agent can, after all, have contradictory beliefs). The account, in explaining the incompatibility, should not rely either on 1) the claim that belief reports are to be understood 'de re' so that the belief ascriptions are somehow not reflective of the real underlying cognitive facts about the agent, nor 2) the claim that the belief ascriptions are superficially misleading as to the real nature of the beliefs they allegedly talk about. There should be no reliance on the principle in philosophy of mind that says that ascribing contradictory properties to an object in thought depends on the agent have two distinct 'ways of thinking' of the object in question. There should be no reliance on the claim that our belief reports somehow manage to achieve reference to the 'ways of thinking' that the philosophy of mind principle says mediates between agents

and the objects their beliefs are about. (Arguments for these desiderata will be given in Chapter Three).

The response to the puzzle should be able to account for the apparent plausibility of Kripke's arguments at each stage, in particular, why it seems, first, true to say 'Pierre believes that London is pretty' and then later to say 'Pierre believes that London is not pretty' without it ever being right to say 'Pierre believes that London is pretty and that London is not pretty'. The account must then be contextualist (in the sense defined in 1.5), so as to avoid the radically revisionary proposal that the law of conjunction does not apply in the context of belief reports.

The account should not rely on discrediting (DP) or (TP) since both of these principles have been seen to be inessential to the derivation of the puzzle.

1.8

Suggesting A New Response to the Puzzle

Introducing a Principle Which Governs our Practices

My response to the puzzle starts from the empirical observation that the following generalisation is true of our belief reporting practices:

For any agent A whom we consider to be rational, and for any sentence S, we do not issue a report of the form 'A believes that S and that not-S'.

Postulating that our belief-reporting practices embody (PRRA) is not arbitrary. The constraint has an important function, as Joseph Moore (1999) observes⁸. It serves to distinguish between two kinds of epistemic situation which we surely do want to distinguish. On the one hand there are situations structurally like Pierre's where the agent thinks that there are two objects when in fact there is

⁸ Moore does not draw the strong conclusions about the existence of this principle that I do.

only one and so can attribute contradictory properties to the object in thought and in speech (since the agent does not realise that the names he uses are coreferential.) but where the agent's ascribing such conflicting properties arises from his false empirical beliefs. We might say that external circumstances have conspired against him. On the other hand, there are those situations of genuine irrationality which would have to arise through carelessness in thought or 'cognitive dissonance'. These involve not a conspiracy in external circumstances, but some kind of failure of internal functioning. It is plausible that our ordinary belief-reporting practice would track such a distinction, for even though both types of case would require the agent to reorganise his thoughts, in the former type of case, exemplified by Pierre, she could in principle do this by learning new facts. In the latter type of case, she should in principle be able to do this by purely a priori means.

Here are examples which I think exemplify this distinction and which are in accordance with (PRRA). Pierre is intuitively rational, he is a leading logician, and given the story about him we understand that he finds himself in a non-ideal epistemic situation due to a conspiracy of circumstances which have resulted in him not realising that 'Londres' and 'London' name one single city. We have a strong inclination not to say 'Pierre believes that London is pretty and that London is not pretty'. Kripke expresses this intuition when he says: "We cannot convict him of inconsistency. To do so is incorrect". (The weak conflicting inclination to say that the report is true seems to derive entirely from Kripke's argument where he argues that it is true that Pierre believes that London is pretty and argues separately that Pierre believes that London is not pretty. Once the contextual-sensitivity of the reports is made explicit the inference that Kripke actually makes is revealed as fallacious, since it does not follow from the fact that 'Pierre believes that London is pretty' is true in one context and that 'Pierre believes that London is not pretty' is true in another context, that there is a context in which they are both true. I submit that our practices embody a surface constraint along the lines of (PRRA)).

An example of the other type of case is one I heard in conversation recently. 'He believes that The Earth was created two thousand years ago and that it was not

created two thousand years.’ It occurred in the context of talking about a friend who seemed to have ‘compartmentalised’ his religious beliefs. The point of the report was to draw attention to the agent’s irrationality and conveyed a degree of epistemic censure which seemed appropriate because the reporter thought that the agent should surely be able to see for himself (by a priori means) that he needed to change his beliefs.

Why a rational agent cannot have contradictory beliefs

I claim that a rational agent cannot have straightforwardly contradictory beliefs because our practices of ascribing beliefs embody a constraint which prohibits the attribution of contradictory beliefs to an agent who is intuitively rational. Or at least there is a constraint that says we can never say of an agent presumed to be rational something of the form ‘A believes that S and that not-S’, without immediate further qualification and explanation. (So, I concede that we may occasionally be able to say, of Pierre, for example, ‘Pierre believes that London is pretty and that London is not pretty, but this is only because he thinks that there are two Londons, etc....’ but this addition to the ascription would be a necessary part of the context of the belief ascription).

The ‘because’ in the first sentence of this subsection is meant seriously. The coherence of this explanation depends on saying that what an agent believes, or what beliefs he has, depends on what our practices allow us to say in the form ‘A believes that S’. This is an inversion of the standard picture which I argue against, according to which what beliefs an agent has is ontologically prior to our belief ascriptions and grounds their truth or falsity as the case may be. I intend that the discussion in Chapter Three and the view I defend in Chapter Four should make this picture seem more plausible than it might otherwise be.

The Notion of ‘Rationality’ at the Centre of the Puzzle

I need to say something more about the notion of rationality at the centre of this discussion, for the Principle of Reporting Rational Agents would seem to

require (and I believe that it does) that we can have a grasp of whether or not an agent is rational which is prior to deciding whether he has contradictory beliefs or not (since I have deliberately conflated the distinction between having contradictory beliefs and having a report of the form ‘A believes that S and that not-S’ be true of one.) Notice also that Kripke’s argumentative structure also seems to depend on a notion of rationality that does not rule out *simply by definition* that Pierre has contradictory beliefs. For there really does seem to be a difficult question about whether Pierre has contradictory beliefs; it is a matter of some debate whether we, on the one hand, reject the conceptual connection between rationality and contradictory beliefs (and this option seems to presuppose that we can still make some sense of the intuitive notion of rationality even while allowing that the agent in question does have contradictory beliefs), or on the other hand, reject the principles that lead to a contradictory attribution.

I believe that there is such an intuitive notion of rationality according to which Pierre does count as rational and which does not definitionally entail that the agent does not have contradictory beliefs.

The idea of being a good reasoner is at the centre of the notion. Rationality seems to require that one is able, by introspective means, to correct any epistemic tensions that arise in one’s cognitive situations, but only where this does not essentially depend on the agent being in possession of particular empirical knowledge. The intuition that Pierre is a good reasoner then is not impugned by the fact that he ascribes in thought contradictory properties to one object, for he is in no position to see, by introspection alone, that he is ascribing contradictory properties to a single object. If we direct attention away from the alleged contents of Pierre’s mind, since what the contents of his beliefs are is the very question at issue in Kripke’s discussion, and onto how we (as outside observers of his history) understand Pierre, we can see that it is very easy to make sense of Pierre. His verbal behaviour is readily intelligible to us; we know that he assents to contradictory sentences, that he does so because he does not know that they are contradictory and that his not knowing this is not due to lack

of attention to detail, carelessness or cognitive dissonance. His actions appear coherent to us.

An agent is rational in the relevant sense if her overall behaviour is coherent, given an overall view of her situation. Pierre's behaviour is coherent from an overall perspective. The emphasis on overall behaviour is important, for in Pierre's case, his verbal behaviour, taken in isolation, appears not coherent. The notion of coherence is admittedly vague, but I think that it is already a familiar notion to us. It is clear that the notion requires more of an agent than having consistent beliefs; it requires also that her non-verbal behaviour is not dissonant with her verbal behaviour. Consider again the agent who (apparently sincerely) avows that she believes that elderly people should always be treated with respect, and yet whose behaviour in relation to them seems dissonant with these avowals. Or Ruth Marcus' case of the woman who seems, on paper, an expert mathematician, always saying 'yes' to a whole range of true sentences of arithmetic and always saying 'no' to the false ones, and yet who always counts things wrongly and who always gives the wrong change when working in her shop. Her overall behaviour is incoherent. And we would say that she is not rational.

It is important to notice that the notion of rationality suggested here avoids some of the metaphysical assumptions about mind that have been explicitly or implicitly assumed in classical responses to Kripke's puzzle. One very important assumption is that a rational agent can both believe and disbelieve that certain object or property x is such and such, or ascribe in thought contradictory properties, p_1 and p_2 , to an object only if there are distinct ways of thinking of the object, call them m_1 and m_2 , such that the agent believes x to be such and such under m_1 and disbelieves that the object is such and such under m_2 , or that believes that x has p_1 under m_1 and has p_2 under m_2 . On Frege's classical view these ways of thinking are part of the content of the belief; on more recent views, there are not part of the content (where belief is seen as a two-place relation) but nonetheless are mental representations which constitute one of the elements of the belief relation (where belief is seen as a three place relation). What these views have in common is that belief is seen as a relation to a

representation and that ascribing contradictory beliefs to an object in thought involves distinct representations.

It has been usual to add on top of this metaphysical view a view about belief ascriptions according to which a belief report of the form 'A believes that S' is true when it relates the agent to the relevant inner representations. According to these views then a report which is true of a rational agent must relate her to distinct representations. A philosopher who holds this must explain how a belief sentence can achieve this.

This traditional notion of rationality sees rationality as having to do with the representations inside the agent's head as opposed to the intelligibility of the agent to a reporter on her beliefs. The traditional view of ascriptions sees them as somehow referring to these representations.

The notion of rationality that I have discussed makes no assumptions about the nature of the psychological states which are alleged to underlie the belief ascriptions. The focus is placed instead on the intelligibility of the agent to the reporter. The sentences of the form 'A believes that S' that the reporter uses to talk about the agent are not assumed to make reference to such underlying states as it does not assume that there are such underlying states.

Nothing in what I have said implies that there is nothing going on inside the agent's head which is relevant to the truth of the belief reports. There is clearly a functioning brain and there may be all kinds of mental goings on including mental images and sounds, other kinds of phenomenal experiences. I only deny that we need to assume that there is any systematic unity of inner representations which belief reports are somehow targeted on.

1.9

How my response relates to other popular types of response

I want to conclude this chapter with some remarks about how my response to the puzzle fits into the logical space of possible solutions to the puzzle as it is

construed here. I have said that the puzzle is best characterised as involving the tension between Pierre's presumed rationality, on the one hand, and, on the other, a seemingly plausible characterisation of our belief reporting practices which apparently leads us to attribute contradictory beliefs to Pierre, or at least, issue belief reports of the form, 'Pierre believes that S and that not-S'.

Responses to the puzzle can be grouped, first of all, according to whether or not they accept that there is a genuine tension here. Consider those views which deny the tension. This can be done by focusing on the ascriptions and saying that the ascriptions themselves are somehow not suitable for assessments of an agent's rationality. This is a well-known strategy and can be developed along several different lines. One way is to invoke the so-called *de dicto*/ *de re* distinction and say that the ascriptions in question are to be understood *de re*. There is, then, still thought to be a level of mental representation at which assessments of rationality can be made, but the *de re* reports are thought of as abstracting away from this content in such a way that they fail to fully specify the content of the belief held, and so by themselves cannot be used in assessments of rationality. The advantage of this view is that it allows the philosopher of mind to maintain his view that Pierre does not after all have contradictory beliefs, where beliefs are thought of as having a content which is transparent to the agent in such a way that he can see, by mere introspection, that his beliefs are contradictory and so incorrect. Nathan Salmon can be seen as offering a view that falls into this category, and I examine his view in Chapter Three.

Then there are those views according to which the belief report is thought to need supplementing with a logical form which somehow makes semantically relevant mental representations which are part of two non-contradictory contents. A wide variety of semantic and/or pragmatic mechanisms have been suggested to account for how the representations which form part of the content of the beliefs (the level at which rationality is to be assessed) are put at semantic

issue. Mark Richard's is one such view and the one which I shall discuss in Chapter Three.⁹

The views above rely on a distinction between beliefs proper and the sentences we use to talk about those beliefs and it is assumed that there is a level of content, not visible in the surface of the reports, which nevertheless restores rationality to the agent.

Another quite different view denies the tension between rationality and the truth of an ascription of the form 'A believes S and that not-S', but only by proposing a radical revisionary view in epistemology, according to which it is simply not the case that rational agents cannot have contradictory beliefs. This is sometimes argued for by claiming that since externalism about the content of belief states (and other mental states) is determined by factors outside of the head, then an agent can have contradictory beliefs simply if he is ignorant of the factors determining the content and that in general he would not have such knowledge by mere introspection and so should not be convicted of irrationality just because external circumstances have conspired against him in such a way that he has a pair of contradictory contents as contents of two of his beliefs. It is sometimes said by proponents of such a view that the idea that rationality and having contradictory belief are incompatible rests on a faulty Cartesian epistemology which fails to acknowledge the truth of externalism.

I shall not discuss any such theory because the main motivation of such a theory seems to be to account for a false datum, specifically that we can unproblematically ascribe contradictory beliefs to a rational agent. If the arguments and claims of Section 1.8 are correct this is not the case, and so is not a fact that needs explaining. Furthermore, the delicate issues in the epistemology of belief that are raised by such theories are beyond the scope of this dissertation.¹⁰

⁹ Crimmins (1992) and Crimmins and Perry (1989) offer another such theory. David Sosa (1996) offers another variation on this idea.

¹⁰ Owens (1989) is a prominent example of such an approach. Although I do not treat such epistemological theories here, I must admit that the view I advance in Chapter Four would seem have consequences for the epistemology of belief, and I would ultimately like to try to work out some of these consequences. Considerations in epistemology have not constrained my view.

What each of the above types of view have in common is that they all try to reconcile the truth of the belief ascriptions with the fact of the agent's presumed rationality, but if what I say above about our never making belief ascriptions of the form 'A believes that S and that not-S', then this is simply not a task that we should set ourselves. I take seriously the empirical observation that we never say such things of an agent presumed to be rational. If this datum is taken as a starting point that it alters the shape of the task before us. My view does not depend on the standard distinction between belief states and the ascriptions we use to talk about them, but instead depends on a deliberate conflation of these two things. And this conflation allows me to account for the incompatibility of the belief ascriptions with an agent's rationality without relying on a Cartesian epistemology. On my view, the incompatibility does not rest on *any* epistemological view, but on a feature of our *belief ascribing practices*.

The task before us becomes the task of explaining why our belief reporting practices do not allow us to make ascriptions of the form 'A believes that S and not-S' of a rational agent.

But even those philosophers who have seen the task in this way have responded to it in a way which is tightly constrained by the combination of the view in philosophy of mind, described above in Section 1.8, according to which an agent can attribute in thought contradictory properties to an object only if she thinks of the object in two different ways, and the view in philosophy of language that such representations are referred to by the embedded clause in a belief report. Frege's original theory would be a theory which embodied such a view. The argument for the truth of the report 'Pierre believes that London is pretty and is not pretty' is then thought to fail when it is held that such sentences do not manage to secure reference to the relevant representations.

1.10

Conclusion

In this chapter I have tried to move slowly through Kripke's puzzle in order to draw out what I think is really at stake. I have argued that Kripke's puzzle does not really depend on the principles he says it depends on, but rather on a certain picture of belief and belief-reporting, according to which belief is a relation to a contentful representation and that belief reports report on this relation, the relation being thought of as a real psychological relation which obtains between the agent and the content, prior to what a reporter might want to say about the agent. I have suggested how Kripke's derivation of the puzzle can be blocked by reversing the picture: what an agent believes depends on what ascriptions are true of him in a given context, and the truth of these ascriptions does not depend on relating him to the content of his allegedly prior belief. This reversal of the picture suggests that we should consider first the puzzle in its metalinguistic guise, and when we do that we see that the puzzle does not seem to arise. I have maintained the view that rationality and having apparently contradictory reports true of one are incompatible by, instead of trying to explain why a rational agent cannot have contradictory beliefs, trying to suggest why contradictory belief reports can never be true of a rational agent. In order to make good on this proposal I have discussed a notion of rationality which is not defined in terms merely of what an agent believes, but has to do with intelligibility to a belief reporter.

The next chapter is a critical evaluation of several theories which I take to exemplify a good range of diverse theories which all embody the assumption that I have said lies behind Kripke's puzzle.

Chapter Two: A Grammatical Interlude

2.1

Introduction

In this section I want to consider the grammar of ‘A believes that S’ and its two nominalisations ‘A’s believing that S’ and ‘A’s belief that S’. I shall also consider one other nominalised form of this verb, that is, expressions which include ‘state of belief’. Each of these expressions is commonly found in the philosophical literature on belief and belief ascriptions.

I examine the grammar and meaning of the above verbal expressions because I think that we can draw some tentative conclusions about how to think about belief and belief ascriptions from considering the grammar of such expressions. I say ‘tentative’ conclusions, because there are no straightforward entailments between facts about the grammar of ordinary constructions and conclusions having to do with metaphysics, or the relation between metaphysics and semantics and methodology (the kinds of conclusions that I am ultimately interested in). Nevertheless, I think I can offer some support to some of the central claims that I wish to defend.

Ultimately, I wish to defend the view that our study of belief ascriptions should not presuppose that belief ascriptions depend for their truth on the existence of an independently specifiable set of beliefs that an agent has. I think that such a picture of independently specifiable beliefs is encouraged by focusing on philosophers’ talk of beliefs using some of the nominalisations that I will suggest are unnatural or derivative upon locutions containing ‘believes that’. I will draw, in Section 2.2, a tentative conclusion about the ontology of believing. I will also argue, in Section 2.3, that some other substantial metaphysical theses about belief are not supported by what we ordinarily say (not that there are not other reasons for these substantial metaphysical positions).

2.2

Zeno Vendler's classification of verbs

When a philosopher is said to be working on the semantics of belief ascriptions it is usually meant that the object of his inquiry is really something narrower: sentences of the form 'A believes that S', where 'A' is a stand-in for an expression that refers to an agent (paradigmatically a normal human being) and 'S' is a stand-in for a sentence. So, for example, 'Robert believes that animals should never be killed for food' serves as an instance of the phenomenon he is interested in. It is usually assumed that such sentences occur absolutely naturally in ordinary English. Ordinary usage is supposed to be the object of his inquiry. With this in mind, it is worth noticing that sentences containing 'believes that' can often sound cumbersome in a natural context and that one is more likely to hear sentences of the form 'A thinks that S', 'A thinks S' or 'A believes S' (in these last two there being no explicit occurrence of 'that'). Everything that I say about belief ascriptions in this thesis is to be understood as applying to utterances of each of the three sentence forms mentioned in the previous sentences, and may even sound more natural when directly applied to those forms.

'Believes' is standardly classified as a stative verb, other standard examples of stative verbs being 'having', 'wanting', 'being green'. Zeno Vendler (1957), for example, classes such verbs together and contrasts them with what he calls 'verbs of activity', 'verbs of accomplishment' and 'verbs of achievement'. His basis for this classification has to do with the way they relate to time. Vendler says that the question 'For how long did he V?' can only sensibly be asked when 'V' is replaced by an activity verb, e.g. 'For how long did he sing?' and that the question 'How long did it take to V?' can only sensibly be asked when 'V' is replaced with an accomplishment verb, e.g. 'How long did it take him to recite the poem?' Both types of verbs admit of continuous tenses (e.g. 'he was singing', 'she was reciting the poem'), but neither stative verbs nor verbs of achievement do. It makes no sense to say 'I was believing that S' (a stative verb) nor 'I was recognising him' (an achievement verb). Associated with such types

of verbs are the question forms ‘Do you V?’ and ‘Did you V?’. Stative verbs are distinguished from achievement verbs, according to Vendler, by the fact that stative verbs are not predicated for specific moments of time but rather for periods of time, although these periods are rarely specified precisely. It is much more usual, for instance, to hear someone say ‘he believed that Vermont is liberal for years’ than to hear someone say ‘I believed that from 8pm until midnight’. Vendler applies a simple linguistic test to distinguish achievement verbs from stative verbs. The question ‘At what time did you V?’ only makes sense for achievement verbs.

As we have seen, Vendler uses the criterion of whether a verb admits of a continuous tense form to distinguish between verbs of activity and verbs of accomplishment on the one hand, and verbs of achievement and stative verbs on the other hand. I suggest, however, that we should see stative verbs as an even more fundamentally different category than Vendler’s use of this criterion suggests. I suggest that we should modify his criterion to distinguish stative verbs from all other kinds of verbs. Instead of asking ‘Does ‘V’ ever usually admit of a continuous tense?’, which seems to be Vendler’s real question, we should ask ‘Does it ever make sense to use ‘V’ in a continuous tense form, whatever the perspective we have on what it is that we are describing?’ If the answer to this question is ‘No’ then we have a stative verb. Application of this criterion will have different results from Vendler’s, for even those verbs which he classifies with stative verbs since they do not usually appear in a continuous tense form, will count as having a perfectly natural continuous tense usage when the action is described from a certain perspective. Let me provide an example.

Presumably, ‘spitting out a button’ would count as a verb of achievement for Vendler since it would not normally occur in a present tense form, for the simple reason that it is a relatively instantaneous act, not normally something that one could be thought of as being in the process of doing. But I say ‘relatively instantaneous’ because it must be conceded that any action which might, for ordinary purposes, count as instantaneous could be viewed from a perspective from which it did not appear so. Suppose, for instance, that a video recording is made of a child spitting out a button and that this is played back to an audience

at slow motion. It would seem to make perfect sense, to say, pointing at the screen, ‘The infant is now spitting out the button’. Anything we would normally count as instantaneous – explosions, coughs, sneezes, blinking - can be viewed in such a way that they do not appear instantaneous. I submit that there is, however, no context in which it makes sense to say ‘the agent is now believing that S’.

Perhaps an ontological conclusion can be tentatively drawn here. Vendler observes that the fact that certain verbs do not admit of continuous tenses suggests that those verbs do not designate processes going on in time. If we agree that there is no context in which it makes sense to say ‘A is now believing that S’ and we agree that any event which may appear instantaneous from one perspective can be viewed from another perspective in which it is not counted as instantaneous, then we can see that the reason that ‘A is now believing that S’ never makes sense is not because believing happens instantaneously. Believing is not an event or enduring process of any kind. If this is true, then, a fortiori, it is not an event or process that underlies the truth of an ascription.

2.3

Nominal forms

I now move on to consider the nominals ‘belief (that S)’ and other less common nominalisations. Lynne Rudder Baker (1995) says that the term ‘belief’ is a nominalisation of ‘believes that’. She does not say precisely what she means by ‘is a nominalisation of’, but she makes many remarks throughout her book ‘Explaining Attitudes’ which suggest that she has in mind the view that in an important sense locutions which contain ‘believes that’ are more basic than locutions which contain ‘belief that S’. I shall explain what I mean by ‘more basic’ shortly.

‘Believes that S’ can be nominalised in various ways. The first two ways that I wish to inspect are 1) the gerundive nominal ‘believing that S’ and 2) the derived nominal ‘belief that S’.

Concerning the former, philosophers often try to use ‘A’s believing that S’ as if it were a count-quantifiable noun. Several philosophical theses are standardly explained by using it in this way. Despite the prevalence of this usage in the literature, this usage seems very awkward and is (almost) never encountered in ordinary contexts. The linguistic apparatus of pluralisation, quantification, demonstration and identification and the use of definite and indefinite articles do not seem to apply at all naturally to the gerundive nominal form ‘A’s believing that S’;

*A’s believings that S

*There is a believing by A that S

*That believing by A that S

*A’s believing that S is identical to A’s being in neural state XYZ

all sound extremely odd.

‘A’s believing that S’ only seems to occur at all naturally in subject position with such predications as ‘was very surprising’, ‘is useful to us’, etc. Any sentences containing ‘A’s believing that S’ involving one of these predications (let ‘F’ stand for any one of them), which seem at all natural here seem naturally paraphrased by ‘The fact that A believes that S is F’. This strongly suggests that ‘A’s believing that S’ normally designates a fact or state of affairs. Notice that the normal way of designating this fact involves the reappearance of the verbal form ‘believes that’.

There is also the noun ‘belief’. This noun can be naturally understood both as mass-quantified and as count-quantified. It is mass-quantified as it occurs in ‘the phenomenon of belief’ ‘A Puzzle about Belief’, ‘the metaphysics of belief’, ‘her belief in ghosts’. I mention this usage merely to set it aside. I doubt that it enters into any of the ambiguities and possible confusions which I shall discuss next.

I wish to focus more attention on ‘belief’ as it is understood as a count-quantified noun. Many have noted that ‘belief’ understood as a count-noun is

ambiguous. The point is usually put like this: Sometimes 'belief' refers to the proposition believed and sometimes it refers to the state of believing. Some philosophers try to disambiguate this by using the term 'object of belief' to designate the proposition believed and try to use 'the belief that S' to refer to the state of believing. (I shall say more about this 'state of' locution shortly). Unfortunately, 'object of belief' is itself ambiguous. Sometimes it means the object thought about and sometimes the proposition believed. So, for example, if one believed that the moon is a planet, then the object of belief could either be the moon or the proposition that the moon is a planet. I shall avoid the expression altogether and use the unambiguous 'what is believed', which is also more neutral about the correct analysis of the metaphysics and semantics of believing.

It is often said that the expression 'the belief that p' designates the content of the belief. This view seems to gain support when we consider that beliefs can be true, false, verifiable or unverifiable, that beliefs can be consistent or inconsistent, and that one belief can logically entail another. These semantic or logical properties are ones that we naturally ascribe, in the first instance, to propositions. So, perhaps 'belief that p' can be used in such a way that it is synonymous with 'the proposition that p'.

On the other hand, it is common to hear philosophers talk of the 'content of a belief' and this must mean that the belief in question is not to be identified with its propositional content, but rather as something that can be thought of as *having* a content. On this understanding, where a belief is still thought of as an abstract entity, beliefs are not wholly individuated by their content. The hope that p is clearly not the same thing as the belief that p. On this understanding, the belief that p is what is sometimes called a 'propositional attitude', as opposed to a propositional content. It is important to remember that, on both understandings of 'the belief that p', where belief is a count-noun, the expression designates an abstract entity. Philosophers often seem to run these two obviously different usages together, talking of them as attitudes and then saying that they are abstract entities individuated by their content. To avoid possible confusion, I will use the expression 'the belief that p' to designate the

attitude not the proposition. And I will avoid predicating logical or semantic properties of these attitudes for that encourages us to think that we are talking about the contents that those attitudes are attitudes towards rather than the attitudes themselves. I think that both these usages exist in ordinary use and that they are not clearly distinguished there either.

There seems to be a third distinct possible understanding of the nominal 'belief that p'. It seems, as Helen Steward (1994) has noted, that we sometimes have a need to distinguish beliefs held by one person from beliefs held by another person where those beliefs have the same content. George's belief that Mr. Y will be proved innocent is well-grounded whereas Jane's belief that Mr. Y will be proved innocent is not. In a sense, these beliefs must be distinct; they have different subject-relative properties, and, after all, one is George's and one is Jane's. This supports the idea that there is a reading of 'belief that p' (where it is used in a possessive construction of the form 'A's belief that p') according to which the identity of the agent is an essential part of the individuation of what it is that is designated by the expression. It is tempting to think that such possessive constructions are never confused with the two described in the previous paragraph. The very use of the possessive might be thought to show that we are referring to the subject relative belief, but this seems not to be the case, for we can also talk about A's belief that p being true. This suggests, I think, that it is the belief conceived of as a proposition that is in question, rather than the fact that it is A's belief.

It is essential to mention here that expressions of the form 'A's belief that p' have really taken on a life of their own in recent philosophy of mind. Expressions of this possessive form are sometimes taken to designate what Mark Crimmins and John Perry has called 'cognitive particulars'. This choice of expression is revealing, for they do mean to use 'particulars' in the philosophically standard sense of that term. They mean that they are not abstract entities and that they are the kind of thing that can be referred to in indefinitely many non-analytically connected and non-relational ways. The move from thinking of beliefs as subjective relative to thinking of them as cognitive particulars is a fairly subtle move, but it is a move nonetheless. For there is

nothing in the idea that sometimes we need to distinguish between A's belief that p and B's belief that p that leads us directly to the idea that A's belief that p and B's are cognitive particulars.¹¹ All that we are led directly to is the thesis that the fact that A believes that p is distinct from the fact that B believes that p.

It is worth remembering that the kind of locutions which led to us needing to distinguish different subject relative beliefs were predications like 'is (not) well-grounded' or 'was formed yesterday'. On the other hand, recently, philosophers have assumed that other locutions are readily intelligible, locutions like 'A's belief that p is identical to the neural state XYZ' or 'A's belief that p caused her to duck' (where 'caused' is further construed in a particular tendentious way). While the second grouping presuppose the particularity of the subject-relative belief in question, the first ordinary usage grouping is, I think, most naturally dealt with by paraphrasing. 'John's belief that p is well-grounded' is naturally paraphrased as 'John believes that p but he is not warranted in that', where 'believes that' re-emerges and there is no suggestion that a cognitive particular is at issue.

Locutions involving 'state of'

There is also a nominalisation of 'believes that p' which has the form 'the state of belief (or believing) that p' or 'a belief state'. These expressions occur rarely, if ever, in ordinary usage. So, if we wish to understand these expressions we must try to find out what philosophers mean by them.

Since philosophers of mind so often talk of identifying belief states with brain states we could easily be led into thinking of these expressions as having the same logical structure. They do have a superficially similar structure, but we can very readily see a deeper difference.

I suggest that to find the most natural understanding of the expression 'belief state' we can look at other natural expressions involving 'state'. We can speak

¹¹ Crimmins and Perry (1989) think that there are other, metaphysical reasons for positing beliefs as cognitive

of ‘a state of health’, or ‘a state of fitness’, ‘a state of anger’, ‘a state of agitation’, ‘the state of being undressed’, ‘the state of an economy’, ‘the state of my bedroom’. Even this short list shows that what follows ‘state of’ is a motley of expressions. We have derived nominals referring to properties (the first two) and then expressions referring to the state types themselves (‘anger’ and ‘agitation’), we have a gerundive nominal referring to a state (‘being undressed’) and we have nouns referring to concrete entities (‘the economy’, ‘my bedroom’). In the case of the last two, what follows ‘state of’ is the thing which is in the state in question, but what that state is not itself specified by the use of the expression ‘state of..’. I would say that ‘brain state’ is most naturally classified with ‘the state of my bedroom’, where the noun ‘brain’ is to be understood as the object which has certain properties, and that its having certain properties is almost certainly to be thought of as its having certain of its parts in a certain configuration and/or exemplifying certain properties. (Steward 1997)

A brain-state, then, is a state of the brain, that is to say, it is a state that a brain is in, and the state a brain is in (at a time) is a matter of what properties it has (at that time). A belief-state, however, is not a state of a belief. It has nothing to do with what properties the belief in question has. Nor is it clear the ‘state of belief’ can be classified along with any of the other locutions in the list above, for ‘belief’ is neither to be understood as a property or state type itself, for while we can say that ‘A is in a state of anger’ for instance, we cannot say that ‘A is in a state of belief or in a belief state. It seems incoherent to assume that A can be in a belief state simpliciter. It makes sense, although it is awkward, to say ‘A is in the state of believing that S’. The only natural way to express this seems to be with a gerundive nominal, and this suggests that the verb, as it were, does not want to be suppressed with a nominal expression. The verbal form ‘A believes that S’ seems to be the *basic* form.

None of the above proves, of course, that ‘brain states’ cannot be identified with ‘belief states’. I have simply tried to show what this claim might involve. I have tried to suggest that it involves the claim that the fact that an agent’s brain has a

particulars. Although this is beyond the scope of this thesis to argue this, I think the arguments are unconvincing.

certain (complex) configuration or set of properties is to be identified with the fact that the agent believes a certain thing. This view will itself be attacked later on.

The conclusions I wish to draw are that ‘A believes that S’ is more basic than any use of its associated nominals, that ‘belief that S’ is three-ways ambiguous and these usages are often not sufficiently distinguished in the literature, that all claims involving the expression ‘believes that S’ or its associated nominals are best understood as facts about the agent, except that ‘A’s belief that S’ may simply mean to refer to the proposition believed. I have used considerations of naturalness as a methodological starting-point. I have not argued that such considerations are necessarily decisive. By themselves they are not.

2.4

A qualification about ‘facts’

I have suggested that most of the awkward sounding locutions such as ‘A’s believing that S’ should be understood as, if they designate anything, designating facts. But I wish to emphasise that this is not meant to imply that belief ascriptions are made true by facts about what an agent believes if this is understood as meaning there are determinate context-independent facts about what an agents believes that ground the truth of the report. (My talk of facts was meant to be understood in as deflationary a way as possible). What I have tried to prepare the ground to suggest in this chapter is that we should look at our ordinary belief ascriptions using the verbal form ‘believes that S’ if we want to understand our ordinary concept of belief, rather than reify beliefs and then ask questions about how we manage to report on those beliefs.

2.5

Conclusion

In this chapter I have looked at several different types of expression involving ‘belief’, ‘believing’ or ‘believes’ and suggested that ‘believes that S’ is a fundamental form. This suggests, to me, that in order to understand belief ascriptions we should focus on exactly that, belief ascriptions, and not depend

on unusual locutions which may support what I think is a mistaken picture of belief reporting. I have briefly argued that believing is not a matter of having anything going on inside one, no event or process. Moreover, there is nothing about any of the locutions I have examined which supports the view that there is an independent metaphysics of belief to ground our practice of belief reporting.

Most of the considerations in this chapter are not decisive with respect to what I ultimately want to argue. However, the next chapter aims to criticise a range of theories with I think have as an assumption an independent and constraining metaphysics of belief.

Chapter Three: Evaluating Other Theories

3.0

Introduction

In this chapter I will critically examine several quite diverse theories of belief ascriptions. I will try to explain how they apply, or could be applied, to Kripke's puzzle. I will present criticisms of each theory and try to show how they each embody the crucial assumption that this dissertation aims to cast doubt on, and that the way they embody this assumption is central to their failings. The theories that I select are, of practical necessity, only a few of the many different theories that are to be found in the literature, but I think that many of the theories I do not treat directly can, to some extent, be seen as variations on the ones I discuss. I start, in Sections 3.1 and 3.2 with a discussion of Frege, at least under a very popular and influential interpretation. Before I discuss (in Section 3.3) how Frege could have responded to Kripke's puzzle, I want to articulate some Fregean assumptions about the nature and structure of belief ascriptions, in particular how they relate to simple sentences and issues connecting to substitutivity, since it is with Frege and subsequent discussions of Frege that some very fundamental assumptions about belief ascriptions have their source. I will use this section to try to emphasise that these assumptions, although plausible, are nonetheless substantial assumptions that we should be sensitised to in thinking about belief ascriptions. In Section 3.4, I look at Nathan Salmon's theory of belief ascriptions, a theory which was developed in response to a then new theory of the semantics of simple sentences which rejected Frege's account of simple sentences. It will be seen that although Frege's semantics for simple sentences has been rejected some basic assumptions about the structure and function of belief ascriptions was left in place. In Section 3.5 I discuss a more radically different theory of belief ascriptions according to which belief ascriptions do not relate agents to any kind of thought contents at all. It will be seen that this theory faces very serious problems and still,

despite its radical nature, retains the basic assumption I want to cast doubt on. Section 3.6 discusses a ‘contextualist’ theory of belief ascriptions which avoids some of my main objections to the other theories discussed, but has different problems of its own.

3.1

Frege on Simple Sentences

According a very popular and influential interpretation of Frege, due mainly to the work of Gareth Evans (1982) and of Michael Dummett (1973), Frege was the first to formulate a systematic theory of meaning for a fragment of natural language whose aim was to explain how the meaning of sentences depends upon the meaning of their parts. The focus was ultimately on the meaning of sentences because Frege noticed that it is only sentences that can be used to say things capable of being true or false. Frege, then, wanted to be able to explain how the meaning of components of a sentence contributed to how a sentence could be true or false (given the way the world is). He assigned non-linguistic entities to subsentential expressions in order to explain how a sentence would say what it said. These can be called the ‘semantic values’ of expressions since they contribute to the meaning of the sentence. In his earliest work he concentrated on extensional fragments of language (i.e. fragments in which any two coreferential expressions can be intersubstituted anywhere *salva veritate*) and so it was natural to make the decision to assign expressions their referents as semantic values.

In very general and abstract terms, what I have said so far can be thought of as a description of a semantic theory having only to do with the relation between linguistic expressions and the world. The idea of users of a language has not even entered the picture yet, and neither has the idea of the use and understanding of the language. Another distinct level of theory and description having to do with use and understanding was necessary in the semantic theory. To appreciate why Frege thought that such a level of description was necessary we need to look in more detail at two things. We

need to understand, in general terms, what Frege saw his semantic project as being about. Secondly, we need to see why this goal is not met by the theory discussed so far.

To appreciate Frege's overall project and starting point is to see that there could be no question for Frege of regarding a potential language user's use and understanding of her language as separate from an account of the relation between language and the world. For what Frege was primarily interested in was the structure of thought. Thoughts, for Frege, are abstract objects which are the primary bearers of truth and falsity. The meaningful sentences of a language are meaningful only in virtue of expressing thoughts, and sentences are bearers of truth-values only derivatively. Thoughts are what we think when we think something. The thought that a sentence expresses is what we grasp when understand a sentence. Thoughts are non-mental and are essentially public as they are what is conveyed when one speaker of a language says something to another by uttering the sentence expressing the thought and when the other understands what is being said by grasping the thought.

Since Frege's primary interest was in thought so understood, his interest in language was in a sense derivative. But he had to be interested in language nonetheless since it is not possible to investigate thought without investigating language. Frege believed that we cannot refer to, express or convey a thought except linguistically and that we cannot have a thought that we cannot in principle express, at least to ourselves. He also believed that any attempt to investigate thought without studying its linguistic expression would inevitably lead to confusing the psychology of thinking with the study of thought (Dummett 1973).

The study of language, then, does not presuppose the study of thought. On the contrary, the latter is, in fact, impossible without the former. It is absolutely central to Frege's philosophical project, then, that language and thought should be unified. Frege's theory of meaning was a unified theory in the sense that it aimed to provide for each sentence of the fragment of

language under investigation a content which is the meaning of the sentence that expresses it as well as what we think when we think the thought that the sentence expresses. Frege's project was, as Joseph Almog (2004) has recently labelled it, 'translational'. This means that giving a semantic account of a sentence is giving a synonymous sentence which unpacks the meaning of the target sentence. It gives, as it were, the speaker's perspective on the sentence.

So far, I have tried to explain, in general terms, why Frege's treatment of language and thought was (and had to be) unified. Next I will discuss a more specific constraint on that distinct level of theory and description.

Frege begins one of the most famous passages of analytic philosophy with the following:

Identity challenges reflection through questions which are connected with it and are not altogether easy to answer 'a=a' and 'a=b' are obviously sentences of a different cognitive value: 'a=a' holds a priori and is according to Kant to be labelled analytic, whereas sentences of the form 'a=b' often contain very valuable extensions of our knowledge and are not always to be grounded a priori.....If we wanted to view identity as a relation between that which the names a and b signify then 'a=b' and 'a=a' would seem to be potentially not different, in case that is 'a=b' is true. There would be thereby expressed a relation of a thing to itself, one in which each thing stands to itself, but no thing stands to another.¹²

What exactly is the nature of the problem Frege is describing here? Frege asks: how can a sentence of the form 'a=b' if true, differ in cognitive value from a sentence of the form 'a=a'? To say that they differ in cognitive value means that someone who does not already know the truth of 'a=b' could rationally have different epistemic attitudes to the sentences at a given time, accepting one as true and taking the other as false, or be agnostic about its truth value.

That is the form of the basic question, but where is the philosophical problem? To see the problem here it is necessary to state the assumptions

that lead Frege to ask his question. Frege explicitly held (1) that understanding a sentence is grasping the thought that it expresses upon reading it. It can also be assumed, from his overall form of argument in describing his puzzle, that he also held that (2) meaning is transparent in the sense that if a language user attaches a single meaning to each of two expressions, then she must know that she does so. Frege also implicitly held (3) a principle of compositionality according to which if thoughts t_1 and t_2 have the same structure and are composed of the same elements then t_1 is t_2 .¹³ As we have seen, Frege held that the semantic value of a singular term is to be identified with its referent. The question that Frege poses raises a problem then on these three assumptions, since if 'a=b' is in fact true, 'a=a' and 'a=b' will express thoughts composed of the same elements in the same structure. By (3), then, the thoughts expressed will be identical. By (1) and (2), in understanding the two sentences the language user will see that the same thought has been expressed by the two sentences. She could not, therefore, rationally have different attitudes to the two sentences. But it seems an undeniable fact that she can do.

The problem as originally stated by Frege (1948) is presented as a problem about identity, or the identity predicate, but there is a consensus these days that the problem has nothing to do with identity, since the same set of philosophical issues arise for any pair of sentences that differ only in that they contain different coreferential singular terms¹⁴. For example, one could believe that 'Hesperus is a very shiny star' were true and not believe that 'Phosphorous is a very shiny star' were true, and yet be competent with the names. On the same assumptions as presented in the paragraph above it is puzzling how a language user could rationally have different epistemic attitudes to each of the pair of sentences.

Frege's response to this puzzle was to invoke another level of semantic description to account for the possibility that an agent could rationally have

¹² Frege (1948, pg. 209)

¹³ Assuming that Frege implicitly held this assumption seems necessary to reconstruct the argument of the first paragraph of 'On Sense and Reference' quoted above, although it is in prima facie tension with Frege's claims that thoughts are essentially unstructured. (Frege 1956).

different epistemic attitudes to any pair of sentences which differed only in that they contained different coreferential expressions in the same position in a given structure of a sentence.

As formulated above, Frege's Puzzle (as I shall call it after Nathan Salmon's popular coinage) has to do with epistemic attitudes to sentences (believing the sentence to be true or false, or being agnostic as to its truth-value). There can be no doubt that a rational individual can have, in the relevant sense, different epistemic attitudes to the individual sentences of the relevant kinds of pairs. But it is common to find in the literature an importantly different formulation of the puzzle, one which describes the puzzle by asking: how can one think that, say, Hesperus is Hesperus without thinking that Hesperus is Phosphorous?

According to this different formulation, there is no mention of sentences and no direct reason therefore that an answer to this puzzle should require a solution in terms which makes reference to linguistic meanings. But this is not quite the way Frege puts his puzzle. He talks explicitly of sentences and our understandings of them and that is why his problem, understood in the light of his assumptions, requires a *semantic* solution. Given these assumptions, Frege asks, equivalently by his own lights, how can 'a=a' and 'a=b' express different thought contents? But it is important to note that this is an equivalent question to the original question only on the assumption that understanding a sentence means grasping its content so that a difference in possible epistemic attitudes to sentences is understood as a derivative phenomenon, being dependent upon different epistemic attitudes to the contents that sentences express.¹⁵

If two sentences express distinct thoughts, yet agree in point of reference and structure, then the thoughts cannot be fully determined by the identity of their referents and mode of combination. The identity of thoughts must therefore be determined by something else. This something else Frege called

¹⁴ Nathan Salmon (1986) was the first to make this point.

‘sense’. Frege spoke of sense of an expression as a mode of presentation of the expression’s referent, the particular way that a referent is thought of. Essential to Frege’s idea of sense was that it had to be a property of an expression of a public language and that in order to understand utterances containing the expression the language users must all associate with the expression a single particular way of thinking of the referent. The reason for this is that Frege wanted to ensure that, when communicating, a speaker would transmit and the hearer receive a single thought.

In explaining how this notion of sense addresses how pairs of sentences like ‘Hesperus is shiny’ and ‘Phosphorous is shiny’ can differ epistemically, it is enough to point out that the two sentences express different thoughts since ‘Hesperus’ and ‘Phosphorous’ express different senses. And it is not trivial that the two distinct ways of thinking associated with the two names determine the same referent.

It is difficult to give more of a positive account of how postulating distinct senses explains the relevant epistemic differences without falling into the trap of making it sound as though what one grasps when grasping, say, an informative identity sentence like ‘Hesperus is Phosphorous’ is the fact that the distinct senses associated with the two names determine the same referent. This makes it sound like the sentence is about the senses of the names. But this is not the case. The sentence is about the planet Venus. Informativeness then is not to be understood as having to do simply with what the sentence is about, but rather what the thought is constituted out of, and Frege was original in suggesting that these two things are distinct.

Frege’s notion of sense arose, I think, primarily out of a concern to address the puzzle just discussed, although the great power of the concept can only really be seen in how it relates to other theoretical issues which are not directly relevant. What is relevant to this project, though, is how it plays another closely related role in his theorising, that is, in his semantics for

¹⁵ That these two questions are equivalent only under such theoretical assumptions is a point that was first discussed by Howard Wettstein (1989). I will return to the significance of this observation later.

psychological report sentences, including belief ascriptions, and that is what I shall now turn my attention to.

3.2.1

The semantics for belief sentences and the issue of substitutivity

In this section, I will look at a question of substitutivity in belief ascriptions. It is often assumed that the problem I am about to discuss is necessarily closely related to Frege's Puzzle discussed in 2.1. Often, it is explicitly said to be the same problem posed in a different way. In the Stanford Encyclopedia of Philosophy in a section on 'Propositional Attitudes', it says "Frege's puzzle can be posed as a question about propositional attitude reports". I want to use this section to show why, although on Frege's picture this seems to be the case, we have good reason to deny the tight connection. This will provide a motivation for focusing only on belief ascriptions and not questions concerning the cognitive value of simple sentences if it is belief that we are primarily interested in.

Apart from the problem of explaining how a pair of sentences of the forms 'a is F' and 'b is F' can differ in cognitive value, there is an *allegedly* closely related problem of explaining how belief ascriptions such as (1) and (2) below can differ in truth-value when 'a' and 'b' are coreferential expressions.

- (1) Jane believes that a is F.
- (2) Jane believes that b is F.

For it is widely assumed that such pairs can differ in truth-value. The 'problem' here is that the difference in truth-value would seem to be a violation of the principle of substitutivity. The principle is a rule of inference that sanctions the validity of arguments of a certain form. In particular, it sanctions the inference from the premises 'a is G' and 'a is (identical to) b' to the conclusion 'b is G', where 'a' and 'b' stand for singular terms and 'is

G' is any predicate, so that 'Jane believes that.... is bald' could be a relevant replacement for 'is G'. The principle is thought to receive considerable intuitive support from an apparently solid intuitive assumption about what has been called the 'aboutness of language' together with the assumption that the role of terms whose standard function is to designate. The assumption in question is that the use of a singular term enables a speaker to affirm or deny something about the entity designated by the term and the truth-value of what would thereby be asserted would depend only on whether the object in question has the property it is said to have, or lacks the property it is said not to have (Marti 2007). The principle would also receive more theoretical support from the notion of propositional content and the thesis of compositionality. Two terms which expressed the same content would seem to be interchangeable in the sentence if compositionality were true. In recent discussions of substitutivity, the assumption that names and indexicals are directly referential leads to the thesis that substitutivity holds for these kinds of expressions. This will be explained in Section 3.4.

An apparent violation of the principle thus needs to be accounted for. Frege offered a solution that involved the idea the apparent violation of the principle is *merely* apparent. Frege held the view that singular terms in the position occupied by 'a' in 'Jane believes that a is F', do not have their usual referents, but instead refer to the senses that they usually have as they occur unembedded. Much has been written about the difficulties of such a change in reference. Some have argued that there being such a systematic *change* of reference is in itself theoretically undesirable, and some have worried about how the new reference is actually achieved. It is alleged that Frege's solution, would, if not for these difficulties, and possibly others, be a very elegant solution. For, as we have seen, in Section 3.1, Frege's theory for simple sentences assigns to them *thought*-contents. So, a semantic account of belief report sentences which syntactically embed them, according to which 'A believes that S' is construed as a relational statement, with 'A' a singular term, 'believes' a dyadic predicate, and 'that S' a second singular term which refers to the thought-content expressed by 'S', can seem very natural. It can seem very natural because there are independent grammatical

reasons for holding that ‘A believes that S’ is a relational form of statement. Secondly, there is, allegedly, an independently popular metaphysical picture of belief according to which believing is standing in a relation to a representational entity of some kind.

Frege’s theories of simple sentences and of belief sentences are, it seems, implicitly accepted in some contemporary discussions of opacity¹⁶ when it is acknowledged to be a sufficient condition of its being a non-sequitur to move from ‘A believes that a is F, to ‘A believes that b is F’ given the identity of the referent of ‘a’ and ‘b’, that ‘a’ and ‘b’ differ in sense, where sense, even if it does not have all the properties that Frege says it does, is what is taken to solve Frege’s Puzzle. That this is implicitly accepted in contemporary work is evidenced, I think, by the fact that apparent counter-examples will be explained away either by saying that the apparent counter-example is merely apparent because it is either a pragmatic phenomenon or that the belief sentences in question are to be understood as having a *de re* profile. Jennifer Hornsby (1997) exemplifies the former approach when she says: “of course, in practice we often *feel* justified in moving from ascription of a belief with one content to ascription of a belief with what has to be counted another content. It may be *in a strict sense* a non-sequitur, but nonetheless *a safe bet* to move from ‘Harriet believes that the present British Prime Minister is uncompromising’ to ‘Harriet believes that Margaret Thatcher is uncompromising’. It can often be *taken for granted* that if someone believes one thing she often believes another.” (pg.202, italics added).

Of course, if one has already accepted Frege’s theory of belief sentences and his theory of simple sentences, then one will see the idea that the substitutions that are strictly speaking permissible depend on Frege’s notion of sense as developed in response to Frege’s Puzzle as self-evident, but if a philosopher were looking for independent support for Frege’s account of

¹⁶ A context is opaque if substitutions there do not guarantee truth preservation.

belief sentences, then she would have to look at the substitutivity data impartially. I think that this is rarely done.

It is possible, I think, to cast doubt on the idea that facts about what are seen as acceptable substitutions are such as to support Frege's theory of belief sentences by pointing out that there are many examples in which it is legitimate (that is, really, strictly, semantically legitimate, as opposed to apparently or merely pragmatically legitimate) to interchange terms which cannot be counted as having the same sense. In the remainder of this section, I want to suggest that it is possible that the substitutivity data are not such as to support Frege's theory of belief sentences, and this will involve trying to remove one potential a priori obstacle to this possibility by questioning what has become a standard and deeply entrenched, but not usually articulated, view of the relationship between Frege's Puzzle and the substitutivity 'problem'. Discussion of the apparent counterexamples will be deferred to chapter four.

It is often assumed that Frege's Puzzle (construed in its generalised form, and not as restricted to identity) and the substitutivity 'problem' are essentially the same problem, maybe just expressed in different ways, the latter just being the former recast in the formal mode. This can seem natural enough when we think that Frege's puzzle is concerned with an important aspect of the individuation of expressed belief contents, and belief sentences ascribe beliefs. As Jennifer Hornsby (1997, pg. 202) says:

"Frege held that it is a sufficient condition of the difference of two thoughts that it be possible for a person to have some attitude towards one and not the other [...]. We cannot be indifferent to whether we attribute (say) the belief that p or the belief that q to a person if we know that that person might believe that p and not believe that q; and we must use a notion of content which precludes such indifference."

Presumably Hornsby is thinking, in the first part of the quote, of the notion of thought-content that Frege developed as a response to Frege's Puzzle. She then goes on to draw a conclusion from the observation that two thoughts cannot be the same if an agent can have different attitudes to them. This

reasoning would seem to involve nothing more than the principle of the indiscernibility of identicals. The conclusion is a claim about ascribing beliefs, and it seems extremely plausible. Of course, we cannot be indifferent about which of two *distinct* beliefs we ascribe, if we want to ascribe the right one! A conclusion about failures of substitutivity then seems not far away. The conclusion relating to the case in hand would be that 'p' is not intersubstitutable with 'q' *salva veritate*, when 'p' and 'q' are sentences to which a speaker can rationally have different epistemic attitudes in the sense discussed in Section 3.1. I will return shortly to explain why, if this is the line of thought, it is too quick.

Gareth Evans (1982) is very explicit about the connection between ascribed beliefs and Frege's Puzzle. What Evans says further encourages the line of argument that I described above, and which I wish to question. He observes that, although Frege said rather little about the notion of sense, he did link it to notions ordinarily employed in propositional attitude psychology in such a way that there will be a very tight restriction on the individuation of senses. Evans calls this restriction 'Frege's Intuitive Criterion of Difference'. The criterion says that the thought-content associated with one sentence S as its sense must be distinct from the thought-content associated with another sentence S1 as its sense if it is possible for someone to understand both S and S1 and rationally have different epistemic attitudes to them. One could go so far as to say that this criterion does not just happen to be true of the concept of sense, but is (at least partially) definitional of the notion, since it was exactly this problem of cognitive difference that the notion of sense was designed to account for.

It can seem (wrongly, I believe) that the Intuitive Criterion of Difference has straightforward implications for the issue of substitutivity. If its connective is taken as a biconditional, then it appears to give necessary and sufficient conditions for when two terms can be intersubstituted *salva veritate*. If it is taken as a one-way implication, as Evans puts it, then it appears to give sufficient conditions for the failures of substitutivity. I think that no such tight connection exists between Frege's Puzzle and the issue of

substitutivity, even if we accept Frege's Intuitive Criterion of Difference. In order to make sense of this challenge, I think that it is important now to mark a three-way distinction so that we do not conflate issues which are in fact quite separate and assume that there are implications where in fact there are not.

1. As we saw in Section 1.1, Frege's Puzzle is, in its pure form, about sentences of a language: how can an agent understand pairs of sentences differing only by different codesignative singular terms and think that one is true but not that the other one is? Thought of like this, it is a question in the philosophy of language or semantics and is not (in the first instance) a question about belief individuation, although the type of answer that Frege is bound by his theoretical framework to give, makes us tend to think that the question is simply a question about belief contents.
2. Next, there is an issue about belief contents and what an agent can rationally believe. This would most appropriately be seen as an issue in the philosophy of mind, or psychology.
3. Next, there is the question of substitutivity. This is a thesis, not about mind, but about what patterns occur in our belief ascriptions. So, it is a view in semantics, but not in the semantics of *simple* sentences.

Given Frege's theory and the underlying picture which has been inherited by most philosophers working in this area, the three can seem to be inextricably connected. But that is why it is important to try to disentangle the issues.

3.2.2

Separating issues in simple-sentence semantics from issues about thought

My ultimate object of interest is the semantics or ordinary belief ascriptions. There is an orthodox, but rarely spelled out, chain of reasoning connecting Frege's Puzzle to thought content individuation and then that to issues about substitutivity in belief ascriptions. What I will try to do in this section is break, or at least give some cause to question, the link between Frege's Puzzle and thought content individuation.

The distinction between the question of informativeness as applied to sentences and the question of the individuation of thought-contents has been clearly articulated throughout essays in a book by Howard Wettstein⁶. The two questions, 1) how can 'a is a' and 'a is b' have a different epistemic status for a speaker? and 2) how can 'a is a' and 'a is b' express different thought-contents? are only equivalent on the assumption that understanding a sentence is grasping the content that it expresses. That this is a substantial assumption and is not in any way truistic is made clear by considering that one could give an answer to the former question without invoking any notion of content at all.

The answer to the question of how 'a is a' and 'a is b' can differ in informativeness would be a fundamentally different kind of explanation from the Fregean kind. Roughly speaking, 'a is b' could be informative if whatever conditions are sufficient for counting as competent with the terms did not necessarily suffice for knowledge that the terms are coreferential. The idea would be that to count as competent with a term requires relatively little; one could simply pick up and start using the names intending to use them with the same references as those from whom one picked them up, and be able to use them in the appropriate grammatical way and be able to ask, for instance 'Who is a?'. There is nothing in these conditions that requires that the speaker should know a priori that 'a is b' is true. On such a view, there is no presumption that the speaker should know that a is b even while he can use both terms. The idea of content has no place in such explanations,

¹⁷ Wettstein (1995)

which have been described as ‘anthropological’, as opposed to traditionally semantic.

Whether such explanations are acceptable as replacements of explanations which utilise the notion of content is not at issue here, but I observe that such putative explanations would need to be properly assessed and not dismissed a priori, and that is enough to show that the two questions are distinct.

It is not automatic then that we are in a position to draw any conclusions about what an agent may rationally believe from observations about what epistemic attitudes she has towards sentences. For while it is obvious that one can have different epistemic attitudes to the members of the pair ‘Hesperus is bright’ and ‘Phosphorous is bright’, it is not clear that we can always say of such an agent that he believes that Hesperus is bright, but does not believe that Phosphorus is. In a context in which the names are being used by people who ordinarily use both names interchangeably, such an attribution can sound paradoxical. If we imagine the use of the names in these ascriptions to be accompanied by a pointing, then it can seem as if we are ascribing contradictory beliefs to the agent. There may be some contexts in which the ascriptions are understood unproblematically and others where they are not, whereas saying that the agent has different epistemic attitudes to the sentences is always unproblematic. If we consider Kripke’s Pierre who has different epistemic attitudes to the sentences ‘London is pretty’ and ‘Londres is pretty’, believing the latter but not the former, can we say, straightforwardly, that Pierre believes that Londres is pretty but that London is not? Not necessarily, for we have the difficulties with giving any simple straightforward statement of what Pierre believes, even though there is no difficulty in saying what his epistemic attitudes to the sentences are.

More abstractly, if, as was suggested in Chapter 1, what beliefs we can say that an agent can have is a context-sensitive matter (the case for this will be developed in detail in the final chapter), then that observation itself is sufficient to cast doubt on the assumption of such a close connection

between epistemic attitudes to sentences, for what epistemic attitudes an agent has to sentences that he understands would not appear to be a context-sensitive matter.

I think that nothing other than the Fregean assumption that understanding a sentence is just grasping the thought-content it expresses would suffice to ensure the assumed connection between attitudes to sentences and beliefs held. One other candidate would be a principle very much like the disquotation principle (which was discussed in detail in Chapter One). *If* such a principle were true then we would be able to draw conclusions about beliefs held. The disquotation principle is a principle linking assent to a sentence with holding a belief expressed by the sentence. Presumably, though, for one to understand and accept a sentence as true would be enough for one to be in position to assent to a sentence in a way that satisfied the antecedent of an instance of the disquotation schema. But, as I suggested in Chapter, and as I will argue in more detail in Chapter 4, when I discuss the relation between saying and believing, there is reason to think that the disquotation principle is false. The reasons for thinking that it may be false would also apply to this related principle.

It is important to remind ourselves that the Intuitive Criterion of Difference does, as Evans observes, tie the notion of sense to the concept of belief definitionally, but, if we are mindful of the distinction that Wettstein articulated, then we will see that epistemic attitudes to sentences involve the notion 'believes to be true' rather than the simple 'believes'. Of course, most philosophers will agree that that these are definitely two distinct notions, but they say this, I think, because they want to be able to say that one can accept as true a sentence that one does not understand. That is no doubt true, but I question whether understanding the sentence fills the gap between 'believes to be true' and the simple 'believes' that allows us to move from the former to the latter.

So far I have tried to question the relation between Frege's Puzzle and the individuation of beliefs. What I have said is meant to imply that there is no

simple direct route between the two issues. In particular, there is no simple direct route from the observation that one can have different attitudes to a pair of sentences ‘a is F’ and ‘b is F’ to the claim that one can believe that a is F and fail to believe that b is F.

3.2.3

Separating issues about belief in the theory of mind from the semantics of ordinary belief ascriptions

I now turn my attention to trying to cast some doubt on the link between thought content individuation and issues about the semantics of ordinary belief ascriptions.

Epistemic attitudes to sentences are, to speak metaphorically, relatively ‘out in the open’. Beliefs themselves, thought of as the subject matter of psychology or philosophy of mind, are not out in the open. If there is no direct and simple route between epistemic attitudes to sentences and beliefs, one might wonder whether there is anything else ‘out in the open’ from which there is a direct route to beliefs.

Nathan Salmon¹⁸ suggests that there is a very direct connection between our ordinary everyday belief ascribing practices and claims about the individuation of belief contents. He proposes such a connection in what he calls the ‘thesis of the substitutivity of co-informational sentences in belief contexts’. The principle is:

If the information that S is identical to the information that S1, then someone believes that S if and only if he or she believes that S1.

Salmon uses the expression ‘the information that S’ to mean the content of the belief expressed by S. It is true that Salmon has a different conception of content from Frege’s. Although the difference is very important for many

¹⁸ Salmon (1986) pg. 80

discussions and will be addressed as such later on, it is not relevant for the discussion here. For what I want to examine here is only the alleged connections between belief ascriptions, Frege's Puzzle, and the notion of belief contents (however that notion is to be explicated).

Salmon says that this principle is "virtually a logical consequence of the idea that the object or content of a given belief ... is a piece of information and that a sentence encoding that information thereby gives the content of the belief". I do not know exactly what Salmon means by "virtually" here, but it is certain that the substitutivity thesis is not a logical consequence of these ideas. That is it not a consequence is clear if we are again mindful of the distinction between (1), (2) and (3). For 'the idea that the object or content of a given belief is a piece of information' is a thesis in philosophy of mind or metaphysics, and the idea that 'a sentence encoding that information thereby gives the content of that belief' is a thesis in the semantics of simple sentences. Salmon would require an assumption about the structure of belief ascriptions to forge a connection between the views in metaphysics and simple sentence semantics. In particular he would require the assumptions 1) that a belief ascription relates an agent to the content of his belief and 2) that the content of 'that S' is the content that 'S' normally expresses. These assumptions are not inevitable.

I want to go further now though, and question whether, even if we grant these assumptions, we automatically have a simple procedure about how to individuate belief contents. I think that perhaps the answer to this could be 'no'.

Before I explain why I think that the answer may be 'no', I want to say why this is so important. It is on the basis of his substitutivity thesis that Salmon goes on to draw an important methodological conclusion. He says that insofar as the substitutivity thesis is accepted as a plausible principle concerning pieces of information contained in a sentence and the content of a belief, we have an important procedure for establishing that two given pieces of information are distinct. He says: "One may simply rely on our

ordinary everyday criteria, whatever they happen to be, for correctly saying that someone believes or knows something or does not believe or know it. We do not have to be able to specify these criteria, we need only to be able to apply them in paradigm cases..." I take Salmon here to be making a methodological point. It is plausible to assume that Salmon thinks that looking at our ordinary everyday criteria for reporting beliefs is an "important procedure" for establishing distinctness of belief contents because of the advantage that would have a pre-theoretical way of establishing distinctness of beliefs, pre-theoretical because one would only need to know when we would attribute certain beliefs without additionally having to know how we do this.

One could doubt whether our ordinary belief ascriptions could be expected to give us the relevant notion of thought-content. Of course, if we start off with a theoretical notion of content, suitable for solving Frege's Puzzle, or governed perhaps by other theoretical constraints, and assume that our ordinary belief ascriptions are targeted on relating agents to contents appropriately individuated, then there can be no gap between our ordinary belief ascriptions and a philosopher's belief ascriptions. However, if we take our ordinary belief ascriptions as the primary data from which we draw conclusions about belief content individuation, then we have to face up to the possibility that we might be unable to extract any conclusions about content individuation that meet various theoretical strictures.

One way to argue that the philosopher's theoretical notion of belief content is different from our everyday concept of belief implicated in our ordinary belief ascriptions would be to show that there is a constraint on one that is not a constraint on the other. I suggest that there is a constraint on the use of philosophers' use of 'believes' that is absent from our everyday use of belief ascriptions. Philosophers are guided by theoretical constraints that the person on the street may not be. Philosophers are guided, claims Gareth Evans (1982), by what he has called the Generality Constraint. According to this constraint, we (as philosophers) cannot attribute the thought that *a* is *F* to an agent unless we are prepared to attribute to the subject a capacity to

entertain indefinitely many thoughts expressed with the name 'a' and also indefinitely many thoughts expressed with the predicate 'is F'. It makes sense to think that philosophers of mind and language are governed by this constraint, at least once certain assumptions are made about the goal and proper form of a theory of language and mind. The fact that a subject can understand indefinitely many sentences of a language in which she is competent, including ones that she has never encountered before, must mean that those sentences are structured, otherwise one would need infinite resources to understand the sentences. If we assume, with Frege, that understanding a sentence involves associating with it the thought that the sentence expresses, this fact about the systematicity of language has its counterpart in thought. Thoughts will be then conceived of as having the same structure as the sentences that express them. Indeed, it might be said that the idea of systematicity is what gives sense to the idea that thoughts are essentially structured. From a theoretical point of view, it may be that the idea of explaining an agent's competence with language or, more generally his ability to think things, plays a role here. The idea of thoughts as structured plays a role in explaining an agent's capacities, since the idea would be that there would be a common explanation of an agent's being able to think that a is F and that a is G. There would be a single cognitive state whose possession would be a necessary condition for the ability to think both thoughts (Evans 1982). Once one assumes a certain function for sentences of the form 'A believes that S', namely relating an agent to a thought, conceived of as above, that is to say, functioning as predicates in a philosophical psychology, then the Generality Constraint will be a constraint on belief ascriptions. But, the Generality Constraint is, in and of itself a constraint on the concept of thoughts, or as Evans says, 'conceivings', and it may be the case that our ordinary belief ascriptions are not always to be so understood. I will not argue here that ordinary belief ascriptions do not function as predicates in philosophical psychology, but have merely tried to show that the assumption that they do is crucial to link what a philosopher might want to say about belief with what we might ordinarily say with sentences of the form 'A believes that S'.

I submit that understanding our belief sentences is a distinct project from developing a theory of thought.

In this section I have aimed to distinguish three separate issues, often insufficiently distinguished, and trace out their real connections. I have distinguished issues about the informativeness of sentences from issues about thought-contents and I have separated issues about thought-contents from ordinary belief ascription sentences. I have indicated what assumptions would connect the three separate topics, and I have suggested that these assumptions are by no means trivial. In particular, I have tried to correct the common view that Frege's puzzle and the substitutivity 'problem' are closely connected. One should not draw conclusions about substitutivity patterns from facts about informativeness of sentences. And if it is really ordinary discourse that we are interested in, one should be wary of assuming a philosophical theory of thought, and then further assume that ordinary ascriptions relate agents to thoughts (in the theoretical sense).

Relatively few philosophers today think that Frege's theory of simple sentences is correct. The dominant view regarding names and indexicals is that they contribute to the content of sentences containing them, not a Fregean sense, but the individual that they refer to (in a context, for indexicals). Nor is Frege's elegant theory of belief ascriptions widely held. Some of the objections about to be discussed in the next section account for the unpopularity of his theory of ascriptions. Nevertheless, some deep assumptions which have been discussed so far continue to be widely held. Now, having discussed Frege, I shall try to say what these assumptions are.

- First, there is the very basic idea that sentences express contents, and that giving a semantic account of a language involves allowing us to associate with an arbitrary sentence of the language under study its semantic content. That is to say, most semantic work still takes place within the 'translational' paradigm. In Section 2.5, however, I shall look at one account which rejects this paradigm. In Chapter 4, I shall try to present at least the outlines of a

positive account which also rejects the paradigm, and hopefully avoids some of what I see as the undesirable consequences of the theory examined in 2.5.

- There is still the idea that something like Frege's Puzzle needs a solution. I say 'something like' Frege's Puzzle, because now the question of 'informativeness' is not seen as having to do with the Fregean content of a sentence. The idea of informativeness may or may not be seen as a question in semantics. Where it is seen as a question in semantics, the semantic account of a sentence must be designed so that it is consistent with the insight that names and indexicals introduce their referents into the content of a sentence. But, it can be seen instead as a question in psychology. The idea is that a thought is informative only if it involves distinct representations of the objects and properties it is a thought about.
- The basic idea that a belief ascription is some kind of a relational statement is also preserved. Although Frege's simple idea that 'A believes that S' simply relates an agent to the thought expressed by 'S' is abandoned, the idea that a belief ascription somehow manages either to specify a representation, or to quantify over representations. The relevant representations are those consistent with the above-mentioned requirement that informative judgements require distinct representations.

I am about to look at four theories of belief ascriptions and focus on how they can be applied to Kripke's Puzzle. The first is the theory of Frege himself. The others are chosen for their apparent diversity. There is Nathan Salmon's theory, which I consider as an example of a direct reference theorist's take on Kripke's Puzzle. It is sensitive to some of the criticisms I make of Frege in the next section. In Section 2.5, I look at a theory developed by Joseph Almog and independently by Erin Eaker. Radically different though it (apparently) is, I will argue that it inherits the most significant problems of Salmon's account, and has other problems of its own. In Section 2.6, I look at an account which seems promising in that it

seems sensitive to some facts about belief ascriptions that Eaker and Almog's theory seems rather insensitive to.

Each of the above theories retains, I will argue, one very basic assumption, which runs deeper than the three assumptions mentioned above. In fact, Eaker and Almog's theory rejects each of these assumptions, and so is indeed fundamentally different from most theories and the other three I discuss in this chapter. Nevertheless, it retains this one assumption which I think is mistaken. Exactly what this assumption is will, hopefully, become clear in the course of this chapter, but it can provisionally be stated as follows: *there are certain context-independent cognitive facts about agents which are independent of each other and ontologically prior to belief ascriptions and which ground the truth of belief ascriptions. The function of a belief report is simply to describe these facts.*

3.3

What Frege might have said in response to Kripke's Puzzle

In this section I want to consider what Frege could have said in response to Kripke's puzzle. It will be seen that while Frege's theory has the resources to stop Kripke's argument to the seemingly paradoxical conclusion that Pierre has contradictory beliefs even though he is rational, it can only do so in a very counterintuitive way.

Frege could have stopped Kripke's argument after the point when Kripke disquotes Pierre's utterances about London in French by invoking a French language version of the disquotation principle to get 'Pierre croit que Londres est jolie' and then applies the translation principle to get 'Pierre believes that London is pretty'. To remind ourselves, the translation principle says that *if a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth in that other language*. Of course, what is meant is that a correct standard translation which aims at preserving literal meaning preserves truth, rather than *any* translation.

Frege's theory, taken strictly, would prohibit this use of the translation principle. Frege would have observed that 'Londres is London' is an informative identity sentence and the explanation of its informativeness would be essentially the same as the explanation of the informativeness of 'Hesperus is Phosphorous', that is to say, 'Londres' and 'London' would be taken as expressing distinct senses. If we assume that 'a correct standard translation which aims at preserving literal meaning' must preserve sense between translations, as would seem plausible from Frege's perspective, then we already have a restriction on the use of the translation principle. But, even if this were not the case in general, it would surely be the case for belief report sentences, for on Frege's theory the function of a belief report is to relate the agent to the content of his belief which is expressed by the embedded sentence. So, in translating from 'Pierre croit que Londres est jolie' to 'Pierre believes that London is pretty' we would be relating Pierre to two distinct thoughts, so not only would the translation not preserve sense, it would not even preserve the reference of all expressions in the sentence, and that must be a condition on translation. In this particular case, we would be moving from a true belief ascription in French, to a false one in English.

One quick objection to this Fregean strategy would be the observation that the response does not work to the argument leading to apparent paradox in all its forms. We saw in Chapter One that Kripke is careful to show that the translation principle is not essential to the derivation of the puzzle since the puzzle can also be created in a monolingual scenario. But whether or not this objection is satisfactory depends on how we read Frege. If we take Frege to have held that a proper name has a single sense which everyone who understands sentences containing it associates with it, then Frege's objections to translation would only apply to translations between languages.

It is a question whether or not Frege *did* hold this view¹⁹, even though he did say that in a perfect language a single name should have a single sense. But, whatever Frege's intentions, it is now often convincingly argued that senses, in the sense of 'a way of thinking' are almost bound to be idiosyncratic in the sense that different speakers can, and often will, associate different senses with names and still count as competent with a name (Richard 1988). For on any plausible way of understanding 'way of thinking' that is relevant to whether a given sentence is informative or not (we now have to add: 'for a given speaker') it seems implausible that all competent speakers would associate the same 'way of thinking' with a given name, even though there may be some, albeit rather exceptional cases, where there may be a well established convention of associating a particular way of thinking with a name, for example, 'Hesperus' being associated with the sense of 'the star that can be seen in such-and-such a position in the sky in the evening' and 'Phosphorus' being associated with the sense of 'the star that can be seen in such-and-such a position in the sky in the morning' (Kripke 1988).

If we do treat senses as idiosyncratic the original objection becomes relevant to more than just the bilingual form of argument in Kripke's paper, that is, to more than just translations between intuitively individuated languages, because differences in senses associated with given names yield, strictly speaking, different idiolects. The Fregean who allows that senses vary between speakers then could restrict the use of translation by saying that we must 1) count different idiolects as different languages, and 2) only admit translations which preserve sense.

This would have the consequence that the translation principle will apply even to monolingual scenarios since even when one intuitively regards others as speaking the same language as oneself one would be tacitly invoking a homophonic translation of their language into one's own. This new restriction is very restrictive and counter-intuitive. Not only would it mean that we cannot translate many names, it would also mean that in most

¹⁹ Joseph Almog (2005) argues that he may not have.

cases we would not know whether we are translating correctly or not since we often do not know what senses a speaker attaches to a name. Moreover, the restriction would affect, not only the applicability of the translation principle, but also the applicability of the disquotation principle, since that principle would now be revealed as implicitly assuming a homophonic translation principle.²⁰ For example, suppose Smith, who associates with the name 'London' the sense of 'the city where Buckingham Palace is to be found', says 'London has a population of seven million'. When the reporter ascribes the relevant belief to Smith, he must make sure that he associates the same sense with the name 'London' otherwise, on Frege's theory, he will be ascribing the wrong thought-content to Smith.

And even if, by luck, he happened to associate exactly the same sense with the name, that would still be insufficient for an actual belief report to be of any use, for while it may be a true report, it still would not *communicate the right information* unless the audience to the report *also* associates the same sense with the name.

So, it seems that if we read Frege in a way that allows his theory to restrict translation so that it not only blocks intuitively bilingual forms of argument, but also blocks the monolingual form, as it would have to do if it were to solve Kripke's puzzle in its more general form, we are left with some very counter-intuitive restrictions on translation, on disquotation and, most relevantly, on belief ascriptions in general.

I think that the restrictions on translation and disquotation are very worrying. I would argue against the restriction on disquotation even though I have argued in Chapter One, and will argue again, in more depth, in Chapter Four, that the disquotation principle is false. This is because, I think that, while strictly false, or not generally applicable, something like the disquotation principle is used as a principle which provides *prima facie* and defeasible evidence for a belief ascription. The problem with Frege's theory is that we

²⁰ Richard (1988) convincingly argues that if senses are idiosyncratic it will often be a matter of luck if we manage to report correct thought content.

would hardly ever be able to rely on it in practice even as a defeasible principle. So, while I too believe that disquotation should be restricted I do not think that it should be restricted in the same ways that Frege would, nor for the same reasons.

Throughout the discussion so far in this section, the focus has been on Frege's theory of language, especially for simple sentences. It is the idea that speakers associate different senses with names as their meanings and the idea that utterances of those sentences express the speaker's beliefs that has dominated the discussion. With this in mind, two things can be said in response. First, in Section 1.2, I urged a sharp separation between the semantics of what a speaker says, and what belief ascriptions are true of her. In Chapter One, I argued that an analogue of Kripke's puzzle can be created without assuming that an agent has spoken on a given subject. If this is correct, then, even if Frege's theory did not force problematic restrictions upon us, it would still not address Kripke's puzzle in its most general form. For in the case where an agent has not spoken on a subject, translation and disquotation do not apply. We can not appeal to any facts about the semantics of the agent's words to block the puzzle. In a way, this is just as well, since Frege's theory has become rather unpopular as a theory of the semantics of names.

The possible Fregean response looked at in this section has focused only on circumstances where an agent has spoken on a topic, and then it has been argued that the translation principle and the disquotation principle do not apply, so that we cannot derive an ascription of belief from what an agent has said. But we can ascribe beliefs to agent even if he has not spoken on a topic. The focus in this section has, then, so far been too narrow. The problem arising from Kripke's puzzle is quite general and so a solution that limits itself to blocking derivations from what an agent has said will not in general be adequate.

The problem with Frege's theory of belief ascriptions, I have claimed, is not that it does not have the resources to block even those derivations from what

has been *said*, but that that it seems that if we take Frege's notion of sense seriously then the majority of uses of disquotation and (homophonic) translation would be suspect. Similarly, when we try to apply Frege's theory to block versions of Kripke's Puzzle that do not depend on an agent's initial utterance, then we see that Frege's theory is, in general, implausible, and that its implausibility has nothing special to do with our not knowing what sense an agent attaches to a given expression. The difficulty has instead to do with the fact that we often will not know how an agent thinks of the objects and properties that he has a belief about. To remind ourselves, the 'that-S' clause of a belief ascription names, in Frege's theory, a thought. The question is: what thought is named in a given belief ascription? There seems to be a dilemma here. If, on the one hand, the thought is the one that the belief reporter associates with 'S', then it seems that a belief report that is intuitively true could count as false in many cases where the agent does not think of the relevant objects and properties in the same way as the reporter. Surely, not everyone of whom it is true to say 'A believes that London is pretty' thinks of London in the same way.

If, on the other hand, however, the thought referred in the embedded sentence of the belief report is stipulated to be the one that the agent is related to, then there is the puzzle of how this reference is achieved. One could say that the embedded clause is stipulated to have the right reference, but one would generally left in the dark as to which thought this actually is.

If belief ascription necessarily involved ascribing to an agent a Fregean thought then belief reporting would be much more difficult than it actually is. It would require us to know much more about the minds of others than we actually do.

3.4.1

Nathan Salmon's theory of belief ascription

In this section I will describe and evaluate Nathan Salmon's theory of belief ascriptions, which he consciously develops to deal with a puzzle which is very similar to Kripke's Puzzle about Pierre. His is a radical attempt to solve his puzzle within a direct reference framework and so avoids the criticisms that Frege faces due to his positing of senses as the meaning of expressions in simple sentences. Salmon's theory also avoids the main criticisms that I made of Frege's possible response to Kripke's Puzzle, specifically, the criticism that it implies unacceptable restrictions on translation and disquotation.

Salmon is a direct-reference theorist. He takes the content of a simple subject-predicate sentence consisting of a name followed by a predicate to be a singular proposition, so that the proposition expressed by 'Socrates is wise', for instance, consists of the man Socrates and the property of being wise ordered in the appropriate way. This is, of course, fundamentally opposed to Frege's theory, according to which, in place of the individual Socrates we would have a purely conceptual representation which semantically determined the individual Socrates as the referent of 'Socrates'.

On the other hand, while he is working within a fundamentally different framework for simple sentences, Salmon retains the assumption that the content of 'Socrates is wise' is the content of the belief that one expresses in competently and sincerely asserting the sentence, and that an ordinary belief ascription of the form 'A believes that S' is a relational statement, simply relating the agent to the content of 'Socrates is wise'.

On these assumptions, all a speaker does when she says 'London is pretty' is to ascribe a certain property to a certain place. We can accurately translate if we also utter a sentence which ascribes the same property to the same place. If she sincerely and competently asserts 'London is pretty', then we can disquote her, since homophonic translation is unproblematic. Words are

treated, in Salmon's theory as public shared items available for us to use, not merely to mention, in ascribing thoughts to others. These are attractive consequences since they are more in keeping with our ordinary practices of translation and disquotation as they seem to be.

Since Salmon's theory does not appear to have the resources to block the argument that Kripke uses to derive his conclusion that 'Pierre believes and London is pretty' and 'Pierre believes that London is not pretty' are both true of a rational Pierre, one difficulty that Salmon faces is to explain the apparent conflict, for we do, on Salmon's view ascribe contradictory beliefs to Pierre. The singular proposition that London is pretty is the contradictory of the singular proposition that London is not pretty. I will return to how Salmon aims to resolve this apparent conflict in the context of discussing his own puzzle, which raises another apparently even more puzzling difficulty closely related to this one.

Another difficulty is that Salmon's theory has the consequence that if 'Pierre believes that Hesperus is shiny' is true then so is 'Pierre believes that Phosphorus is shiny', since both ascriptions relate the same agent to the same thought by the same relation, and that is all there is to the truth-condition of the report, on Salmon's view. Salmon's theory implies a thesis of substitutivity for coreferential names in the embedded sentence of belief reports. This is thought to be a bad result, because there are robust intuitions that belief reports that differ only by different coreferential names in the embedded clause can differ in truth value. Actually, I think that this problem is a much more serious problem than it is often taken to be. In fact, Kripke's arguments in 'A Puzzle About Belief' are often taken to show that assuming that substitutivity is false is not as compelling as it is often taken to be. I will now digress a little to discuss this because it is of great importance to decide to whether or not substitutivity is true for belief ascriptions, and this is an issue that will remain relevant throughout the dissertation.

Digression on the status of substitutivity failures

There are several arguments in the literature which are presented as supporting the idea that substitutivity fails in belief ascriptions. There are then other arguments which try to show that these arguments (the main argument in Kripke's 'A puzzle about belief', in particular) are not so compelling as they seem to be. If our belief in the failure of substitutivity required these arguments to support it then any counter-arguments would affect our belief. However, I do not think that we require these arguments. I will return to explain this, but first I will consider the arguments often given.

The structure of the argument that philosophers find in Kripke for this conclusion that perhaps substitutivity failures are not so obvious as has been supposed is as follows. Reductio ad absurdum arguments can be given against the idea that substitutivity is true. This is prima facie support for the view that substitutivity is false. However, when we reflect on these arguments we see that they invoke at least one other principle (specifically, the disquotation principle) which, along with other compelling assumptions, can be used to derive a paradoxical result, even without the use of substitutivity. The idea is then that it is wrong to blame the paradoxical result on substitutivity. The reductio arguments do not then work against the conclusion that substitutivity is not true. The prima facie support for substitutivity vanishes.

It is common to find in the literature two other arguments for the failure of substitutivity. One argument relies on a strengthened biconditional version of the disquotation principle. It is clear that an agent Tom can assent to 'Hesperus is shiny' and fail to assent to 'Phosphorus is shiny' even though he is sincere and competent, so we can conclude, using the biconditional DP that 'Tom believes that Hesperus is shiny' is true and that 'Tom does not believe that Phosphorus is shiny' is also true. If we take the latter as the negation of 'Tom believes that Phosphorus is shiny' then we can take the latter ascription as false. Another argument again uses the simple disquotation principle to conclude from Tom's assent to 'Hesperus is shiny'

that ‘Tom believes that Hesperus is shiny’ is true, and from Tom’s assent to ‘Phosphorus is not shiny’ to ‘Tom believes that Phosphorus is shiny’ is false, on the additional assumption that Tom is rational and would therefore not believe that Phosphorus is shiny.

Kripke’s puzzle casts sufficient doubt on the assumptions used in the above arguments to make us wonder whether we could find these *arguments* compelling. However, as I said above, I do not think that we require these arguments because the assumptions on which they rely are more questionable than the conclusion they argue for, more questionable even before we consider Kripke’s puzzle. I think we should take as a hard *datum* that ‘Tom believes that Hesperus is shiny’ can be true in some contexts even while ‘Tom believes that Phosphorus is shiny’ is false in the same contexts.

When I say we should take substitutivity failures as a datum I mean we should reflect on particular examples which provide us with the firm intuition that interchange of coreferential names can result in a change in truth-value. I provide such examples in Chapter Four. We do not need arguments for the failure of substitutivity. We only need intuitions about particular cases, *The failure of substitutivity is therefore a datum that needs explaining*. That is the end of my digression on the status of our belief in substitutivity failures.

Salmon offers a new analysis of belief ascriptions which is designed to resolve a puzzle that he invented. In Salmon’s puzzle, an agent, Elmer is determined to apprehend Bugsy Wabbit, a notorious jewel thief. Elmer learns much about Bugsy from FBI files, including photographs, films and slides, and based on what he sees and hears he forms the opinion (on January 1st) that Bugsy is dangerous. On June 1st Elmer receives another piece of information which makes him less sure that Bugsy is dangerous. Salmon asks us to consider two questions: (a) Before June 1st did Elmer believe that Bugsy Wabbit is dangerous? (b) If so, does he continue to believe that after receiving the extra information on June 1st? Salmon says that the questions

should be answered ‘yes’ and ‘no’ respectively and that the Elmer’s cognitive states are paradigm examples of such states. Given what we have so far, these claims cannot reasonably be denied. But things are not as clear as they seem.

Salmon elaborates his story to include the following information. Shortly after January 1st Bugsy learns that he is being pursued, and in order to avoid detection he undergoes extensive plastic surgery. He also has his voice surgically altered and adopts different mannerisms. In short, he becomes unrecognisable. He does, however, retain his name. Elmer finally, on April 1st, catches up with Bugsy after these alterations, makes friends with him, but never learns this man’s true identity. Elmer sees some of Bugsy’s behaviour and in particular his interactions which another man who appears to be frightened of him and comes to think to himself “I’d better watch my step, this Bugsy is a dangerous fellow”. On June 1st, as we have already been told, he receives the extra FBI information which leads him to say to himself “Maybe Bugsy is harmless after all”.

The puzzle arises when we consider the question ‘Once Elmer has received the extra information on June 1st, how does he stand with respect to the information that Bugsy is dangerous?’ Does he believe it or not?

Focus first, though, on Elmer’s beliefs on April 1st. If we roll the clock back to before April 1st, Elmer certainly had the belief then that Bugsy is dangerous. The reasoning that Salmon gives that makes this seem compelling is that we had already decided, before we learnt the extra part of the story, that Elmer believed that Bugsy was dangerous, but all the additional information concerns events that take place after January 1st, so the original grounds for saying that Elmer believed that Bugsy is dangerous still obtain. On April 1st, Elmer formed the opinion that his friend Bugsy (whom he does not recognise) is dangerous. So we can say that he never changed his mind about Bugsy in the sense that he believed him to be dangerous and still does. There was never a point, before June, at which he gave up this belief. Now, when we turn our attention to the puzzle question

about what Elmer believed on June 1st, we must admit that there seems to be no simple satisfactory answer to the question. For there is an important sense in which Elmer came to believe, in January, that Bugsy is dangerous but now suspends judgement about this. So, there is an important sense in which we want to say that on June 1st. Elmer neither believes that Bugsy is dangerous nor fails to believe it. But it seems completely unsatisfactory and misleading to leave matters there, for there is a very compelling reason to say that Elmer still believes that Bugsy is dangerous. As Salmon says, something exactly analogous to the grounds for holding that Elmer continues to believe that Bugsy is dangerous on April 1st. also obtains on June 1st. For Elmer has not relinquished the opinion that his friend Bugsy is dangerous.

On the other hand, if Elmer had decided on January 1st that Bugsy is dangerous and came to have second thoughts, as he actually did, but had never met Bugsy in the interim and come to have thoughts about him, then we would have no hesitation in saying that Elmer once believed that Bugsy is dangerous but believes it no longer. That is precisely what we did say when we had only the first part of the story. Salmon then argues in a way very reminiscent of what I emphasised was so structurally critical in Kripke's argument. He says that all the information given in the first part of the story was enough to determine that Elmer no longer believes, by the 1st of June, that Bugsy is dangerous, and so that nothing else can undermine this, and then draws on a simple logical principle to validate this: If S entails T, then so does (S and S').

So we seem to have compelling reasons for denying and asserting that Elmer believes, as of June 1st, that Bugsy is dangerous, and yet the story which gives rise to this *is not itself inconsistent*.

Salmon's response to this puzzle is to say that we should in fact say that Elmer believes that Bugsy is dangerous all along. It is not true that Bugsy does not believe this, so that we avoid a contradiction in our belief ascriptions. Salmon is then faced with explaining our inclination to say that Elmer did not believe that Bugsy was dangerous on June 1st, by rejecting the

apparently compelling grounds for denying him this belief and explaining why those grounds seemed compelling.

He does this by providing a new analysis of belief and belief ascriptions. He suggests that 'believes that' may be analysed in terms of a notion of disposition to inward assent to a proposition (conceived in the direct-reference way explained earlier) when taken in a certain way. The cognitive state underlying the truth of our belief ascriptions is a three-place relation holding between an agent, a propositional content and a way of taking that proposition. English does not have an expression expressing this relation, so Salmon invents one, 'BEL'. Salmon then treats the logical form of a belief ascription 'A believes that S' as the existential generalisation of BEL, with the quantifier ranging over 'ways of taking propositions':

'A believes that S' is analysed as 'There is an x such that: A grasps S by means of x and BEL (A, S, x).

Since 'BEL' is a theoretical concept, the terms in this analysis need explaining.

Salmon takes over the Fregean notion of grasping, although his understanding of what grasping involves must be different from Frege's, since his notion of that which is grasped is so fundamentally different from Frege's. Salmon does not give an analysis of grasping in more basic terms, and so it is to be taken as a primitive concept and understood as being a necessary ingredient in thinking a thought. A central epistemological thesis of Salmon's is that one can grasp a proposition twice over, but fail to recognise the proposition as being the same one on the two occasions. The notion of recognition here is basically the same as, or at least derived from, the ordinary notion of recognising a material object. The singular propositions that Salmon thinks that belief is a relation to are propositions which involve material objects as constituents. One can then fail to recognise a proposition because one fails to recognise the material object which is a constituent of it. 'Ways of taking' propositions can then be

understood as (roughly) analogous to appearances of individuals. So one can grasp and think a thought without recognising it as the thought one has already entertained if one fails to recognise it, and one can fail to recognise it because thoughts, rather like people, have different appearances.

Salmon does not give necessary and sufficient conditions for grasping thoughts, but he does say that Elmer is in a position to grasp the thought that Bugsy Wabbit is dangerous. Elmer grasps the proposition that Bugsy Wabbit is dangerous when he takes it one way and inwardly assents to it when he takes it that way. The way in question here is the way he thinks of Bugsy when he thinks of him as the man he overhears in a dispute about carrots. This is taken to be a basic cognitive fact about Elmer that, according to the analysis of belief ascriptions provided above, grounds the truth of the report 'Elmer believes that Bugsy is dangerous'.

The fact that there is also a way of taking the proposition, such that when Elmer takes it that way (that is, the way he takes it when he thinks of Bugsy as the source of the FBI files) he does not inwardly assent to it, is, Salmon says, simply irrelevant to the truth or falsity of the belief ascription. What, then, about the apparently compelling grounds we have for denying this ascription? Salmon says that what we really mean when we deny that Elmer believes that Bugsy is dangerous can only be perspicuously represented by invoking the three-place relation BEL. He says that 'Elmer does not believe that Bugsy is dangerous' is what we say when what we really mean to deny is that the BEL relation holds between Elmer, the proposition that Bugsy is dangerous and the relevant way of taking the proposition. But, says Salmon, given only the two place predicate 'believes' we are forced to negate what is really the whole existential generalisation. It is strictly speaking false, but the closest we can get to what we really want to say when we restrict ourselves to the two place 'believes'.

Salmon also uses these theoretical resources to explain away the apparent difficulty with accepting a substitutivity thesis for names in belief ascriptions. Salmon is committed to the view that 'Tom believes that

Hesperus is shiny' and 'Tom believes that Phosphorus is shiny' state the same fact, given that Hesperus and Phosphorus are one and the same. If they state the same fact, then they must have the same truth-value.

The real difference between the two sentences is a pragmatic, rather than semantic one, according to Salmon. Suppose Tom would assent to 'Hesperus is shiny' but dissent from 'Phosphorus is shiny', then the first ascription can convey, although it does not literally say, how Tom inwardly assents to the relevant thought. The idea is that the two distinct sentences, 'Hesperus is shiny' and 'Phosphorus is shiny' expressing the singular proposition correspond to two ways of taking the proposition, and that the use of 'Phosphorus' instead of 'Hesperus' in the belief ascription conveys the information that Tom inwardly assents to the proposition when it is presented through the lens of the sentence 'Phosphorus is shiny', and in so far as it does this it is misleading, but not strictly false.

Salmon is very brief in applying his theory directly to Kripke's puzzle. But I will briefly summarise how he intends to apply it. According to Salmon, 'Pierre believes that London is pretty' is true and so is 'Pierre believes that London is not pretty'. He maintains that the reports are true together because they simply relate Pierre to the singular contents <London, is pretty> and <London, is not pretty>. Salmon thinks that these facts do not impugn Pierre's rationality because, at the appropriate level of analysis it would be revealed that Pierre has different ways of recognising London, but that these are not relevant to the truth of the belief ascription which by existentially generalising over the distinct representation washes out the specific information about Pierre inner representations.

3.4.2

Criticism of Salmon's theory

I do not think that Salmon's analysis can be correct. I think that the most serious objection to Salmon's theory is one that has standardly been made

against it: it has consequences which conflict with our intuitions about the truth-values of ordinary belief ascriptions. As we have seen, it implies a principle of substitutivity for belief ascriptions of the form 'A believes that a is F' with respect to the position occupied by 'a', when 'a' is a name or other directly referential expression. It is highly counter-intuitive to say that 'Tom believes that Hesperus is shiny' and 'Tom believes that Phosphorus is shiny' cannot differ in truth value at a given time. Moreover, with respect to Elmer's situation and Pierre's situation, Salmon's theory has the consequence that 'Elmer believes that Buggy is dangerous' and 'Pierre believes that London is pretty' are straightforwardly true with respect to the scenarios described, and that even in a context when we are focusing on Elmer's indecision about Buggy, and Pierre's revulsion at the ugly part of London he finds himself in, we can truly say: 'Elmer believes that Buggy is dangerous' and 'Pierre believes that London is pretty', even though he would adamantly deny that the sentence 'London is pretty' is true.

Even stronger conflicts with intuition would arise, it seems, if we were to consider belief ascriptions involving identities, such as 'Tom believes that Phosphorus is Hesperus' which intuitively could surely be false even while 'Tom believes that Hesperus is Hesperus' is, of course, true. And conflicting intuitions may be stronger still in the case of second-order belief ascriptions such as 'John believes that Tom believes that Hesperus is Hesperus'²¹.

According to Salmon's theory, this sentence expresses that John stands in the belief relation to the singular proposition about the proposition that Hesperus is Phosphorus to the effect that Tom believes it. Exactly the same fact is expressed by 'John believes that Tom believes that Hesperus is Phosphorus'. But surely, intuition says, the former could be true when the latter is false.

It is plausible to consider that such results would count as a *reductio ad absurdum* of a theory of which they are consequences. Salmon is, of course,

²¹ Salmon (1986)

aware of the counter-intuitiveness of these results, and devotes a large section of his book in trying to blunt the force of these intuitions.

We should remind ourselves that our objects of analysis are ordinary language belief ascriptions and intuitions about truth-values are generally taken as the data against which we test our theories. Any philosopher would agree that ordinary intuitions about the truth-values of belief ascriptions which describe aspects of ordinary situations are relevant, and sometimes the only real data we have to test our philosophical theories. But what exactly is the nature of this relevance? It would be too simplistic to say that what we find it natural to say about a given circumstance must always be true, because we know that we can always make mistakes. Salmon says, in the context of making a distinction between theory and data, that “the truth values of our pronouncements are not pre-theoretic data” and that “pre-theoretically, all we have is that we speak this way”. He goes on to say “only after we have decided on one theory can we determine the truth-value of our pronouncements”.²²

Salmon seems to me to draw the line between theory and data in the wrong place. It is methodologically more usual to take as a datum the assumed truth of our ordinary pronouncements. We do not assume that we are infallible of course, but it does seem that when we are considering cases which are central cases of a given phenomenon and we take ourselves to be speaking literally, non-elliptically, carefully and reflectively then we must assume that we are right in the majority of cases. For if we did not adhere to this rather minimal methodological principle it would no longer be clear that we were really interested in this aspect of our ordinary language after all. Or is there an even more minimal principle which can still allow us to see what we are doing as investigating ordinary language? Perhaps a philosophical theory designed to explain some aspect of ordinary language could be acceptable even if it had as a consequence that we are strictly wrong about what we say most of the time, if it can give a theoretically satisfying account of some

²² These quotations are from Salmon (1986) pg. 119.

core uses of a given bit of language and then additionally explain, in a non ad hoc way, how we can systematically account for the errors.

Even if this were the right methodological principle, I do not think that Salmon's theory passes this test. For Salmon admits that we often need to specify how an agent takes a proposition. Less theoretically, we often need to know how an agent thinks of an object. It is commonly held that we need to have something like an agent's view on an object if we want to know how the agent will behave in a specific situation, if we want to assess an agent's rationality from their beliefs, or if we want to know what an agent could be expected to say on a given subject. These are some functions that belief ascriptions would normally be thought to be designed to serve.

If it is correct that belief ascriptions have the above as amongst their main functions, it seems strange that belief ascriptions semantically only relate agents to propositions individuated in a way that doesn't allow them to serve this purpose.

It is commonly accepted that English has a locution which relates an agent to an object without specifying how an agent thinks of that object, for sometimes we may be interested only in knowing what object an agent has a belief about. This may be the case when the object of the belief is the topic of the conversation, rather than the agent being the topic of the conversation. These ascriptions are the so-called *de re* ascriptions, and are standardly supposed to have the surface form 'A believes, of a, that it is F'. On a popular understanding of these utterances, these *de re* ascriptions characterise, rather than fully specify a content believed, in the sense that they say something about the proposition believed without saying exactly which proposition it is. Apparent evidence for this idea that *de re* ascriptions merely characterise, rather than specify a content comes from the observation that the term that occurs in the place of 'a' is open to

substitution and existential generalisation because it occurs outside the context under the scope of the operator 'believes that'.²³

What seems strange is that even if we accept that there are genuinely de re ascriptions which merely characterise a content, Salmon's theory leaves no room for there to be a semantically significant difference between ascriptions which characterise and ascriptions which specify a content, as I shall now explain.

On Salmon's theory, the distinction between de re ascriptions which merely characterise a content and belief ascriptions which fully specify a content collapses in the cases where it is a directly referential term replaces 'a'. Salmon denies that de re ascriptions merely characterise a content. He thinks that they fully specify a content which is a singular proposition. For while it is often said that, in a de re ascription, the term 'a' occurs outside the context embedded under 'believes that', Salmon points out that although this is true, what is more relevant is that another singular term, a directly referential one (an anaphoric pronoun in ordinary locutions, and a bound variable in semi-formal existential generalisations), occurs *within* the scope of 'believes that'. Consider a semi-formal existential generalisation: 'There is an x such that: x = a and A believes that x is F'. If this sentence is true, then it follows from basic semantic principles that the open sentence 'A believes that x is F' is true under the assignment of 'a' to the variable 'x'. Given Salmon's simple relational analysis of belief ascriptions, the above open sentence is true if and only if Tom stands in the relation of belief to the content of the open sentence 'x is F'. Since the semantic content of a variable under an assignment is the individual assigned, the semantic content of the open sentence 'x is F' is the singular proposition expressed by 'a is F'.

If the claims and reasoning of the last five paragraphs are correct then there seem to be the following consequences. On Salmon's theory, there is no distinction between de re ascriptions and content specifying ascriptions

²³ In the final chapter I will try to cast some doubt on the idea that there are real de re ascriptions outside of artificial philosophy contexts.

when the embedded term is directly referential. This means that there is a redundancy in ordinary English of the locution ‘A believes, of a, that it is F’, since it says the same as ‘A believes that a is F’. Moreover, while we have two locutions that say the same thing, we have no locution which semantically does what Salmon, by his own admission, says we often want a belief ascription to do, that is, say how an agent thinks of an object. So, with ‘A believes that a is F’, we have a linguistic form which is redundant and insufficiently expressive, and conflicts with our ordinary intuitions about truth-values.

It would seem very plausible that if 1) ‘believes that’ functions in the way that Salmon says it does, failing to provide any information about an agent’s way of thinking of an object, or at least giving some kind of information about the agent that is not available from a merely *de re* report and that 2) we often need to know how an agent thinks of an object, then English would already contain some expression analogous to the BEL operator, but it does not, and that is why Salmon had to invent the operator. This seems to suggest that belief ascriptions are not to be understood in the notion-neutral way that Salmon suggests.

These considerations together provide at least a strong *prima facie* case against Salmon’s analysis of belief ascriptions. What then is to be said in favour of Salmon’s analysis?

It seems that Salmon thinks that one of the main considerations in favour of his theory is that it preserves what he says is “an intuitively appealing picture that is entrenched in philosophical tradition.”²⁴ This is the picture that depicts belief as a type of inward assent to a piece of information, a content. I would suggest that the intuitive appeal of such a picture derives more from a picture in which propositions are conceived along Fregean lines, rather than as singular propositions.

²⁴ Salmon (1986), pg. 80

Salmon's analysis starts from a certain notion of content according to which contents are singular propositions. This analysis of content is informed by considerations only having to do with the truth-conditions of simple sentences. Salmon then assumes, without any argument, that belief is a relation to such contents, and then further assumes that ordinary belief ascriptions specify this relation. These are exactly the moves which I tried to suggest are not inevitable in Section 1.2 of this chapter. We have now seen some specific reasons why the moves should not be made so quickly.

There is another criticism to be made of Salmon's theory, which is that it seems to be unable to account for the contextual-sensitivity. I will discuss this criticism in the context of discussing the next theory since the same criticism arises for the next theory I shall discuss and the advocates of the next theory have more to say on this point.

3.5.1

A more radical no-content theory: Erin Eaker and Joseph Almog

I shall look at a theory of belief ascriptions which abandons the orthodox assumption that belief ascriptions relate, in one way or another, agents to some kind of inner representation. In recent years, an alternative theory of belief ascriptions has been defended by Erin Eaker (2002 and 2004), and separately by Joseph Almog (2005). In their theory, they abandon the orthodox assumption that belief reports relate agents to any kind of inner representation or content. In this section I describe their theory and explain how radically different it is from orthodox views, and I will discuss some of its appeal, then I will discuss how they relate it to Kripke's Puzzle, before criticising the theory by drawing out some of its undesirable consequences.

Strictly speaking, although Almog and Eaker have developed views on belief ascriptions that means that their ideas can be grouped together as belonging to a radically new way of thinking about belief ascriptions, Eaker's work has been more narrowly focused on belief ascriptions, while

Almog's work has focused on developing a new program for giving a content-free semantics for a range of sentence types. Almog looks at belief ascriptions as part of this program, and develops view about what he calls 'co-ordinative reports' which are attitude reports of the form 'A and B both think that a is F', or 'A thinks that a is F and B thinks that a is G'. It is Eaker who gives a more detailed account of the truth-conditions for belief ascriptions of the form 'A believes that S', and so it is her definition that I will discuss, but in situating her work in its contexts I will be using ideas that Almog has developed.

According to Eaker (2004),

A belief ascription of the form 'A believes that a is F' is true iff A has, in thinking about the individual which is designated in the report context by the reporter's use of the singular referring expression 'a', ascribed to it the property of being F

The first thing to note about this truth-condition is that, according to it, the belief report does not relate a thinker to a content or any kind of inner representation. Not only does it not specify a relation to anything like a Fregean thought, it does not even relate the agent to a de re content. The truth-conditions are, as Eaker calls them, 'de re truth conditions', but this should be understood as meaning that the function of the report is to relate the agent to the individual she thinks of, and specifies what property she ascribes to it, and that is all.

One way to appreciate the nature of this semantic account of belief ascriptions is to see how the account, at least in Almog's hands, is part of a much larger semantic project of providing a new paradigm for doing philosophical semantics. The project entails rejecting the classical form of semantic theory which Almog says is 'translational' in the sense in such a theory each target sentence of the language under investigation is assigned an entity articulating 'what it says' or its 'content'. Almog instead intends to replace translation with non-translational truth-conditions. The truth-conditions provide the condition in which the sentence is true, but do not articulate what the sentence says. Thus liberated from having to articulate

what is said, the semanticist can, in stating the truth-conditions, use concepts which the user of the sentence may not even have. Almog says that the truth-conditions are given by what he calls (following Donnellan 1974) ‘the omniscient observer of history’ (henceforth OOH).

The metaphor of the OOH is central to Almog’s picture of semantics so it is worth explaining what is involved. Almog says the OOH articulates the truth-conditions of the target sentence. This terminology is a bit misleading, for articulating the truth-condition clearly does not require omniscience. Omniscience would be required to know the truth –*value* of the sentence, but not to know its truth-*condition*. If it did we could never do semantics. What is meant, I think, is that the OOH has the conceptual resources to give the semantics that the ordinary language user need not have to use the sentence. The OOH will use such concepts as ‘causal chains of reference-preserving links’, ‘speakers intentions’, the ‘agent of the context’. The conceptual advantage the OOH has over the ordinary language user is less apparent in the case of the semantics for belief ascriptions, since the truth-conditions for these sentences does not seem to invoke any theoretical concepts. In fact, we will see shortly that Eaker is committed to the view that her notion of ‘thinking about’ and ‘ascribing a property to a thing in thought’ are to be understood in an ordinary way. Nevertheless, the truth-condition does not articulate what is said. The OOH’s language then is richer than the ordinary language user’s since it contains extra theoretical resources. On the other hand, there are limitations on what the OOH can say. He cannot use empty names and cannot use definite descriptions referentially which there is no referent, he can only mention them in mentioning the sentence whose truth-condition he is giving.

3.5.2

The Application of Eaker’s and Almog’s account to Kripke’s Puzzle

Eaker’s response to Kripke’s puzzle is interesting because it is one of the few (or only) that questions the assumptions that belief reports relate thinkers to a content.

Kripke begins his presentation of the puzzle by stipulating that the only reading of 'believes-that' reports that he is interested in are de dicto reports which give the content of an agent's belief. He specifically says that he is not interested in de re reports (or de re readings of 'believes that' sentences). Eaker's response is important because it reminds us that the reports in questions are de dicto is an assumption that must be seen as defeasible in the discussion that the paper engenders. It is no less a defeasible assumption just because Kripke stipulates that it must be true. We should question whether it is even methodologically legitimate to make such a stipulation while simultaneously relying on ordinary intuitions about truth-values to guide us, for to do so could be seen as entirely begging the question, in the context of a debate about the semantics of ascriptions, in favour of a certain analysis.

Eaker, in effect, sees Kripke's puzzle as a reductio of Kripke's assumption that the belief ascriptions involved are to be understood as specifying a de dicto content.

Consider the following four sentences:

- 1) Pierre believes that London is pretty
- 2) Pierre believes that London is not pretty
- 3) Pierre believes that London is pretty and Pierre believes that London is not pretty
- 4) Pierre believes that London is pretty and that London is not pretty

Eaker's account first of all commits her to the truth of (i) and (ii), since the truth conditions for (i) and (ii) are, respectively, 'Pierre, in thinking about London has ascribed to it the property of being pretty' and 'Pierre, in thinking about London has ascribed to it the property of being not pretty'. Intuitively, Pierre has fulfilled both these conditions. If (i) and (ii) are both true then so is (iii), since this is simply derived from (i) and (ii) by the law of conjunction. It is interesting to note that neither Eaker nor Almog comment on the truth-value of (iv). But even without any view on the truth of (iv)

Eaker's account endorses those reports whose truth has seemed to be incompatible with Pierre's rationality.

According to Eaker, (iii) need not ascribe any cognitive flaw to Pierre, and is perfectly compatible with his rationality. The explanation that she and Almog give is essentially as follows. First they claim that description that Kripke himself gives of Pierre's case, that is, the description that Kripke gives from an external OOH viewpoint *by itself* grounds the truth of the reports. The viewpoint is external in the sense that it does not involve ascribing and contentful mental states to Pierre. It avoids making any explicit assumptions about Pierre's inner mental representations and indeed involves no explicit commitment to the claim that there *are* any. The external description uses such expressions as 'Pierre is inclined to think that it is pretty', 'Pierre has not changed his mind', and 'Pierre is unimpressed so is inclined to assent to the sentence 'London is not pretty'. It is on the basis of such statements, and the rest of the story, which we can see is clearly compatible with Pierre's rationality, that we can say that Pierre has, in thinking about London ascribed to it the property of being pretty and the property of being not pretty. But this is all, say Almog and Eaker, that the truth of the reports depends on.

Kripke pre-empts a response to the puzzle which resolves it by saying that the externally given description of Pierre is intuitively compatible with his rationality. He says:

In one sense the problem may strike some as no puzzle at all. For, in the situation to be envisaged, all the relevant facts can be described in one terminology without difficulty. But, in another terminology, the situation seems impossible to describe in a consistent way.¹⁵

The first terminology referred is the external viewpoint description. The other terminology, in which, says Kripke, it is impossible to describe the situation consistently, is the terminology of 'believes-that' reports. But Eaker and Almog have given a way to understand this terminology so that

¹⁵ Kripke (1988, pg. 102)

the situation can be described in a consistent way. They do this by abandoning the de dicto content assumption and by making the semantics of the reports depend only on those externally given facts that Kripke admits are consistent with Pierre's rationality.

Almog (2005) adds the point that the conjunctive belief report (iii) does not ascribe conflicting contents to Pierre, because it does not ascribe any contents at all. He says that only if (iii) ascribed to Pierre a conjunction of thoughts of the form 'a is F and a is not F', where 'a' has a univocal content would the report be incompatible with his rationality.

Before turning to my criticism of Eaker and Almog's theory, I want to contrast it with Salmon's theory which was discussed in Section 2.4. I do not want to underestimate the differences between the two types of account, and I will discuss one very important difference which will form the basis of reason to favour Eaker and Almog's account over Salmon, but I will also argue that are more similar than might be supposed considering that that one relates agents to contents and one does not. For that *sounds* like a very large difference.

Eaker's and Almog's theory can sound very different from Salmon's if we focus on the way they account for the compatibility of the truth of (iii) with Pierre's rationality. Eaker and Almog say that there is no incompatibility because (iii) does not ascribe any conflicting thoughts to Pierre. Salmon's account, on the other hand, says that (iii) does ascribe contradictory thoughts to Pierre, specifically it ascribes to him belief in the singular proposition consisting of London and prettiness in the appropriate order and in the singular proposition consisting of London, negation and prettiness in the appropriate order. These two propositions are contradictories. This makes it appear that Salmon now has a further explanatory task. He must offer an account of how an agent can have contradictory thoughts despite being rational. Eaker and Almog seem to have no such task. Superficially, then, the two types of account seem very different. But I suggest that we should look at exactly what is meant by this talk of contradictory thoughts. For it is

not clear what is really at stake in an argument between Eaker and Almog on the one side and Salmon on the other side about whether or not belief ascriptions like (iii) ascribe contradictory thoughts.

Salmon says that having the belief that London is pretty is standing in the relevant relation to a singular content. Almog and Eaker say that it is not. But what does it mean to say that an agent stands in the relation of believing to a singular content?

On the one hand, Salmon gives an account of believing according to which it is analysed as a disposition to inwardly assent to an entertained proposition. It is tempting to visualise what this involves and this leads to us thinking of a proposition as an inner object and assenting to it as an episode that we could perceive. But Salmon rejects the idea that a singular proposition is literally any kind of inner object, and rejects a phenomenological account of believing. More clarity is had when we step back from this metaphorical way of thinking of belief and look at Salmon's semantic analysis of belief ascriptions. For Salmon ultimately thinks we can derive the thesis that belief is a relation to a singular content from an ordinary *de re* belief ascription of the form 'A believes, of a, that it is F', an assumption about the semantics of 'it' and a natural assumption about the grammatical structure of an ordinary belief ascription. This argument has already been given in Section 2.4, where I said that the distinction between a *de re* belief ascription which merely characterises a content and a belief ascription that fully specifies a singular content collapses on Salmon's theory. Salmon would presumably say that the *de re* truth-conditions that Eaker gives for 'believes-that' ascriptions actually specify relations to singular propositions. If 'Pierre, in thinking of London, ascribes to it the property of being pretty' is to be understood as paraphrasable into 'Pierre thinks, of London, that it is pretty', as seems natural, then Salmon will read this as a sentence which relates Pierre to the singular content that he thinks is expressed by 'London is pretty'. This move depends only on two semantic assumptions. Firstly, that anaphoric 'it' in the truth conditions is directly referential (an assumption with which Eaker and Almog would agree) and secondly, that 'that London

is pretty' is a term referring to the semantic content of 'London is pretty'. Assuming that Eaker and Almog would not disagree that the singular proposition exists, then we can see the difference between Salmon's view that (iii) relates an agent to contradictory contents and Eaker and Almog's view that it doesn't depend on any substantial difference in metaphysics, but on a different grammatical understanding of belief reports. If this is correct, then we can ask whether there is any further explanatory task that Salmon is faced with in accounting for Pierre's rationality that Eaker and Almog do not face.

Salmon postulates distinct mediating representations to account for how an agent can have contradictory thoughts. In doing this he can be seen as holding onto one entrenched Fregean assumption that I isolated in Section 2.2. These representations he calls 'guises' or 'ways of taking' propositions. He emphasises that they are not like Fregean modes of presentation which are integral to the semantic content of sentences and make up the thoughts that the sentences express. They may be psychological and subjective. Nevertheless, he says that they have conditions of individuation and are entities which are quantified over in the logical form of belief ascriptions. Why does Salmon postulate the existence of such entities? There seems to be a deep assumption at work here. It is the assumption that an agent can only have different attitudes to a given content only if he fails to recognise that he grasps the same content twice, and that this can only happen if he has different representations associated with the same content. Since, as we saw in Section 2.4, Salmon says that failing to recognise a content is parasitic upon failing to recognise an individual that is part of that content, let us focus on recognising an individual. Salmon thinks that if Pierre fails to recognise that Londres is London (as we might say) then he must have two ways of thinking about London, two distinct representations of London. Eaker casts doubt on this basic assumption. She says that even though we will often find differences between 1) how an agent who thinks that there are two entities when there is in fact just one is thinking of it when he is prepared to ascribe one property to it and 2) when he is not prepared to do so, we need not always be able to find such differences. All that need to be

the case is that the agent thinks there are two entities. Pierre-like puzzles can arise even when there is no difference in information associated with two names of the single entity, for the agent need have no names for the object. There also need be no information associated with two encounters of the entity, two pictures of it, or two presentations of the same picture. All that needs to be the case is that the agent thinks there are two objects, when there is just one.

Almog's and Eaker's account does not involve a commitment to distinct mediating representations which in Salmon's theory are quantified over in the logical form of the belief ascription. I think that this is a considerable advantage of their account. Nevertheless I do not think that their account is correct.

3.5.3

Criticisms of Eaker and Almog's theory

I will now examine some difficulties for their account. First, their theory licenses substitutions of coreferring names in belief contexts. We have already seen why this is a problem. But, not only does their theory imply a principle of substitution concerning names, but it also implies a substitution principle for any co-referring singular referring expressions. So, on their account, if 'John believes that Gordon Brown is no longer the Prime Minister' is true, then so is 'John believes that the current Prime Minister is no longer the Prime Minister'. We can acknowledge that there may be a de re reading of the latter ascription which is true, But Eaker and Almog do not say that. They would say that the sentence is true simpliciter. For they do not distinguish different readings of the sentence. Both Eaker and Almog give an interesting battery of examples to help to undermine the commonly held view that de dicto reports are to be taken as paradigmatic and de re reports are somehow deviant, or at least exceptional. But even if we are persuaded by their examples that de re reports are more pervasive than has traditionally been recognised, we still only have a reason to question the assumption that

de dicto reports are paradigmatic. We have been given no good reason to think that de re reports are paradigmatic, since the literature is already replete with examples which suggest that the de dicto reading (or, more accurately, the reading on which we would not accept substitutions as truth-preserving) is also pervasive. But Almog and Eaker do not merely suggest that de re ascriptions are paradigmatic, they completely ignore any other readings.

Another apparent difficulty with Eaker's truth-condition, although I do not mean to suggest that it is insurmountable, is that while the object language sentence 'A believes that a is F' will sometimes use a name that has no referent in place of 'a', the semanticist's truth-conditions will not make use of such names. Nor can the semanticist even use a definite description which fails to designate an individual, given that the predicates in the truth-condition are 'thinks about x' and 'ascribes to x the property F', for these predicates presumably fall into the class of predicates that Almog describes as 'existence-receptive', meaning that their meaningfulness requires the existence, at the time of application of the predicates, of the subject to which the predicates are applied. Eaker acknowledges the existence of such a problem in a footnote, but does not suggest how the account can be modified to deal with the problem. She only suggests that for an account of thought ascriptions involving non-existents that is suggested by our everyday practices, we should look at Almog's account. When we look at the relevant section in Almog, this is what we find.

Almog treats sentences of the surface form 'A believes that a is F' as essentially of the form 'a is believed by A to be F', in which 'a' is the logical subject. This way belief ascriptions are subsumed under his general truth-condition for sentences of the form 'N is F', where 'N' is a subject term, and 'is F' is any kind of predicate. The general truth-condition is:

A sentence of the form 'N is F' used a context c is true at a time t iff the source of the chain traced by the OOH leading to our use of 'N' at c is F at t.

Almog does not define ‘source’, but I think that the intuitive idea is that the source is object or event which stands at the beginning of a causal chain of that leads up to our use of an expression. He gives as an example, in a case relevant to a belief ascription, Aphrodite, the non-existent goddess, and the belief ascription ‘Aphrodite is believed to be in town’. Regarding the source in this case, he says “I do not have enough information about the real source of our use; I take it to be Greek tales of an irresistibly alluring divine being.”¹⁶ The tales are the source then. So ‘John believes that Aphrodite is in town’ is true just in case those Greek tales are believed by John to be in town. This cannot be right. Similarly, with the predicates that Eaker uses in her truth-conditions, we have ‘John believes that Aphrodite is in town’ is true just in case John, in thinking about those Greek tales, ascribes to them the property of being in town.

It is not merely that Almog has given a wrong candidate for the source here, for whatever the source is plausibly thought to be it will be something, in the majority of cases, which is very remote from the agent in space and time, and furthermore something the agent has no inkling of. If he has no inkling of it, how can he ascribe, in thought, any property to it?

I have argued that Eaker’s account, as it stands, is not adequate to deal with the whole class of potentially meaningful belief ascriptions, specifically, not with those with non-existent subjects, and that the suggested direction is not clearly the right one.

Another possible problem with Eaker’s account arises because of the precise wording she uses in giving the truth-condition for the belief ascription. To remind ourselves, she says that a belief ascription of the form ‘A believes that S’ is true iff A *has, in thinking about x, ascribed to it* the property of being F. It is the italicised part of the truth-condition that I wish to focus on here. What I want to draw attention to is the fact that in the belief ascription we have a verbal expression that has a very different logical grammar from

¹⁶ Almog (2004, pg.412).

the verbal expression in the truth-condition. While ‘thinks about x’ and ‘ascribes to x’ describe events that occur at specific points in time, or at least occur during specific intervals of time, ‘believes’, on the other hand, is a stative verb and expresses something that someone does not do at a specific time, or during a specific interval. Once this difference has been noticed, one might already be suspicious about whether a truth-condition which contains ‘has ascribed F to x’ could possibly be right for an ascription whose truth, on the face of it, has nothing to do with whether one specific kind of mental event has happened.

Strictly speaking, Eaker’s truth-condition has the consequence, for example, that John believes that Tony is bald if he has ever thought it, assuming that this is an acceptable (just more idiomatic) way of saying that John has ascribed the property of being bald to Tony, that it does not matter what properties he has ascribed to Tony more recently. Perhaps this response is correct, but superficial, since it might seem that Eaker could simply amend her truth-condition by simply adding the clause, ‘and A has not since changed his mind’. This, however, cannot be quite right because if the agent John has not changed his mind, he may simply have forgotten all about Tony and his baldness. In which case it would not be right to say that he believes that Tony is bald. Perhaps we could add to the truth-condition ‘nor has A forgotten about a and its F-ness’.

But it seems that adding qualifications to the truth-condition in this way misses the main point of the flaw in Eaker’s condition. The main flaw is related to the fact that there is the difference in logical grammar between the ascription and its putative truth-condition. To pre-empt a point that I will substantiate more fully in Chapter Four, and to connect back to a point I discussed in some detail in Chapter One, one consequence of using an event verb like ‘ascribes (to x)’ as giving the truth-condition is that it would seem to make the truth of a belief ascription depend on some one simple other fact about an agent in such a way that there is no conceptual bar on an agent having two conflicting beliefs ascribed to them in a conjunctive belief ascription of the type that is considered problematic in Kripke’s puzzle. This

is due to the fact that this fact (about what an agent has ascribed to an object) is, according to Eaker's account, *sufficient* grounds for the truth of the ascription, and so is not (and cannot be) made false by some other fact's obtaining. In particular, it cannot be made false by the fact that the agent, at some other point, ascribed a conflicting property to the object. In Chapter One, I said that it was this notion of sufficient grounds which is essential to Kripke's argumentative strategy. But why is this a problem? Eaker, after all, explicitly says that it is a virtue of her account that the conflicting reports are independent. In fact, she gives the following as a *constraint*: 'the truth conditions for each report must not be given in such a way as to contradict the possibility that the other report is true'¹⁷.

I think that this is a problem because there is good reason to suppose that the belief ascriptions that a reporter can apply to an agent are not always independent of each other. In Chapter One, I presented a principle called the Principle of Reporting Rational Agents according to which reporters cannot correctly apply to an agent who is presumed to be rational any ascription of the form 'A believes that S and that not-S'. If a reporter were to use such an ascription she would be trying to convey something about the agent's *irrationality*. If this principle were correct, this would mean that, in a Kripke-style case, even if we had grounds to say 'A believes that S' and grounds to say 'A believes that not S', we could never have grounds to say 'A believes that S and not S'. If we accept that belief reports are contextually sensitive, we can intelligibly say that 'A believes that S' and 'A believes that not S' can both be true at a given time but are never true in a single context, when A is a rational agent.

It is in my view a weakness of Eaker's account, and of Almog's, that their belief report semantics leaves out the contextual sensitivity of belief reports. For them, a belief ascription is made true by a single cognitive fact's obtaining, where the obtaining of this fact is not itself thought of as a contextually sensitive matter. For them, there must always be a univocal

¹⁷ Eaker (2004) pg.27

context-free answer to the question ‘Does A believe that a is F? If A has ascribed F to a in thought, the answer is ‘yes’, if not, the answer is ‘no’.

But that there may not always be a univocal, context-free answer to the question ‘does A believe that a is F?’ can be illustrated by considering the question as applied to Pierre in two different specific contexts. Suppose that there are two belief reporters, Bob and Carol, and that they both omniscient observers of history, so that they agree on all the facts pertaining to Pierre’s mental life, in particular, facts about what Pierre would say when presented with certain pictures, sentences, questions, and any other behavioural dispositions which might be thought relevant to knowing what an agent believes. Suppose that Bob is asked ‘Does Pierre believe that London is pretty?’ by an one of Pierre’s London neighbours who wants to know if Pierre is planning on staying in London, since he has noticed that Pierre looks a bit glum every time he steps outside his front door. Bob knows the purpose of the question and answers ‘no, he does not’. Carol is asked by someone who is responsible for designing the brochure that Pierre first looked at back in Paris ‘Does Pierre think that London is pretty?’ Carol knows that this person is only interested in whether or not she should take some new photographs of Buckingham Palace for the next edition of the brochure, or stick with last years. Carol answers ‘oh, yes, he thinks it’s beautiful.’

If we agree that Bob and Carol have both given correct answers, then we have a good reason to say that what they have said is true. Moreover, they do not need to disagree about any facts about Pierre, for by hypothesis they do not, and in fact accept all the same facts, yet still give opposite answers to the question ‘Does Pierre believe that London is pretty?’ The facts about Pierre then do not determine the answer that they each give. This does not, by itself, prove that what they say is true. But I think that it gives us good reason to build into the semantic account of belief reports the fact that belief reports are context sensitive and do not depend simply on facts about the agent.

I said at the end of Section 3.4 that the criticism that belief ascriptions are not context-sensitive applies to Salmon's theory and Eaker and Almog's theory. The criticism applies equally to Salmon's theory because he makes the truth of a belief ascription depend on whether an agent grasps a certain singular proposition via some guise or other and nothing more. There is nothing to suggest that whether an agent does this or not allows for belief ascriptions to be context-sensitive, since, presumably, given the little that Salmon says about grasping propositions, whether or not an agent does grasp a given proposition is a context-independent fact which suffices for the truth of a given ascription. If it suffices for the truth of an ascription, then a given ascription cannot be made false by the fact that some other independent cognitive fact obtains, namely, specifically that the agent grasps some other contradictory singular proposition.

Both Salmon's and Almog and Eaker's theories allow for the possibility that ascriptions of the form 'A believes that S and that not S' can be unproblematically true of an agent, and both in fact endorse such an instance of this in the case of Kripke's Pierre. Yet, this seems to me to falsify a datum. The datum is that we would not normally make such an ascription in the case of Pierre. The fact that such an ascription is puzzling depends not on any particular theory about what makes two beliefs contradictory or why we can have or not have contradictory beliefs and yet be rational. It just depends on our intuition that such an ascription would seem to be very puzzling and is in fact very rare. Salmon's and Eaker and Almog's theories predict that such reports would be common and normal. But what we need is an explanation of why they are uncommon and puzzling. I sketched my explanation of this at the end of Chapter One.

3.6.1

Mark Richard's contextualist account of belief reports

I now turn my attention to an important account of belief ascriptions which has been designed to take account of the context-sensitivity of belief

ascriptions. The account predicts the failure of substitutivity for belief ascriptions (at least in certain contexts) and also accounts for the contextual-sensitivity of belief reports.

I shall consider the theory put forward by Mark Richard (1990) in his book 'Propositional Attitudes'. Richard's theory begins with what I think is a very intuitively appealing idea: that the truth-value of a report depends not only on relating the agent to the objects and properties that he thinks about, but also on the words we use in the report itself and that in what way the words can affect the truth-value of a report can vary from context to context. Richard focuses on one example that I think manages to isolate the key feature of belief reports mentioned in the last paragraph of the preceding section.

We are to consider Mutt and Jeff who agree on everything that seems relevant to the question 'Does Odile believe that Twain is dead?' They do not however agree in their answers. When Mutt was asked this question it was because someone wanted to know whether or not Odile would list Twain in a list of dead Americans. Mutt knew that Odile accepted the sentence 'Twain is dead' and so said 'yes'. Jeff was asked the same question by someone who could not understand why Odile who, in pointing at a picture of Twain, says that she wants to meet him. Jeff knew that Odile rejected 'He's dead' (as uttered when pointing at the picture) and so answered 'no'. I agree with Richard's intuition that both Mutt and Jeff say something elliptical for something true. Richard's account is built to allow us to honour this kind of intuition.

The following is a brief, but hopefully fair and adequate, summary of Richard's theory. Richard claims that the 'that S' clause of a belief report determines an entity that he calls a RAM (an abbreviation for 'Russellian Annotated Matrix') which is a pairing of a natural language sentence with the proposition that it expresses, where 'proposition' is understood along Russellian lines, so that 'Twain is dead' is the singular proposition containing Twain and the property of being dead, in the appropriate order.

There is then a thesis in psychology: each agent has a Representational System consisting of RAMS of all and only the sentences she accepts as true. There are correlation functions that map RAMS to RAMS in such a way that the content of the RAMS is preserved in the mapping. A context of utterance of a belief ascription provides restrictions on which correlation functions are acceptable. These restrictions are determined by the intentions and interests of the speaker and audience. A restriction is specified in such a way that it determines for an individual (the agent, A) what is to count as an appropriate translation of part of the agent's RAM. A restriction can then be represented, for instance, as:

(R): A: 'a' \rightarrow 'b'.

where this means that when ascribing beliefs to A in the context in which (R) is operative, 'a' translates part of A's RAM that are expressed by only by 'b'. (Since it is always the case, on Richard's theory that Russellian content must be preserved in the mapping, this fact can be suppressed from the notation).

Finally, a belief report of the form 'A believes that S' is true in a context c iff there is an acceptable correlation function (in c) that maps the RAM that is determined (in c) by the 'that-S' clause of the report to some RAM in A's Representational System.

How does this apply to the example of Odile? Mutt says 'Odile believes that Twain is dead' in a conversation about whether Odile would list Twain among deceased Americans. In this context, the restrictions on correlation functions are such that the RAM determined by 'Twain is dead' can only be mapped onto a RAM in the agent's Representational System that contains the name 'Twain'. This seems intuitively acceptable since what is of concern in this context is whether Odile would use the sentence 'Twain is dead'. In terms of the notation for restrictions given above, we would have here

(R) Odile: 'Twain' \rightarrow 'Twain'.

Since, Odile does have the relevant RAM, the report is true. On the other hand, in Jeff's conversation there would be a different set of restrictions at work. Specifically, the restrictions would be such that 'Twain is dead' would not be an acceptable translation of the same RAM that made the other report true. So, in this context, the report would be false.

Richard's theory has the flexibility to allow that in some contexts, at one extreme, it might be necessary to use the words that the agent would himself use, and that there may be contexts, at the other extreme, where there are no restrictions on correlation functions, so that any 'that S' clause which in the context of the report determines a Russellian proposition which is part of a RAM in the agent's Representational System would be acceptable. In such contexts, the report could effectively be seen as the kind of report that Salmon says is the normal kind of case. In between these two extremes, we may have some restrictions on correlations. Which restrictions exist in a given context determine which terms can be intersubstituted *salva veritate* in that context. So, we have a more complex view of substitutivity here. Traditionally, substitutivity has been seen as an all-or-nothing affair, in the sense that either any substitution of co-referring terms is acceptable (in *de re* reports), or no substitutions are, unless they preserve sense. But Richard's correlations are not translations and have nothing to do with the preservation of sense, or any other kind of conceptual content. This seems intuitively right.

Applying Richard's account to Kripke's Puzzle, it seems that his account is able to preserve much of what I suggest should be preserved. We want to be able to account for why 1) 'Pierre believes that London is pretty' seems true when Kripke takes us through the puzzle and why 2) 'Pierre believes that London is not pretty' also independently seems true, but why the conjunction of 1) and 2) seems false, and also why the question 'Does Pierre, or does he not, think that London is pretty?', taken out of any natural context, and posed as a philosophical question, seems like a badly posed question.

It seems plausible to say, as Richard says, that one way in which restrictions on correlation functions become operative in a context is that if a speaker is focusing on how a speaker would express his beliefs, and thinks that his audience is focusing in this way also, then restrictions come to be operative in such a way that only a very limited range of 'that-S clauses are acceptable representations of the agent's RAMs. When Kripke takes us through Pierre's story, focusing on the beliefs he has in France before he comes to England and which he would express in French, it might be reasonable to expect correlations to be restricted in such a way that 'London' can represent Pierre's 'Londres', in which case, (1) 'Pierre believes that London is pretty' would be true in this context. Later in the account of Pierre the reporter and audience are focused on what Pierre thinks in England and what he would express in English, so it might be reasonable to expect the restrictions on correlations to be such that 'London' is mapped onto 'London', but not onto 'Londres' in which case (2) 'Pierre believes that London is not pretty' would be true, and the fact that Pierre had the RAM $\langle \langle \text{'Londres'}, \text{London} \rangle \langle \text{'est jolie'}, \text{is pretty} \rangle \rangle$ would be irrelevant in this context. Moreover, as Richard says, the most natural way of interpreting a conjunction of (1) and (2), and especially natural for the report (3) 'Pierre believes that London is pretty and that London is not pretty', would be with a single correlation restriction throughout the report, so that the two reports would not be true together and (3) would always be false.

3.6.2

Criticisms of Richard's theory

I agree with Richard's view that the intuitions about the context-sensitivity of reports in the Mutt, Jeff, Odile example are solid intuitions that we should try to explain, rather than explain away, but I am not convinced that his account is correct. I now consider some objections.

There is one major objection that has been made by Sider (1995) and Soames (2002) and developed in some detail by Nelson (2005). This is a

problem which arises from a rather technical aspect of Richard's theory, specifically his claims about restrictions on correlation functions. The problem has been called 'the problem of conflicting restrictions' (Nelson 2005). The problem can arise whenever the belief reporter is confused about the identity of the agent to whom the belief is ascribed in the belief report. I will show how the problem can arise in Kripke's scenario.

Pierre is also known in certain settings as Peter. Suppose that a belief reporter, R, is unaware of this and thinks that when people are talking about Pierre, sometimes calling him Peter, sometimes Pierre, two different people are being talked about. Suppose that R thinks sees Pierre, in Paris, looking at a picture of London and saying 'That's such a pretty city'. Suppose he hears of Peter, who is normally now based in London, not realising he is Pierre, who says many negative things about London, including 'London is such an ugly city'. Suppose that R finds himself in a context where is talking to those who know Pierre and wants to talk about his beliefs about London and contrast them with, as he would say, Peter's beliefs. He might plausibly say 'Pierre believes that London is pretty' and 'Peter believes that London is not at all pretty'. In the terms of Richard's theory it would seem plausible R's communicative intentions are such as to generate the restriction that for the former report R can use 'London' to represent the relevant part of Pierre's thoughts which he would express with 'this city pictured before me', and that for the latter report, 'London' would translate the relevant part of Pierre's thoughts that he would express with 'the city I live in'.

The problem is that no correlation function can satisfy both of these restrictions. The former restriction requires that 'London' translates only as 'This city in the photo' while the second restriction requires that 'London' only translates 'The city I live in'. Clearly, no correlation can satisfy both of these incompatible restrictions. Given Richard's truth-condition for belief reports, no belief sentence which has a term referring to Pierre in subject position and 'London' in the embedded subject position can be true relative to such a context, even the report 'Pierre believes London is itself'. The problem here does not depend on exactly what 'London' can be translated

as, but only on the fact that it translates differently under the two restrictions. This is what makes the restrictions incompatible.

Nelson considers several possible solutions to this problem and discounts them. I will not discuss these here, but only focus on what he thinks is the idea that forms the basis for a genuine solution. What is the source of the problem of conflicting restrictions? On Richard's theory, restrictions are on correlation functions that are functions from pairs of an individual (however conceptualised) and an annotation to an annotation. (All of the other responses to the problem retain this assumption, and that is why I think they do not work). When reporter is confused about the identity of the agent, thinking there are two people when there is in fact one, he can have two different sets of communicative intentions with respect to that individual. He might, as it were, think to himself, for Pierre I will use 'London' to translate this and for Peter I will use 'London' to translate that.

Nelson thinks that the correct solution requires that the correlation function embody this insight by not being determined by an individual and annotations but by an individual under a mode of presentation and annotations, so that a restriction would be represented as

(R1) $\langle A, \text{a mode of presentation} \rangle: 'a' \rightarrow 'b'$.

So, as applied to the Kripke scenario above we would have:

(R2) $\langle \text{Peter, thought of as Peter} \rangle: 'London' \rightarrow 'the city I live in'$

(R3) $\langle \text{Peter, thought of as Pierre} \rangle: 'London' \rightarrow 'the city in this photo'$.

These restrictions can co-exist as they are not incompatible.

I do not think that this solution can be correct. In the scenario I described all that was required to generate the puzzle was that the reporter thought that Pierre and Peter were two people. It has already been claimed, in the discussion of Eaker and Almog's theory, that an agent can think that there are two different people where there is in fact one without necessarily

having distinct modes of presentation associated with the individual. Nor does it seem necessary for the reporter to have different intentions with respect to what he takes to be two individuals that he associated two distinct modes of presentation with the individual. But if modes of presentation are not generally available to them they cannot be used as a general solution to the problem of conflicting restrictions.

Ultimately, I think that the problem is an artefact of the assumption that there are distinct representations that the reporter must somehow be able to specify or provide a translation of. Without such an assumption, the whole idea of correlation functions cannot even be made sense of.

The problem as I see it then stems from the fact that Richard's theory embodies the same assumption I see as underlying each of the theories of belief ascriptions that I have discussed so far as well as the overwhelming majority of theories I have not discussed. It is the central assumption that this dissertation aims to cast doubt on, namely that there are individuable internal representational states which ground the truth of ordinary belief ascriptions. In Richard's theory, this is his system of RAMs. In Richard's description of the psychological side, the agent accepts various sentences, each of which has a Russellian interpretation. For Richard, this set of RAMs encodes all the facts about the believer that are relevant to the truth or falsity of a belief ascription about her. There is, then, a prior psychological reality here that the belief ascriptions report on, and whether or not these RAMs exist in the believer's mind is not itself a context-sensitive matter. The contextual-sensitivity comes in when we are restricted about how we can represent what an agent believes with our 'that-S' clauses.

I have two related worries about building such ideas into the semantics of belief ascriptions. First, there is an epistemological concern. Jennifer Saul (1999) considers the case of a contemporary of ours, Alice, who has just heard, in a philosophy class, that Hammurabi believed that Hesperus appeared in the evening and that Hammurabi did not believe that Phosphorous appeared in the evening. Alice can report these beliefs and say

something that is intuitively true even if she is completely ignorant about Hammurabi, apart from knowing about these particular beliefs of his. She could even mistakenly think that Hammurabi would have used the terms 'Hesperus' and 'Phosphorous'. In fact, he spoke only Akkadina and would not have known these names. But Alice takes his beliefs to be linked to these names. If Alice does make this mistake then it is plausible to assume that her intentions in making these reports determine a restriction on correlation functions in such a way that 'Hesperus' in her ascriptions must represent 'Hesperus' as part of one of Hammurabi's RAMs. But Hammurabi did not know this name and would not have had the relevant RAM. But this would imply that Alice's report 'Hammurabi believed that Hesperus appeared in the evening' would be false. But intuitively it is true. Alice's ignorance about Hammurabi's language, and therefore, about his RAMs leads to her having intentions which restrict correlation functions in such a way that an intuitively true report turns out to be false.

This problem generalises if we assume that reporters are often ignorant about what sentences an agent accepts.

There is also an ontological concern. It seems to me to be incorrect to claim that an agent must have a relevant RAM in order for a belief ascription to be true of the agent. I think that having the relevant RAM is neither necessary nor sufficient for the truth of a belief ascription. I will only argue here that it is not necessary. Why I think it is not sufficient should become clear in the final chapter. But to give a hint. I think, and I will try to argue with a series of examples in the final chapter, that no one single fact generally determines the truth value of a belief ascription.

I think that it is not necessary because we often make intuitively true belief ascriptions even when we know that the agent does not accept any relevant sentences. The example I am about to give will be very controversial, and some will say that my intuitions about the truth of such an ascription are simply wrong. Nevertheless, I want to pursue the type of example that I have in mind because I think that it allows me to make a methodological point

that I think it is important to remind ourselves that we are supposedly committed to, or should be committed to. The methodological point is that we should take very seriously our ordinary pre-theoretical intuitions about the truth-values of ordinary belief ascriptions. My example involves the ascription of a belief to an animal. Regarding ascriptions of belief to animals, Richard says:

“It is supposedly obvious that some dogs think that bones are fun to play with. All I can say is that many people cease to think it is obvious that dogs can have such beliefs upon hearing the argument, and most people acknowledge being made uncomfortable by the argument [that such beliefs require the concept ‘bone’ which dogs allegedly lack].”¹⁸

I see my cat, Mozzy, who has followed me into the kitchen and put the milk carton back into the fridge. Mozzy seems thirsty, making the sound that he always makes when he is about to start drinking but can’t get to his milk or water. He walks up to the fridge door, and makes a crying sound. The carton of milk was actually virtually empty and I had just left a few drops of milk to make a cup of coffee later. My housemate knows that the milk has all but gone. She comes into the kitchen and looks at Mozzy. She’s wondering why he’s crying. I say ‘Mozzy thinks that his milk is in the fridge’. What I say seems to me perfectly natural, perfectly true. Had my housemate come in a few seconds earlier and seen Mozzy following me as I put the carton back into the fridge, she would also agree that what I said was true.

But, of course, Mozzy does not have a language consisting of sentences and so does not have any RAM consisting of a sentence that he accepts and its Russellian content, so on Richard’s account the report is false. Maybe this line of thought is too quick, because Richard only says that having a RAM involves accepting a sentence as a preliminary account of RAMs. What he is essentially committed to is the claim that for a report of the form ‘A believes that S’ to be true, there has to be some sort of isomorphism of referential content between ‘that S’ in the ascription and some inner state of the agent. He uses the term ‘referential content’ to name the relation that obtains when

¹⁸ Richard (1990) pg. 254

for each simple constituent x of the 'that-S' clause there is a constituent x' of the agent's inner state such that x' determines in the context of the ascription the content of x in the context.

This notion is still rather vague because Richard does not (try to) specify what it is for a part of an inner state to determine a particular content. A complete account here would be an account that provides an answer to the question of intentionality and is a huge independent philosophical project. In the absence of such an account we can neither directly confirm nor refute Richard's theory with the example of the ascription about my cat.

However, if we assume that having a state with a part that determines a particular content 'm' is the same as or entails having the concept 'm', as Richard seems to, then it could be argued that Mozzy does not have the concept 'milk' and that is why any ascription with 'milk' occurring in the embedded clause of an ascription will be false. Adapting an argument from Stephen Stich that purports to show that dogs do not have the concept 'bone', we could give three reasons for thinking that my cat does not have the concept 'milk'. First, Mozzy does not have enough of the same beliefs about milk that we who use the word 'milk', or any normal translations of it, do. Secondly, nothing in Mozzy's behaviour suggests that he groups samples of milk into a single class which excludes other substances, in particular water. Thirdly, Mozzy is not able to distinguish ersatz from real milk.

The underlying problem, according to Richard's view, would be that if the three claims were correct then it seems that no discriminatory capacity of the cat's, which we would identify as its concept 'milk', determines an extension even remotely close to the extension determined by our concept 'milk'.

I am inclined to think that the three claims about Mozzy are correct and also that this gives us good reason to say that Mozzy lacks the concept 'milk'. So I will grant these two assumptions. What I want to challenge is not the claim that Mozzy lacks the concept 'milk', but that the truth of an ascription

including 'milk' in its 'that-S' clause does depend on Mozzy having that concept. If we take as our starting point the intuitive truth of the ascription in its context, then we can use the claim that the cat lacks the concept 'milk' to argue against the claim that the truth of the ascription depends on possession of the concept.

Returning to Richard's quote about the attribution of a belief to a dog, I think that what Richard says here is actually consistent with my claim that the ascription that, say, 'Fido thinks that this old bone is fun to play with' can be true in many natural contexts of utterance. First, Richard does not even go so far as to say that people change their minds about the truth-value judgement of such an ascription, only that it ceases to be so obvious. Moreover, if one asks, after giving the argument about concept possession, 'But does Fido *really* believe that bones are fun to play with?', we actually change the subject, and I dispute the relevance of the answer to this new question to the truth of the original question. First, the word 'really' is added in, and the question changes grammatical form from 'Does Fido (really) believe that...?' to 'Does Fido (really) have the belief that...? I claim that there is a change in context partly brought about the addition of the word 'really' and the fact that a philosophical argument has just been given. If this is correct, then the fact that an ascription is false (or as Richard says: 'not obviously true') in one context, says nothing about whether it is true in the original, natural context. We must remind ourselves that Richard is committed to the context-sensitivity of belief ascriptions.

The change in grammatical form also signifies, I think, a change from a natural context to a change into a specifically philosophical or psychological context. As I suggested in Section 2.2, there is a difference between 'A believes that S' (understood as an ordinary market-place ascription) and 'A has the belief that S' (understood as a theoretical question). As support for the significance of this change in grammatical form, I would like to draw our attention to what would most naturally be understood by the question 'Does Fido really believe that his bone is fun to play with?'. It does not, in this

form, sound a question about animal concepts. It suggests the following (italicised) contrast ‘Does Fido think the bone is fun to play with *or does he just pretend to keep his owner happy?*’

In Chapter Four I will say much more to support the truth of this intuitive ascription, but for now I have tried to suggest that our intuitive judgement about the truth of this ascription is correct and is even consistent with Richard’s quoted remarks, which he takes to count against the intuitive judgement. This concludes my argument that RAMs are not necessary for true belief ascriptions.

3.7

Conclusion

In this chapter I have evaluated four theories of belief ascriptions and have argued that they are each unsatisfactory as accounts of belief ascriptions. Although the theories and my criticisms of them are rather diverse, I think many of the criticisms can be unified by relating them to what is a common source. The common problem is that each theory assumes that a belief ascription reports on some single fact about an agent. In Frege’s theory there is the question about which sense the agent grasps and assents to; in Salmon’s theory there is the question about which singular proposition an agent assents to; in Eaker and Almog’s theory there is the question about which object an agent applies, in thought, a certain predicate to; in Richard’s theory, there is the question about whether an agent has a given RAM in her representational system. Each of these questions depends on some one cognitive fact which either obtains or does not obtain irrespective of any contextual considerations about the setting of a given ascription.

With Frege’s theory, there was the difficulty of knowing and specifying which sense an agent was related to, so that correct ascriptions became a mystery. With Salmon’s and Eaker and Almog’s theory, belief ascriptions became too uninformative, had counter-intuitive truth-values and were insensitive to ordinary facts about the ascription of contradictory beliefs.

Each of the above theories made it the case that there should always be a univocal context-free answer to the question ‘Does A believe that S?’ since on each theory whether or not the relevant fact obtained is itself a context-free matter. Richard’s theory allowed for the context-sensitivity of ascriptions but still made them depend on a single cognitive fact about the agent. But I argued that the obtaining of such a fact was not generally knowable and was not necessary for the truth of a belief ascription.

Some desiderata for a positive account of belief ascriptions have emerged:

1. A theory of belief ascriptions should take seriously our intuitive truth valuations. Neither substitutivity nor its failure should be taken as the norm. Substitutions can often preserve truth, but failures of substitutivity are also too prevalent to be dismissed. (Salmon’s theory and Eaker and Almog’s theory are not fully adequate in this respect.)
2. Belief ascriptions are context-sensitive and this fact should be built into the semantics. (Frege, Salmon and Eaker and Almog do not allow for this in any clear way.)
3. We should not presuppose that belief ascriptions depend on a picture of the mind which is itself controversial. (Richard’s theory presupposes an account of the mind that denies the truth of some intuitively true ascriptions.)

In chapter four I aim to give the outline of an account which honours these desiderata, and will also give a very central role to the role of a belief ascription in our day to day lives.

Chapter Four:

Proposing an Account of Belief Ascriptions

4.0

Introduction

In this chapter I shall put forward my positive account of belief ascriptions, which, I think, avoids the problems of the other theories discussed in the previous chapter. Sections 4.1 and 4.2 set out the main desiderata for a semantic account of belief ascriptions. Section 4.3 describes the general form of the semantic account that I give. Section 4.4 describes what I think is a useful model for thinking about the semantics of belief ascriptions (that is, seeing each ascription as an answer to a specific question) and in doing so tries to say something about the function of ordinary belief ascriptions. The actual statement of the truth-condition for a belief ascription is given in Section 4.5 in terms of the model described in 4.4. Section 4.6 describes the metaphysics that is presupposed by my account, explains more about the kind of questions that belief ascriptions would normally be seen as answers to, and shows how the context-sensitivity of ascriptions arises within this setting. In 4.7, I present several examples in which I apply the considerations of Sections 4.4 – 4.6. I try to use these examples to elicit intuitions about the truth-value of the ascriptions to support the claims made in the above sections. Each example will be used to highlight a different point and across the examples I will vary the different kinds of factors to show all the different sources of context-sensitivity of the ascriptions. Section 4.8 tries to explain the terms given in the statement of the truth-condition a little further and in doing so aims to defend my account against one intuitive objection. In 4.9, I contrast my type of contextualism with Richard's in order to show clarify how the context-sensitivity arises in my account and to show how my

account is fundamentally different from those theories I criticise in Chapter Three. The *de dicto/ de re* distinction, generally given a very prominent place in the discussion of belief ascriptions is discussed in 4.10 where I try to deflate its significance.

4.1 The central desideratum

This chapter aims to provide an account of belief ascriptions which abandons a widely held and rarely explicit assumption that I have tried to cast doubt on in this dissertation (in Chapter 3) and which I have argued underlies Kripke's puzzle about belief (in Chapter 1). This is the assumption that a belief ascription depends (at least in part) for its truth on some one single context-independent cognitive fact about an agent. In Chapter Two we have seen theories (Frege's, Salmon's, Almog and Eaker's) according to which some single cognitively specifiable fact, which is itself a context-independent fact about the agent, fully determines the truth-value of a belief ascription. We also saw a theory, Mark Richard's, which gives a central role to the obtaining of some single cognitive fact about the agent, although on Richard's theory, this fact does not by itself determine the truth-value of the report, since how a reporter represents this fact is a context-independent matter. In both types of theories, a central role is given to some single psychological fact about an agent which can be stated in a context-independent way. I have explained why I think this assumption is problematic.

Philosophers of language have tended to say that belief reports relate an agent to a proposition (whether propositions are conceived of as singular propositions containing individuals and properties, or as thoroughly conceptual entities), and may or may not specify what this means in more basic psychological terms. Even those philosophers of language who abandon the idea that belief ascriptions relate agents to any kind of proposition make the truth of a belief report depend on one single cognitively specifiable fact which can be stated in context independent way. Eaker and Almog, for instance, say that the truth of a belief ascription

of the form ‘A believes that a is F’ depends only on the fact that the agent has, in thinking about a, ascribed to it the property F.

Although I have not explicitly addressed the subject of belief as a topic in the philosophy of mind, rather than in the philosophy of language, it is worth noting that philosophers of mind (such as Ruth Millikan, David Papineau, and Fred Dretske) tend to think that believing something is a matter of having a representation of some sort in one’s mind. Again, the specification of the content of such a representational state is a context-independent matter. These philosophers of mind may say that they are concerned with the cognitive phenomenon of belief, and they may not be explicitly addressing the semantic description of belief sentences (as the philosophers of language do and as I try to do), but it is all too easy to conflate these projects and assume that what philosophers of mind are investigating is the cognitive grounding of our ordinary belief ascriptions. I suggest that the assumption which I have said lies behind all the theories that I criticise may encourage a conflation of what I think are two distinct projects: the study of the content inner representational states, and the study of ordinary everyday belief ascriptions. It is the latter only that I am interested in.

The central desideratum of my semantic account of ordinary belief ascriptions is that it should not make belief ascriptions depend on any single context-independent cognitive fact.

4.2 Other Desiderata

There are also the following desiderata that I want my account to answer to.

1.

I want my account to account for the intuitive truth-values of our ordinary belief ascriptions. Whatever burdens of explanation that my account accumulates, I do not want the burden of explaining away highly counter-

intuitive and pervasive intuitions about truth-values. I take it as a sort of bottom-line methodological starting point that our intuitive evaluations should, at least for the most part (there will be *some* intuitions that I think are mistaken and shall want to explain away), be taken as the data to which the account should be ultimately responsive.

2.

The account should give a central role to the context-sensitivity of belief ascriptions as a general class, and not merely of those which contain an obviously indexical element. This second desideratum really follows from the first. For I think that our intuitions about truth-values depend on how the context of a given report is specified. That this is the case is becoming increasingly orthodox in the literature on the subject, and has already been defended to some degree in my earlier chapters, but I will try to strengthen the case with a detailed discussion of some new examples in this chapter. The way that the contextual-sensitivity arises in my account will be quite different I think from the way it arises in other accounts, for in other accounts the contextual sensitivity cannot be explained, except in terms of the assumption that I have said in the previous section that I want to reject.

3.

The account should be psychologically realistic in the sense that making true ascriptions should not require that the ascriber have more knowledge about the agent than he actually does have; in particular, it should not require detailed knowledge of an agent's mental representations (if such things there be) or an agent's 'way of thinking', where this last is to be understood along Fregean lines.

4.

The account should recognise that belief ascriptions are used to serve specific purposes, such as explaining and predicting the agent's actions.

5.

This dissertation started with a detailed description of Kripke's puzzle about belief and said that any theory of belief ascriptions must ultimately come to grips with it. So I want my account to come to grips with it, but not necessarily within Kripke's strictures, since those strictures arise, in part, from what is, I think and have argued, a mistaken assumption. In application to Kripke's puzzle, I want my account to i) account for why 'Pierre believes that London is pretty' is intuitively true when he first gets us to consider that report, ii) account for why 'Pierre believes that London is not pretty' is intuitively true, iii) account for why 'Pierre believes that London is pretty and that London is not pretty' is intuitively false even though the ascriptions referred to above are both intuitively true. I would also like to be able to iv) offer some principled rationale for rejecting Kripke's question 'Does Pierre, or does he not, believe that London is pretty?' since it seems that we are, as Kripke insists, not able to give a simple and unproblematic answer to this.

4.3

Giving the truth-condition

Before giving a statement of my account I would like to make a general point about what sort of semantic project I see myself as engaged in. In giving my semantic account I will give a truth-condition for belief ascriptions. I will give a biconditional of the form 'An utterance of 'A believes that S' in a given context is true if and only if'. What replaces the dots will be an attempt at a statement of a condition necessary and sufficient for the truth of the ascription. The truth-condition will be a condition for the truth of the report but is not meant to be an analysis of an ascription. It is not an articulation of the content of the ascription, not a synonym, and does not give the logical form. This means, among other things, that the ordinary language user need not be able to give the condition himself. The vocabulary I use in stating the truth-condition is to be thought of as occurring in a philosophical metalanguage, not in the

language of the ordinary belief ascriber, even though I will state the condition using ordinary English words.

Giving a truth-condition in this sense differs in a significant way from giving the content of a report. For when we give the content, it is presupposed that there is just one correct content which is *the* content of the report. However, there is no such presupposition when the aim is only to give a truth-condition. There is then a very real question about what is to count as an adequate account of the report, for there is no correct *one*. Of course, the truth-condition must be extensionally correct, but this is obviously not a sufficient condition. The following would be, although extensionally correct, completely uninformative: 'A believes that S' is true iff A believes that S. We would like our account to be informative, so a disquotational schema will not do. But what exactly should we expect the truth-condition to contain? Erin Eaker (2004) makes an interesting claim about giving the truth-condition for a belief report. Remember that Eaker said 'A believes that a is F' is true iff A has in thinking about a ascribed to it the property of being F. She emphasises that 'thinking about' occurs unanalysed in her semantics. In her defence she says that the task of providing an analysis of belief and of the thinking about relation is not a part of providing a semantic account of a 'believes that'. She says that whatever may, metaphysically speaking, constitute having a belief, or thinking about an object, the semantics of the report is not the place for this theory to appear. She offers, apparently in support of this claim, the following analogy. She says that one would be surprised to see an analysis of vaporisation in the semantic theory for the sentence 'The Martians vaporised Mary because they thought that she was an Earthling spy'. She then asks, rhetorically: 'Why should we should we expect an analysis of Martian belief in the theory?' I agree with the intuition, but I think that the analogy is not apt.

Our object of study is belief ascriptions. They have been selected as an object of study because we think that they form a natural semantic class, exhibiting what may appear to be puzzling logical behaviour and because

they may be thought to shed some light on issues in the philosophy of mind. No doubt, if we were interested in vaporisation-sentences (because we thought they formed a special class worthy by themselves of study), then we would like our semantics to tell us more about vaporisation than is given to us by the schema ‘‘A vaporises B’ is true iff the relation of vaporisation hold between A and B’. I hope that my account will cast some light on belief, metaphysically speaking. The only plausibility that Eaker’s claim has is if we are trying to give the logical form of belief ascriptions. But I am not, and nor is she.

I do not think that there is any principled and general way of setting an upper limit on what a semantic account of belief ascriptions should include; there is only the lower limit discussed above. I would like to say as much as I can. And so I will have something to say about the metaphysical underpinnings of our ascriptions in Chapter Four.

4.4

How to think about belief ascriptions

In this section I will state and explain my account of belief ascriptions in an abstract way, before looking in detail at examples in Section 4.7 in order to both illustrate and support my theory.

I suggest that it will be useful as a heuristic for understanding what I want to say about belief ascriptions that I make an idealising assumption about belief ascriptions. It is the following: a belief ascription is always an answer to a question. In the typical, or canonical, situation, there will be three parties involved; there is the agent who will be designated, as she has been so far, by replacements for the schematic ‘A’ in ‘A believes that S’. There is the reporter who I will represent with ‘R’ in my abstract account. There is also the audience, a speaker who I will designate with ‘B’. B asks a question of R and R responds with an ascription of the form ‘A believes that S’. I do not think that there is a definite limit to the range of questions

to which an ascription can intelligibly be thought of as an answer, but I wish to mention three central kinds.

1) Often the question will be of the form 'Why did A do X?' where 'do X' is to be replaced with some verbal phrase which expresses an intentional action or some non-intentional reaction of an agent to some situation. I assume that an answer in the form of an ascription, when it is given as an instance of a response presupposing an intentional action, is given as a reason-giving causal explanation. This question only arises when B assumes that A did in fact do X. Once the question has arisen it can either be rejected because B believes that A did not do X or it can be answered. There may be many forms of possible answer, but 'A believes that S' would be a central and typical form of response.

2) An ascription can also be given as an answer to the question 'Does A believe that S?' This form of question could again be a request for an explanation of some particular action in, for instance, a case where B has suggested an answer to his own 'Why did A do X?' question and is essentially asking if that is the correct explanation. For example, B has just seen A pour milk into the sink and asks 'Does A believe that the milk is sour?' More typically, however, B will want to be able to predict A's behaviour or reactions in various situations or to various kinds of triggers. B may be interested in what A would do in one specific situation or may want to know something about A so that he will know what A would do in various different situations. In the typical case, B will be interested in some aspects of A's behaviour and not in others, although it may not be determinate exactly what aspects of A's behaviour B is interested in. B will not have in mind a definite list of possible scenarios for each of which he will want to know what A's expected behaviour. I consider this use of an ascription a very central case and most of my discussion of the examples to follow will be examples in which an ascription is given to provide information so that B can predict some of A's behaviour.

3) There is at least one other intelligible purpose that B may have in asking ‘Does A believe that S?’, and that is to find out more about A, but not in such a way that we come to know what A would say or do in various situations. If B were to ask, for example, ‘Does John still believe that God exists?’, B may not be at all interested in any aspect of A’s behaviour. We are often interested in what people believe without being interested in anything else about them. Typically we may want to know about A’s character or personality or want to take a moral stance towards that person, or we may be interested in what they think about us, but not care about what they would actually say or do.

I shall concentrate on questions of the first two kinds (requests for explanations, and requests for predictive information) because I think that it is possible to trace the context-sensitivity to particular kinds of source in the case of these two kinds of questions. The basic idea will be that whether or not an answer to a question is a good answer depends on in part on the purpose of the question. When the question is a request for an explanation, a good explanation will be a good answer, and hence a true ascription. When the question is a request for predictive information, an answer will affords the right kind of prediction will be a good answer, and hence a true ascription.

4.4.1

Proper questions

I assume that a question arises for B only if B has a specific purpose in asking it. I will say that a question is *proper* when and only when it has a specific purpose. If B has no specific purpose, then even if she did for some reason ask ‘Does A believe that S?’ there would be no saying whether R’s response was a good answer, or even counted as an answer at all. I make this model of belief as answers to questions a central feature of my semantic account of belief ascriptions. I claim that a belief ascription can only be evaluated for truth or falsity at all if it is an answer to a proper question. As a corollary, considered in abstraction from a possible proper

question, a belief ascription cannot properly be considered either true or false. This corollary may seem counter-intuitive, and I will try to account for its counter-intuitiveness in my discussion of Example 1. I also claim that a given ascription may be true when given as a response to one question and false when given as a response to another question, hence the context-sensitivity of belief ascriptions. This will be illustrated in 4.7.

The question (with its specific purpose) giving rise to the belief ascription is then a crucial aspect of the context in which an ascription is to be evaluated. What else must be included as elements of the context in which a belief ascription is evaluated? Apart from any elements which are not specific to the evaluation of belief ascriptions but are also required for the evaluation of utterances with other indexical elements, there is the shared background knowledge of B and R.

It is important to note that questions are not, for my purpose, to be identified with the particular interrogative form used in asking the question, for questions will be distinct when the purposes behind them are distinct, even though the same verbal form may be used to ask the question. (this will be seen in the examples of section 4.7). A question then should be thought of as including its purpose.

4.4.2

Contexts for belief ascriptions

A context consists of a time, an agent, a reporter and an audience, a question with a specific purpose, and the background knowledge of the speaker and audience.

4.5

Stating the truth-condition

In this section I will give a brief statement of the truth-condition of a belief ascription using the model I have described above, before fleshing out

what kinds of factors are relevant to the asking and answering of questions to which an appropriate response is a sentence of the form ‘A believes that S’, looking at various examples of the relevant answer and question pairs and then defending the truth-condition against some likely objections.

An utterance of a belief ascription of the form ‘A believes that S’ is true in a context c if and only if it provides a good answer to the proper question of its context.

What counts as ‘good’ depends on the goals of the reporter and audience and the background knowledge of the audience, and will be explained via discussion of the examples of Section 4.7 and directly in Section 4.8.

An answer is good if it gives the questioner the information that he is looking for, which is determined by the purpose of the question.

4.6

The metaphysical background

The whole practice of asking questions about what an agent has done or will do in various circumstances and the practice of ascribing beliefs to answer these questions presupposes a very complex set of patterns to be found in an agent’s behaviour. Without some pattern we could not find each other intelligible at all. I now want to say something about such patterns as they play a central role in our belief ascribing practices, most obviously when we ascribe beliefs to answer questions which are asked by B, so that she can make some predictions about A’s behaviour. I now want to say something about these patterns before showing how they feature in belief ascription.

So far I have used the term ‘behaviour’, but what I want to consider is actually something much broader, which, for lack of a better expression, I shall call ‘an agent’s mental life’. An agent’s mental life exhibits certain patterns which are normally easily discernible to us. To say that an agent’s

mental life exhibits certain patterns immediately raises two questions: ‘Patterns in what?’ and ‘What kind of patterns?’ The patterns I have in mind are patterns in capacities and dispositions to do, say, think and feel certain things. Dispositions can be characterised by conditional statements of the form: if a certain condition C obtains, then an object O will exhibit property P. Roughly speaking, we can say that O has a given disposition just in case the relevant conditional is true. For example, paracetamol has the capacity to cure headaches when swallowed because it is true that when swallowed (condition C), the paracetamol (object O) cures a headache (exhibits P). For an example relating to an agent, John is disposed to get very irritable when trapped in the underground because it is true that when trapped on the underground he will tend to snap at people and huff-and-puff, say. Consideration of these examples invites immediate consideration of some complexities. The statements I have given as examples of dispositions may seem overly simplistic, for it has to be conceded (and it standardly is conceded in the literature on this subject) that conditional statements such as these are only strictly true *ceteris paribus*. It is very easy to concoct examples which C obtains but O does not exhibit P, and yet we want to say that O does have the relevant disposition.

Not all ascriptions of dispositions are couched in overtly dispositional language. Many commonly used dispositional predicates – ‘is fragile’, ‘is soluble’, ‘is malleable’, ‘is an acid’, ‘is irritable’, ‘is stubborn’, ‘is hot-tempered’, make explicit reference neither to disposition to manifest certain properties nor to the conditions in which such properties would be manifested. It may not even be possible for a competent user of such predicates to provide an analysis in overtly dispositional terms of the form ‘if C then O will have P’. What, for instance, is an irritable person supposed to do and in what circumstances exactly?’ It is not that we cannot think of anything to say in such cases, but that it is not clear exactly what conditions and manifestations should be included.

Even in cases where a disposition is ascribed in overtly dispositional terms, there can be a significant difference in how specific the dispositional ascription is. Consider, for example, the differences between the following:

1. John is hot-headed.
2. John is disposed to respond very angrily to minor provocation.
3. John is disposed to shout and swear in response to minor provocations.
4. John is disposed to shout and swear in response to a trivial argument.
5. Ceteris paribus, John is disposed to respond angrily to minor provocations.
6. John is disposed to respond angrily to minor provocations unless he is drunk.

In (1) John is described with a predicate which is not explicitly dispositional. (2) might plausibly be thought of as an analysis of (1), but leaves it open as to what is to count as an angry response and what counts as a minor provocation. (3) might be thought of as including examples of angry responses and minor provocations. It is not clear what the entailment relations between these four sentences are. (1) might be thought of as entailing (2), and possibly as being entailed by it. (1) would probably not be thought of as entailing either (3) or (4) since there are other ways of being hot-headed. Whether (3) or (4) suffice for the truth of (1) may be a matter of dispute. (5) explicitly states that the manifestation of the disposition occurs other things being equal, and is intended to be equivalent to (2) which is assumed to implicitly be restricted in this way. (6) might be one plausible way of understanding (1). It explicitly refers to a condition such that if that condition obtained, then even if the disposition was not manifested, the general dispositional ascription would not be falsified.

The above brief discussion is intended to highlight two complications for any kind of account which understands some phenomenon in terms of dispositions and capacities, or at least gives them a central role in the

account. First, there is the question of how dispositions are to be individuated. Is John's hot-headedness one disposition or is it really a cluster of more specific dispositions, and if the latter, which ones exactly? Since dispositions can be specified to different degrees (consider the differences between (2), (3) and (4)), do we have different dispositions here or different specifications of a single disposition? Perhaps these questions pertaining to individuation could be given some definite answers if there was something like a normal, canonical form for disposition statements. But I think that the variety of normal uses exemplified above suggests that there is no privileged form of disposition ascription.

It would seem that there is no determinate end to a list of more specified disposition ascriptions that are taken to be equivalent to a disposition ascription involving a simple predicate.

The above two points, (that there may be no principled way of individuating dispositions and no determinate list however we individuate them) would make certain kinds of accounts of belief or belief ascription given in terms of dispositions difficult. The features of disposition ascriptions would provide at least a *prima facie* difficulty for any account which said, for example, that 'A believes that S' is true iff A has all of the dispositions in a given list, most of them, or a certain number of them. Nor do I think that a belief ascription is equivalent to, implies, or is implied by any single dispositional ascription.

So far, I have considered a particular mental predicate, 'is hot-headed' and discussed some dispositional ascriptions. I now want to turn to an example more germane to my topic. Let us consider the predicate 'believes that London is pretty'. There are dispositions that we ordinary belief ascribers are apt to associate with having this belief. I use the word 'associate' as a deliberately vague word because I want, for the moment, to remain neutral about exactly what the relation is between a given belief ascription and the dispositions that we associate with it.

The dispositions that an ordinary speaker associates with having the belief that London is pretty are roughly those dispositions that she would cite, given enough time and perhaps some prompting, when asked the question ‘What would someone who believed that London is pretty be likely to do, to say, to think and to feel and in what circumstances?’ This particular belief might be associated with the disposition to say that London is pretty, to assent to the sentence ‘London is pretty’, the disposition to draw certain inferences, given certain further information, for example the disposition to draw the conclusion that England has a pretty capital given the information that London is the capital of England, the disposition to feel some degree of surprise when someone says ‘London is a very ugly city’, the disposition to consider London as a possible place to visit.

This list could be continued indefinitely. I have merely tried to give some illustrative examples. It seems that for any arbitrary belief that we can think of, we seem to be able to produce a list of dispositions which we would associate with having this belief. That we are able to do this even for beliefs that we have never considered before shows that we do not learn lists of beliefs and dispositions that are to be associated with them one by one. However we do it, it seems that being able to do it is a necessary part of being competent with the locution ‘believes that’.

No doubt different speakers would if asked, produce different responses if asked for a list of dispositions associated with having a given belief, but I will assume that there would be considerable intersubjective agreement if speakers were presented with a candidate list for a given belief and asked of each disposition on the list, whether it should be associated with the given belief. I will assume therefore that we can meaningfully talk of *the* list of dispositions associated with a given belief. The list is the one that everyone would agree on, at least if prompted.

4.6.1

Dispositions and belief ascriptions²⁵

Before saying exactly what role such associated lists of dispositions play in my semantic account of belief ascriptions, I want to say a few things about how such dispositions have traditionally been understood. Howard Wettstein, in his book, 'The Magic Prism', talks of such associated dispositions (he calls them 'coherences') and says that they have been given very different roles in the philosophy of belief and belief ascriptions. He says that 'at one extreme' is Frege, who sees 'believing that p' as standing in a particular relation to a grasped content, so that believing becomes not essentially embodied, so that the coherences are not essential to having a belief. At most, the 'coherences' are the causal consequences of having a given belief.

It is not really so clear, though, that Frege would have said that the 'coherences' are causal consequences of having a given belief rather than constitutive of having it. Frege was interested in a semantic treatment of certain types of sentences, where semantics was understood, as we have seen, as having to do with the analysis of a sentence. He said that the referent of an embedded clause was a proposition, and said various things about the nature of such propositions, but his interest was not primarily metaphysical, so he could remain neutral about what it means to, say, stand in the relation of acceptance to a proposition. More recently, Nathan Salmon, has endorsed such a picture of belief and said a little more. He says that belief is a type of inward assent to a piece of information, "that to believe that p is to covertly concur with, to endorse mentally, to nod approval to the information that p when p occurs to you." These are metaphors and perhaps they do suggest that having a given belief is matter of something internal to the mind which cannot essentially have to do with dispositions.

²⁵ Scwitzgebel (2002) proposes a context-sensitive account of ascriptions where what a person believes depends on what dispositions are relevant in that context. He does not, however, draw on the concepts of explanation and prediction as I do, and his account has consequences that I do not accept. I am not sure that he sees belief

At the other extreme, there are understandings of belief and belief ascriptions which give the ‘coherences’ as a very central role. Gilbert Ryle (1949), it is often said, held that ordinary belief ascriptions could be analysed in terms of dispositions to do, think and feel certain things. Ryle did not actually give many examples of such analyses, and in his short discussion of belief he gives lists of what seem like dispositions that a competent user of ‘believes that’ reports would plausibly associate with a given belief ascription, or at least agree with once she had been presented with them. He said:

“Certainly, to believe that the ice is dangerously thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people’s assertions to that effect, in objecting to statements to the contrary, [.....] . But it is also be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to ward other skaters. It is a propensity to make certain theoretical moves, but also to make certain executive and imaginative moves as well as to have certain feelings”. (pg. 134-5).

But of course, in light of my observations above, these lists cannot be complete and are to some extent arbitrary. Geach (1957) pointed out a difficulty for any analysis of propositional attitudes in terms of dispositions. He observed that what an agent will be disposed to do, say, think, feel given a particular belief will depend on what other beliefs the agent has. This certainly implies that a simple atomistic analysis of belief in which a belief is paired off with a single disposition is a hopeless project. More recent developments in philosophy of mind are sensitive to Geach’s point and have resulted in considerably complicated analyses of belief. I won’t discuss these later developments here, but observe that they all claim that belief ascriptions can be analysed in terms of dispositions or are functionally specifiable (this would mean that ‘A believes that S’ can be analysed in terms of dispositions to produce certain effects in certain conditions and to be caused by certain inputs).

ascriptions as ESSENTIALLY answers to questions, as I do. What he takes to be hard cases I take to be the norm.

I do not think that any of these positions get the relation between ‘the coherences’ and the truth of belief ascriptions right. In my view, they make the connection between the coherences and the ascriptions too tight. For on each of these views, there is a list (maybe a very long and complicated one) which is both necessary and sufficient for the truth of a given ascription. On the assumption that whether or not these complex dispositional and functional descriptions apply to an agent is a context-independent matter, then the truth-value of the belief ascriptions will also be a context-independent matter. My discussion of the examples in the next section will try to defend further the claim that it is not.

I will now try to state what role the ‘coherences’ have in the semantics of belief ascriptions. Belief ascriptions are *not* to be understood as semantically equivalent to any ascriptions of dispositions to an agent. So that one cannot say, in general terms, that ‘A believes that S’ is true if and only if , where the blanks specify some list of dispositions, however the dispositions are spelt out and with whatever caveats.

I think that the coherences are presupposed by our general practice of ascribing beliefs, in the sense that they are metaphysically necessary for the practice to be possible at all. Without the coherences, we could not predict, explain or understand agents, and belief ascriptions play a central role in these phenomena.

4.6.2

Dispositions and asking questions: two sources of context-sensitivity

Let me focus on a question asked by B whose purpose is to obtain an answer which will allow B to make predictions about an agent. I assume that in the typical case B will not want to be able to predict all of an agent’s actions and responses in all possible scenarios, for we all know that this is not possible. It is not possible because there are too many factors relevant to determining what an agent will do in a given situation. I assume that the question ‘Does A believe that S?’ arises only when B has already

made a significant number of assumptions about A, for without these assumptions, knowing whether or not 'A believes that S' is true provides no information about what A would do. This is for the reason mentioned above, given by Geach, that a given belief cannot by itself be expected to issue in any particular behaviour. Assumptions must be made about the agent's other beliefs, desires, and other attitudes, as well as assumptions about his environment. Some of these assumptions are normal across most contexts of question. For example, it would be normal to expect that a person does not want to be severely injured and knows that if he were to be hit by a car he would be seriously injured. It is on the background of such assumptions that one can make predictions about an agent's likely actions if given the further information that he believes that he is crossing a dangerous road. Some assumptions would be peculiar to a given context of a question. For example, it is only if we know that A believes he is with someone who understands English that he is likely to express his beliefs to them in English. Whether a belief ascription allows us to predict a given range of actions depends on other assumptions. What other assumptions are made will vary with the context of the question. Whether a given belief ascription allows us to predict a given range of actions then is a context-dependent matter.

The same goes for explanations. A belief ascription does not explain an action simpliciter. Explanations, like predictions, are essentially *situated in a context*.

What explains in one context will fail to explain in another context. And what fails to explain in one context can explain in another context. I shall discuss belief ascription as action explanation in Example 7.

This sort of context-sensitivity of predictions and explanations described above is one source, I will argue, of the context-sensitivity of belief ascriptions. But there is another very important source which arises only in the case of belief ascriptions used to allow predictions of an agent's behaviour. There is an important asymmetry between explanations and predictions, which is obvious enough once pointed out. It is that

explanations are explanations of a specific action as it falls under a given description, or perhaps of why a certain kind of action or response is instantiated in a given situation. Sometimes prediction is like this, and sometimes it is not. Sometimes we want to know whether A will do some particular kind of thing in a specific situation, but sometimes we want to be able to predict a range of an agent's actions and responses in a range of circumstances, and we may not always have a precisely determinate range in mind. But I consider that the question 'Does A believe that S?' arises typically for B only when B is interested in some aspects of A's behaviour and responses in some restricted range. I claim that belief ascriptions are tailor-made for such cases and will try to explain this more fully in my discussion of Examples 3-6. This variation in range of behaviour and responses that is the focus of B's interest between different speakers is another important source of context-sensitivity in belief ascriptions which are responses to requests for predictive information.

4.7 The examples

Example 1

My first example is one which concerns what I will call an ideal agent. He is ideal in the sense that he is perfectly logically rational, behaves rationally and never even makes the kind of mistakes that Kripke's Pierre does. He never misidentifies things, never thinks that what is one thing is actually two, and vice versa. The ideal agent is intended to be understood in such a way that it would be correct to ascribe to him any disposition that would typically be associated with having a certain belief. Suppose that this agent, A, replaces Pierre in Kripke's scenario. Unlike Pierre, A knows that the city he has previously called 'Londres' is the same city he now lives in and calls 'London'. He would not assert contradictory statements about London, and would come to think that parts of London are pretty and that parts are ugly. Let us suppose that although he finds parts of London pretty and parts of it ugly, he thinks that overall he can discount the ugly parts, because even the most beautiful city has its less attractive

areas. He assents to ‘London is pretty’, buys tickets for his friends to visit, recommends to others, etc.

Now, if Kripke were to ask, concerning this agent, ‘does A believe that London is pretty?’ then it seems clear that the answer would be ‘yes’. This might not seem to be a very interesting result, but it might seem to be in tension with my earlier claim that a belief ascription is a response to a proper question, and can only be evaluated with respect to that question. But in asking ‘Does A believe that London is pretty?’ I have not specified a purpose and it does not seem to be obvious from the context of my asking what the purpose is, so how can we feel so sure that the answer to the question is ‘yes’? My response to this is to say that in some cases, specifically those cases of ideal agents, who can be correctly described as having all the dispositions associated with a given belief, the ascription of the relevant belief will be true in all contexts. Even without having yet given any account of how the context of a question ‘Does A believe that S?’ narrows down the range of dispositions that is relevant to the ability of an ascription to allow B to predict an agent’s behaviour and responses, we can see that if the agent has *every* disposition that is associated with a given belief, then any subset of associated dispositions which would suffice for the truth of the ascription in its context is to be found among the agents’s disposition. I do not know how common ideal agents are, but I suggest that it is by focusing on ideal agents that has made it seem as though the truth or falsity of a belief ascription is a context-independent matter. The appearance of context-independence is, I suggest, an artefact of focusing narrowly on such examples. But instead of saying that the truth of an ascription to an ideal agent is a context-independent matter, we should simply say that we know a priori the meta-linguistic truth that for ideal agents the ascription will be true in all contexts.

I now turn to look at examples of agents who are less than ideal in order to show how I think context-sensitivity arises in realistic cases. In Examples 2 and 3 I will be focusing on the context-sensitivity that arises from differences in B’s interests, not from differences in B’s background

knowledge. So, in these examples I hold the background knowledge constant across the different contexts that I will look at. I will vary the background knowledge only in Examples 6 and 7, while holding constant the focus of interest there.

Example 2

There is another structurally similar example, originally provided by Eric Schwitzgebel (2002). In this example, Ellen has been learning Spanish for years and based on her exposure to the language is willing to sincerely assent to and assert that all Spanish nouns ending in ‘-a’ are feminine, which is the rough-and-ready rule that she has been taught. Ellen, has, however, come across certain words ending with ‘-a’ which are not feminine, but masculine, and she has learnt to use these words correctly with the appropriate grammatical agreements. She treats them as masculine. If one of the counter-examples to the rough-and-ready rule came to mind, she would not assent to nor assert that all nouns ending in ‘-a’ are feminine, but the counterexamples never come to mind when she is considering the question in the abstract. Clearly, Ellen is not, then, in my terminology, an ideal agent. Perhaps this is because she is not, at least in this matter, sufficiently reflective.

Consider the question ‘Does Ellen, or does she not, believe that all Spanish nouns ending in ‘-a’ are feminine? I say that it depends on the context. Suppose that B is asking the question because he wants to know whether she is likely to give a good oral presentation in Spanish on some topic, say, what she did in the summer holidays. B asks R ‘Does Ellen believe that all Spanish nouns ending in ‘-a’ are feminine?’ R knows that this is why B is asking the question and she replies ‘No, she does not believe that all Spanish nouns ending in ‘-a’ are feminine. This answer allows B to make the correct relevant predictions about Ellen’s behaviour, and so is a good answer. This is, therefore, a true report. I think that if B were to find out all the relevant facts about Ellen’s dispositions and capacities then he would say that R had spoken truly when she answered his question.

Suppose now that the context is altered so that R and B are talking about a lesson that Ellen has been asked to give to some new students of Spanish who are beginners. B wants to know how Ellen will perform as their teacher, then the same question ‘Does Ellen believe that all nouns ending in A in Spanish are feminine?’ would elicit a different correct answer. In this context, the true report would be ‘Yes, Ellen does believe that all Spanish nouns ending in ‘-a’ are feminine. I think that if B were later to find out all the relevant facts about Ellen’s capacities and dispositions, then he would agree that this had been a true report.

Example 3

I suggest that Kripke’s Pierre is an example of a non-ideal agent much like Ellen in the previous example. Again, this means that he does not have all the dispositions and capacities associated with a given belief, in this case, the belief that London is pretty. Pierre is, admittedly different from Ellen in an important way. Intuitively, Ellen could come to have more of the associated dispositions merely by reflecting on her knowledge and behaviour, at least in principle. Pierre could not come to see, merely by reflection, that the city he calls ‘London’ and the city calls ‘Londres’ are really just one city. I do think that this difference can, in some contexts, be semantically relevant to what belief ascriptions are likely to be true of an agent in various contexts. I shall return to this point at the end of the discussion of this example.

Pierre has some dispositions that would be associated with the belief that London is pretty, and some that would be associated with the contradictory belief. In a context where the first set of dispositions was relevant the report ‘Pierre believes that London is pretty’ would be true. In a context where the second set of dispositions were relevant, the same report would be false. Imagine, as an example of the former, someone asking ‘Does Pierre believe that London is pretty?’ where this person works for the

tourist board and is responsible for the design of the brochure that Pierre's view of London was originally based on when he was back in Paris. What the questioner wants to know is whether Pierre like the pictures of London in the brochure, would he be likely to recommend London (at least, the famous parts) to others, would he be likely to look at and use the brochure again in the future? He may not have an absolutely determinate list in mind, but he is interested in at least these sorts of things. If R knew his intentions in asking, he could say 'Yes, Pierre believes that London is pretty'. I believe that in this context the report would be true. But if the context were one in which, back in London, one of Pierre's neighbours asked 'Does Pierre believe that London is pretty?' because he wanted to know how Pierre feels when he walks around the neighbourhood and what he would say about London on the 'phone to his family back in France then, it would be right to say 'No, Pierre does not think that London is pretty'. It would be a true report. I think that the reporter would give this answer even if he were aware of all the relevant facts, so it is not the case that he gives the report out of partial ignorance.

What about *Kripke's* question to the reader of 'A Puzzle About Belief': 'Does Pierre, or does he not, believe that London is pretty??' Kripke, as we have seen, thinks that any answer to this question is unsatisfactory. It is this fact that leads Kripke to say that Pierre's case "lies in an area where our normal apparatus for the ascription of belief is placed under the greatest possible strain and may even break down".

I suspect that the appearance of a breakdown of the practice has more to do with Kripke's asking a question which would not normally arise in the way that Kripke has posed it. Why does Kripke ask this question? What is it that he wants to know about Pierre? The answer is, strangely, *nothing*. This, I think, is exactly where the problem with Kripke's question lies. There is nothing that Kripke wants to know about Pierre that he does not already know. Kripke is omniscient with respect to Pierre's dispositions and capacities. If the truth of 'Pierre believes that London is pretty' depended on some context-independent fact about Pierre that is not given

in the story, such hidden fact about an inner representation, then even given all the other facts about Pierre, the question might still arise, but it is just this view of what the truth of a belief ascription depends on that I have been concerned to reject. If, on the other hand, a belief ascription is merely a convenient way of summarising some other information about Pierre's possible behaviour and responses in various situations, then since we already have that information, and know that Kripke has it, the question would not really arise. It is not, in the terminology developed earlier, a proper question. If I tell you everything there is to know about Pierre's dispositions, and you still ask me 'Yes, but does Pierre, or does he not, believe that London is pretty?', then I have to say that I do not understand your question.

It might be said, in response to this, that the question could still be asked by someone who did not know about Pierre's dispositions. But, if there is no determinate purpose in the question, the question is not proper and should be rejected. If, as seems possible, the questioner simply wants a summary statement of Pierre's overall cognitive situation, then this is where we have to leave the vocabulary of belief ascriptions. If there is a breakdown it is here. But it is important, I feel, to acknowledge that the breakdown occurs because of a combination of the fact that Pierre is not an ideal agent with the particular purpose of the question.

In my next two examples, I want to discuss the issue of the relevance using a particular name in a belief ascription, but before I do that I want to say something general about the related issue of substitutivity, then I will use the examples to illustrate and support what I say.

I think that most of the debate about substitutivity occurs on the background of a framework that I reject, and, furthermore, one that is rejected by an increasing number of philosophers working in this area. It is a dominant view that belief ascriptions are opaque, at least on a certain reading. Opacity is a property of an open sentence, that is, a sentence with a space that requires a singular term or other expression to produce a

grammatically complete sentence. I will focus on the case where a singular term is required because that is what has dominated the discussion. Let the open sentence be represented by ‘... is F’. Suppose that ‘a’ and ‘b’ are arbitrary singular terms that refer to the same item, then if ‘a is F’ and ‘b is F’ can differ in truth-value, then ‘... is F’ is an opaque construction. If an open sentence is not opaque then it is transparent. Substitutivity (of referring singular terms) is said to hold for transparent constructions and to fail in opaque constructions. The open sentence ‘... is so-called because of his size’ is a classic example of an opaque construction. ‘... is tall’ is an example of a transparent construction. These two examples are standard and I do not dispute them.

But how can it be meaningfully asked: ‘Are belief ascriptions opaque?’ Belief ascriptions on the view here are context-sensitive. How is the question to be couched so that it accommodates this fact? If an ascription is true or false only relative to a context, then we cannot talk of the truth-value (of an open sentence when completed with a singular term) simpliciter. I suggest that we only talk of the truth-value of a particular utterance of a belief ascription sentence. The same idea of substitution does not quite make sense in this context, but we can talk about what the truth-value would have been had a different term been used. This seems like a perfectly natural extension of the idea of substitution. Moreover, there is no a priori reason to assume that if an ascription would change truth-value on substitution of co-referring terms in some contexts that it could do so in all contexts. Perhaps in some contexts a ‘substitution’ of ‘b’ for ‘a’ would change truth-value, and for other contexts it would not. This is what I think is the case, and I will try to illustrate this in the following two examples.

Example 4

The Roman orator Cicero is also called Tully, but only relatively rarely and by those who are classically educated. Suppose that A would dissent from ‘Cicero is Tully’ because she thinks that in the relevant settings these

names are used to refer to two different people. She would assent to 'Cicero was born in 106BC'. Suppose B asks R, 'Does A believe that Cicero was born before 100BC?' If the context were one in which it was wondered what A would say in her school quiz, then the answer to the question should be 'yes'. And I take this to mean that 'A believes that Cicero was born before 100BC' would be true in this context. But what about the answer to the question 'Does A believe that Tully was born before 100BC?'

It seems to me that consideration of standard examples of belief ascriptions in the context of debates about substitutivity would incline us to say that the answer to the latter question is 'no'. Sometimes it is said that the fact that A would assent to 'Cicero was born before 100BC' and dissent from 'Tully was born before 100BC' is sufficient to say that 'A believes that Cicero was before 100BC' is true and 'A believes that Tully was born before 100BC' is false *simpliciter*. Such claims have been made against a background assumption of context-independence. I think that the most we are entitled to say so far is that there may be some contexts in which switching the names 'Cicero' and 'Tully' in the question would lead to different and opposing correct answers; 'substitutions' would fail. But that the substitution would alter the truth-value in some contexts, obviously does not imply that it would do so in all contexts. I think that, in fact, regarding the second question, 'Does A believe that Tully was born before 100BC?', the context is so far inadequately specified to give a definite answer.

If I am correct in my claim, which I will try to substantiate in the next paragraph, that the context described above is inadequately specified to assess the truth-value of the ascription in question, this shows that what matters to the truth-value of a belief ascription goes beyond what the purpose of the report is, for in this example, I have already fixed what the purpose is. In this case, R and B are focused on A's verbal dispositions, in particular, on how she will perform in her quiz. It might be tempting to think that since the focus of interest is on her verbal dispositions, and since

her verbal dispositions vary precisely with respect to the two sentences ‘Cicero was born before 100BC’ and ‘Tully was born before 100BC’, that this means that in this context, as it has been specified so far, ‘substitution’ of one name for the other would alter the truth-value. But it is not so clear.

Example 5

Suppose it was assumed by both R and B that the quizmaster was most likely to ask questions about Cicero using the name ‘Cicero’, then I think that ‘A believes that Cicero was born before 100BC’ would be true, while ‘A believes that Tully was born before 100BC’ would be false. This means that the substitutions in this context fail to preserve truth. But suppose instead that R and B both standardly use the name ‘Tully’ for Cicero, and furthermore, that they both know that A is confused in thinking that there are two different people here. In that case, given the assumption about the purpose of the report, I would say that the ascription ‘A believes that Tully was born before 100BC would be true’, her dissent to ‘Tully was born before 100BC’ notwithstanding. It would be true in this context because it would, given the relevant assumptions shared between R and B, lead B to make the correct predictions about the behaviour of A’s in which he is interested, at least if all things worked out as they should.

I think that of the three assumptions,

- (1) The question will be asked in the quiz using the name ‘Cicero’
- (2) R and B normally use the name ‘Tully’ for Cicero
- (3) A is confused about the identity

they are individually necessary for the truth of the ascription, and, given the purpose of the ascription, jointly sufficient. If (1) were abandoned, the report would be false since it would lead to the wrong prediction; it would lead to the prediction that A would get the question in the quiz correct, but she would not. If (2) were abandoned, the use of ‘Tully’ in the embedded ascription would tend to generate the prediction that A would assent to

‘Tully was born before 100BC’, but she would not. Regarding (3); in many contexts, especially those specifically focused on an agent’s verbal dispositions, such an ascription would lead to the expectation that A would assent to ‘Tully was born before 100BC’, but the background assumption about A’s confusion forestalls that possible implicature in this context.

It may be objected to my discussion of the previous two examples that it not really so clear that the intuitions that I have said I have are universally shared. A number of people I have discussed the examples with agree with my intuitions, but a number have also said that they have less clear intuitions. In response to this I would say that even if there are doubts about the intuitions that I have regarding the above two examples, if the doubts are provoked by the changes in the contexts that I describe, then we can explain the shakiness of the intuitions as being due to precisely the complexity and abundance of factors that the belief ascriptions are sensitive to. This would mean that some of my more specific hypotheses about how various factors influence the truth of an ascription are wrong, but would still allow me to maintain my more general claims about the kinds of factors that are relevant to the truth of an ascription.

I now want to look at an example of a belief ascription as an explanation of an action.

Example 6

Suppose that B sees A suddenly hide behind a wall. R who also sees this sees their nephew Mike pick up a stone and move as if to throw it at A. B also sees Mike, and recognises him, but does not see what he does with the stone. B asks ‘why did A just suddenly hide behind that wall?’ R answers: ‘because she believed that Mike was about to throw a stone at her’. In this context, the ascription intuitively seems true. The ascription seems true even if we know that A did not know that it was Mike who was about to throw the stone, perhaps mistaking him for someone else or simply failing to recognise him, so she dissents from ‘Michael was about to throw a stone

at me'. I think that the ascription is intuitively true because it explains the A's hiding.

We can alter the context so that the ascription fails to explain the action. Suppose instead that B fails to recognise Mike in the street and thinks that he is inside watching TV, suppose that R knows that B mistakenly believes this. My intuition is that in such a context, the ascription would be false. The intuition is felt more strongly I think if we focus on the fact that A does not recognise Mike. I suggest that it is because in such a context, R's response 'A believes that Mike was about to throw a stone at her' would leave B baffled. It would fail to explain the action.

How could R explain A's action in this context? She could perhaps say 'A believed that she was about to have stone thrown at her', or perhaps, 'A believed that the man in front of her was about to throw a stone at her'. What is relevant is that from A's perspective someone within what she believes is stone-throwing distance is about to throw a stone in her direction. But there need not be any specific way she thinks this. We do not need to say this in our belief ascription, but the ascription must, in its context, allow B to infer it. I suggest that any way this could be achieved would render the ascription true.

The use of indexical expressions in the embedded sentence of a belief ascription is one very prominent way of making quite explicit the agent's perspective. Castaneda (1966, 1968) and John Perry (1977, 1979) have stressed that belief ascriptions containing indexicals are special. Perry gives examples of first-person belief ascriptions in which he says the use of indexicals to characterise the agent's belief is essential. He describes a scenario in which he follows a trail of sugar on the floor around a supermarket to tell the person he assumes is accidentally spilling it to stop. He comes to realise, after some several laps, that it is his bag of sugar that is leaking. He says "I believed at the outset that the shopper with a torn bag was making a mess. But I did not believe that I was making a mess.

[...] When I came to believe that, I stopped. My change in beliefs seems to explain my change in behaviour.

Perry says that in reporting this belief that he came to have the indexical is essential in the sense that replacement of it by any coreferring term, even a directly referential one, destroys the force of the explanation. For example, if he had said instead “I came to believe that John Perry is making a mess” then he would no longer have explained why he had stopped.

I want to understand exactly what it means to say that the indexical is essential in this explanation. Perry is slightly equivocal on a point that I think is crucial. Sometimes he says that using another term in place of the indexical destroys the explanation *simpliciter*. But sometimes he says, that it would destroy the force of the explanation, *or at least requires certain assumptions to be made to preserve it*. The caveat is important. Perry himself concedes that in order to explain his stopping with ‘I came to believe that John Perry was making a mess’ he would need to add ‘And I believe that I am John Perry’. He says that the only reason that ‘I came to believe that John Perry was making a mess’ seems to explain the action is our natural assumption that he did believe he was John Perry. He says that if he had said: ‘I came to believe that de Gaulle is making a mess’ that would not have explained his stopping at all, but that if he had added ‘and I believe that I am de Gaulle’, the explanations would be on a par. He concludes that ‘I came to believe that John Perry is making a mess’ does not explain the action at all.

The argument is fallacious. To show that given the assumption that he is de Gaulle, the explanations *would be* on a par does not show that they *actually are* on a par. For we do actually assume that he believes he is John Perry and that he is not Charles de Gaulle. I think that the correct thing to say is that whether or not a given ascription explains an action depends on such assumptions essentially. Explanations are essentially situated in a context. For even ‘I came to believe that I am making a mess’ does not explain the action unless we also make assumptions about what

Perry was trying to do. If we had assumed that Perry was very pleased to see a mess and thrived in untidy supermarkets, that ascription would obviously no longer explain his action. The indexical in the ascription that explains the action in a context is neither necessary nor sufficient for the ascription to explain the action. Indexicals are essential only in that some belief ascription, either the explaining one, or one assumed to be true in order to allow the explaining one to explain in its context must contain an indexical in the embedded sentence of the ascription.

I think that what is special about belief ascriptions involving indexicals is that given other very natural assumptions which form part of the necessary background context, the belief ascription is virtually guaranteed, whereas other, non-indexical terms substituted in their place would given only vary natural assumptions destroy the force of the explanation.

4.8

Defending and clarifying the meaning of the truth-condition

In stating the truth-condition for a belief ascription I said that an ascription is true if and only if it provides a ‘good’ answer to the proper question of its context. I would now like to clarify what is meant by ‘good’.

In my discussion of the examples I relied on the idea an ascription which was a good answer to one question could be a bad answer to another question and that the truth-value of the ascription would change accordingly. In the case of Ellen of Example 2, I suggested that when the purpose of the question was to find out whether Ellen would make grammatical mistakes or not, ‘Yes, Ellen believes that all Spanish nouns end in ‘a’ was not a good answer (and so was not true), but when the purpose of the question was to find out what she would say about the grammatical rule about Spanish nouns ending in ‘a’, ‘Yes, Ellen believes that all Spanish nouns end in ‘a’ was a good answer (and so was true). Whether it is a good answer or not had to do with whether it allowed the audience to make the right predictions.

All of the other examples implicitly employed the same reasoning, except for Example 6, which dealt with explanation rather than prediction, so that whether the ascription were true or not depended on whether or not it provided a good answer to the question of its context, which amounted to whether or not it provided a good explanation of an action, since in this example the question was a request for an explanation of an action.

What I have not done so far is give an account of what constitutes a good explanation or affords the basis of a good prediction. Whatever I say by way of explication of such concepts I must not presuppose the notion of truth, otherwise my account of belief ascriptions would be circular. For then I would be saying that a belief ascription is true just in case it provides a good answer to its proper question and a good answer involves giving a good explanation or good basis for prediction when that is what is required, and that a good explanation must be a true one, and a good basis for prediction must be true. Such an account of the truth-conditions of belief ascriptions would presuppose a prior grip on whether they were true.

On the other hand, without presupposing the notion of truth, I may face objections such as the following. Suppose John picks up an umbrella and someone who sees this asks 'Why did John pick up the umbrella? We can suppose that this is a proper question arising out of a desire for an explanation of John's action. If I were to answer 'John believes that it's raining' that may seem like a good explanation. It seems like a plausible explanation and may satisfy the questioner. Nevertheless, intuitively, the ascription may be false. For intuitively, John may have believed that Mary had left the umbrella behind and wanted to take it to her. So how can I rule out the first ascription as a good explanation without presupposing the truth of some particular ascription? (The difficulty arises because it is standardly thought that good belief explanations presuppose the truth of the ascription which explains, whereas I want to account for the truth of an ascription in terms of whether it explains).

I am unable to give a complete answer, but I think that ‘John believes that it’s raining’ cannot count as a good explanation because it would mislead the questioner about what else to expect in John’s behaviour, even though such future behaviour was not required in the question. For to assume that the ascription is false, we must assume that most of John’s other behaviour would not be consistent with the ascription. To count as a good explanation, the ascription has to be able to account for John’s other behaviour. Is there a difficulty in suggesting this response however? For have I not already implied that in asking a proper question which is a request for prediction, the audience is only interested in certain aspects of an agent’s behaviour, rather than the whole of an agent’s behaviour?

If we consider Ellen again, we cannot require that the ascription is consistent with all or most of her behaviour since we have assumed that conflicting ascriptions can be true of her, depending on the interests of the audience.

My response is to say that we must treat explanation ascriptions differently from prediction ascriptions. In the case of an explanation request, the ascription must not suggest that the agent would exhibit other behaviours that he would be unlikely to. In the case of prediction ascriptions, the ascription can in fact only be sensitive to what the audience is interested in. He will not be misled about other aspects of an agent’s behaviour since he in asking a question understood to be a request only for predictions about certain aspects of the agent’s behaviour.

I think that such claims are given some support by considering whether the audience would, later apprised of all the facts about the agent’s dispositions, say that the report was a correct answer to his question or not. It is my intuition, that he would say that ‘John believes that it is raining’ was not a correct answer because it suggests that John would do other things that he would not in fact have done. In the case of a request for information about whether an agent would act in some particular way, an answer would count as correct if it allowed the correct prediction, whether

or not it was consistent with other aspects of the agent's behaviour, for the proper question is already understood as making much of the agent's behaviour irrelevant.

4.9

How my form of contextualism is different from Richard's

I have presented and discussed several examples of belief ascription scenarios in order to support the view that the truth-value of a belief ascription depends on contextual factors. But, even some of the theories that I see as falling under my target of criticism imply that belief ascriptions are context-sensitive. I now, therefore, need to explain how the context-sensitivity in my account is of a quite different kind from the context-sensitivity from the context-sensitivity of the theories that I consider fundamentally mistaken. Since I have discussed Richard's theory already, I will draw the contrast explicitly between my theory and his.

According to Richard's theory, belief ascriptions can vary in truth-value between contexts, specifically according to what counts as an appropriate translation of an agent's representations, which depends in turn on the communicative intentions of the belief reporter. But although the communicative setting of the ascription is given a central role in Richard's theory as it is in mine, the role is really quite different. For in Richard's theory the intentions determine what counts as an appropriate translation of some determinate and independently specifiable cognitive fact, specifically a fact about what representations the agent has in his mind. We only need to look at the way the restrictions on correlation functions are stated to see that his account presupposes that there is a simple fact of the matter about what representations an agent has in his representational system. Richard begins his description of the representational system as being like a blackboard with writing on it (what is written on it being the representations), but this is only a metaphor, and so cannot really convince us that the representations are something that we really have a prior grip on. He talks provisionally about sentences that an agent would accept, but

then realises that this will not serve his purposes, and he finally talks about ‘representations’, but does not explain how we determine what representations an agent has nor how ordinary belief ascribers come to know this.

My account however gives no role to mental representations at all, and does not assume that there is some single fact about an agent which a belief ascription manages to specify, whether directly or via some translation. On my account the contextual sensitivity arises from the fact that belief ascriptions are answers to specific questions, and what counts as an answer is itself a context-dependent matter.

When a belief ascription is an explanation of an action, the context-dependence of the ascription derives from the fact that what counts as an explanation varies from one context to another; we saw in the discussion of Example 6, for instance, that whether an ascription explains depends in part on background assumptions of the audience. What explains in one context can fail to explain in another, when only the background assumptions of the audience are varied. Sometimes, a belief ascription is an answer to a question which is a request for an explanation, so that it is a good answer, and is true, only when it is a good explanation. The truth-value of a belief ascription can therefore vary according to differences in the audience’s background assumptions.

When an ascription is given as an answer to a question which is designed to elicit information to allow the audience to predict some range of an agent’s possible actions, whether the ascription counts as a good answer depends on whether it allows the agent to make the right predictions, so again the truth-value of the ascription depends on an agent’s background assumptions and on his focus of interest. She will be interested in some but not all of an agent’s potential behaviour and reactions.

There is no single fact about the agent which is relevant to the truth of a given ascription in every context. What facts about the agent’s mental life

are seen as underlying a given ascription is itself a context-independent matter. That is how my account differs from Richard's and why it can be seen as abandoning the central assumption that I have, throughout the dissertation being trying to cast doubt upon.

The facts about an agent which are relevant to the truth of an ascription are a motley variety of facts about an agent's dispositions and capacities and are only selected by the context of the report itself.

4.10

De dicto and de re

Much of the literature on belief and belief ascriptions foregrounds a distinction between de dicto and de re belief and between de dicto and de re belief ascriptions. I will not discuss the distinction as it applies to belief directly, and focus only on the alleged distinction as it applies to ascriptions. So far in this dissertation I have scarcely mentioned this distinction, except in connection with Nathan Salmon's theory in Chapter 2. This might seem strange because it is so central to most discussion in the literature of ascriptions, and is intimately bound up with issues concerning substitutivity, and in particular, because it might be thought that my discussion of my examples and of the intuitions about truth-values that they are designed to elicit depend on how we read the ascriptions, on whether we read them as de dicto or de re. It might have occurred to a reader of my discussion of Example 6, for instance, that whether we think the ascription 'A believes that Tully was born before 100BC' is true or false depends on whether it is read de re or de dicto. Might it not be true read de re, and yet false read de dicto? So, shouldn't I have clarified which reading I was thinking of?

I have deliberately ignored the distinction because I do not think that it should be seen as central to the semantics of belief ascriptions as it is usually thought of as being. I do not have any decisive argument against

the coherence of the discussion, but I want to try to clarify the distinction and to try to cast some doubt on its relevance.

Sometimes the distinction is made syntactically, *de dicto* reports being of the form 'A believes that a is F', *de re* reports being of the form 'A believes, of a, that it is F'. If that was all that the distinction amounted to, then *de re* reports would seem to be excluded immediately I said that I was interested in reports of the former form. But the issue is not really avoided so easily. The distinction is meant to be a semantic and logical one, rather than one having to do with surface syntax.

Often, the above syntactic form is merely intended to be a canonical form for a belief ascription for which certain logical rules are stipulated to hold. Specifically, the position represented by 'a' is stipulated to allow substitutions of co-referring terms *salva veritate*, and to be open to quantification, while *de dicto* reports not allow substitution of all co-referring terms *salva veritate* and are not open to quantification. Often, it is said that *de re* reports are those which relate an agent to an item and make no implications about how the item is thought of/ conceptualised, while *de dicto* reports are those which relate an agent to a propositional content which specifies how an agent conceptualises an item.

But these distinctions are not equivalent. It does not follow from the fact that a belief context is opaque that it is *de dicto* on the second distinction. I think that all of the ascriptions that I have considered in my examples are opaque, or more correctly, will not allow, in a given context, *any* co-referring terms, even co-referring names, to be substituted for one another. But that does not mean that they relate the agent to a thought-content. It does not depend on how it would affect which proposition might be expressed by the embedded clause; it does not depend on specifying how the agent conceptualises an item. According to my account, changing a term in the embedded sentence of a belief ascription can change the truth-value of a belief ascription because it can change a good answer to a question into a poor answer to the same question. The explanatoriness of

an ascription can be sensitive to what terms we use in the ascription, but this need have nothing to do whether or not the change in terms preserves, or fails to preserve thought-content. Consider again Example 6. Whether or not 'Mike' can be intersubstituted for 'the man in front of her' depends on whether the audience, B, knows that Mike is the man in front of A. It does not depend on 'Mike' and 'the man in front of her' having the same content, or somehow managing to specify the same content. Failures of substitutivity do not imply that we have a *de dicto* ascription, if this is understood as meaning that we somehow convey an agent's thought-content or say what an agent would say.

It is usually intended that the *de dicto*/*de re* distinction is an exclusive distinction. It is less often explicitly said, but I think that it is intended that the distinction is exhaustive. If there is a real actually instantiated distinction here, I admit that it would be exclusive on either way of drawing the distinction above. But I obviously deny that it is exhaustive. In general, belief ascriptions are sensitive not to an agent's way of thinking (at least as this is generally understood), but to whether or not the ascription serves its specific purpose in context.

The question for me is not whether we can ultimately make sense of such a category of *de re* ascription where this is understood to mean an ascription where the way the object referred to by 'a' is irrelevant to the truth of the report, but whether it is at all commonly used. I will try to suggest that it is not all common. Some philosophers argue that *de re* reports are common, or are even the norm. There are at least three forms of argument that might plausibly be thought to show that *de re* reports are quite common. I do not think that any of them are ultimately convincing.

First, there is the view of those who think that a belief ascription of the form 'A believes that a is F', where 'a' is to be replaced with a name, are *de re* reports even though we can that they relate an agent to a proposition, in this case a singular proposition. Such a view is possible because, as I argued in Chapter 2, the *de re*/*de dicto* distinction (at least drawn is the

relevant way mentioned above) collapses when we are considering such singular propositions. But this view that *de re* reports are common comes from what I have tried to argue is an unwarranted theoretical assumption about the form of a belief ascription, without good empirical grounds. Such theories run counter to the data in a substantial way.

The second form of argument is more empirically grounded. It is the form of argument used by Eaker and Almog. They provide many examples where substitutions of co-referring terms which differ in content preserve truth. They focus on terms which differ in content because terms which express the same content are expected to preserve truth even on the *de dicto* assumption. However, even if substitutions of some terms that differ in content preserve truth, that, by itself, has no tendency to show that all do. But it is this latter that would be required to prove that the ascriptions are *de re*. If my theory is correct, the set of terms that can be intersubstituted *salva veritate* depends in each case on the context of the ascription, specifically on the audience's background knowledge. In some cases, two content-equivalent terms could fail to be intersubstitutable, and in some cases, two terms which differ in content could be intersubstitutable. If an ascription were truly *de re*, every co-referential would have to be intersubstitutable in a given context. I have seen no convincing examples of such cases. Eaker and Almog's form of argument seems to depend on the tacit assumption that if a report is not *de dicto* then it is *de re*. My examples show that a report may be neither.

Thirdly, it might be thought that ascriptions that employ names, pronouns, demonstratives and referential uses of descriptions do not aim to specify the content that the agent is related to, and are therefore *de re*. But this argument would only be valid if we assumed what I think is a false dichotomy. The alleged dichotomy is that we must either 1) specify how an agent conceptualises an object or 2) use a *de re* ascription. My examples can be used to cast doubt on this dichotomy. In Example 7, where A hides behind a wall, the ascription 'A believes that he (pointing at the man) is about to throw a stone at her', may be true and does not specify how A

thinks of him, and yet not any co-referential term could be substituted here *salva veritate*. An ascription can fail to be *de re* then if the wording matters to the truth of the ascription, without that being because the wording specifies an agent's way of thinking of the relevant object.

Not only are the arguments for the widespread occurrence of *de re* reports not convincing, once we acknowledge that ascriptions are designed to serve very specific purposes and do not merely try to specify some descriptive psychological state of affairs, we would have to wonder what purposes a *de re* ascription could serve. I do not suppose it is impossible to imagine some contexts where we might want to do this, but outside of a philosophical context, I doubt that we would often want to do this.

4.11

Conclusion

In this chapter I have presented the beginnings of an account of the truth-conditions of a belief ascription. The truth-condition makes the truth of an ascription relative to a context. Belief ascriptions are seen as answers to specific questions. The purpose of the question and the background assumptions of the audience are two distinct sources of the context-sensitivity of ascriptions. I apply my account to examples which are used illustrate various features of the account and show how the truth of an ascription is sensitive to many different factors in the context.

I have shown how the truth of an ascription depends in a complex way on an agent's dispositions and capacities to do, think and feel certain things, but have suggested that no reductive analysis of a given belief into specific dispositions is correct, so that my account cannot be seen as embodying the central assumption that this dissertation has been concerned to cast doubt upon, namely that belief ascriptions specify that a certain single cognitively specifiable fact about an agent obtains

I have tried to show that the de re dicto is not an adequate distinction to describe real belief ascriptions. I have suggested that very few reports are de re, but that does not mean that most are de dicto, as this is traditionally understood.

Conclusion

This thesis arose out of a consideration of Kripke's puzzle about belief. I wanted to understand why Kripke's argument seems (at least at a first reading) to work in spite of the fact that it leads to an apparent paradox. I wanted to try to identify the picture of belief and belief ascriptions that seemed to me to underpin it and much other semantic work on belief ascriptions. Identifying and articulating the assumption which underlies the picture was my aim in Chapters One and Three. In Chapter One I tried to make it plausible that such an assumption is at the heart of Kripke's paper. Chapter Three aimed to show that such an assumption is shared by several diverse theories of belief ascriptions and that this assumption is a source of problems for these theories.

Chapter Two tried to focus our attention on what it is natural and unnatural to say using various expressions using 'believes' and its cognates. The aim here was to suggest that perhaps the best way to understand belief ascription and the ordinary concept of belief is to focus on ascriptions and see how they actually behave in various real scenarios, rather than presuppose a metaphysical position according to which beliefs are reified or given a primary role and then assume that belief ascriptions talk about such 'beliefs'.

Chapter Four tries to say something positive about how belief ascription works after the analysis and criticism of the first three chapters. The account I give there is intended primarily to take seriously our intuitions about the truth value of ascriptions as they occur embedded in specific scenarios, and to show that an account of belief, which is free of the assumption that I say I find everywhere else, can be developed.

I have not tried to give an analysis of belief ascriptions, but merely a condition for their truth. I realise that in doing so I have produced an account

which does not adhere to many of the goals of other theorists working in this area. But I have hoped to give an account of belief ascription which is thoroughly context-sensitive, realistic and person-level oriented (as opposed to subpersonal).

References

- Almog, J. 2004 'The Proper Form of Semantics' in M. Reimer and A. Bezuidenhout (eds.). *Descriptions and Beyond* (Oxford: Clarendon)
- Almog, J. 2005 'Is a Unified Theory of Language and Thought Possible?' *Journal of Philosophy* 102 no.10: 493 -531
- Baker, L. 1995 *Explaining Attitudes* (Cambridge: CUP)
- Crimmins, M. 1992. *Talk About Beliefs*. (Cambridge, Mass: MIT Press).
- Crimmins, M. and J.Perry. 1989. 'The Prince and the Phone Booth: Reporting puzzling beliefs'. *Journal of Philosophy* 86: 685 -711.
- Donnellan, K.S. 1979 'Belief and the Identity of Reference' *Midwest Studies in Philosophy* vol. 14 issue 1: 275-288
- Dummett, M. 1973 *Frege: Philosophy of Language*. (London: Duckworth).
- Eaker, E. 2002 *Reporting Attitudes* PhD thesis UCLA
- Eaker, E. 2004 'Reporting Attitudes'. Unpublished paper.
- Evans, G. 1982 *The Varieties of Reference* (Oxford: Clarendon Press).
- Frege, G. 1948 'Sense and Reference' *Philosophical Review* 57: 209-230.
- Frege, G. 1956 'The thought: a logical enquiry' *Mind* 65: 289-311
- Kaplan, D. 2001 'De Re Belief', unpublished paper given in 2001 San Marino Kaplan conference

Kripke, S. 1988 'A Puzzle About Belief' in S. Salmon and S. Soames (eds.) *Propositions and Attitudes* (Oxford: OUP) 102 -148

Geach, P. 1957 *Mental Acts: their Contents and their Objects* (London: Routledge)

Hornsby, J. 1997 *Simple mindedness: in defence of naïve naturalism in the philosophy of mind*. (London: Harvard University Press)

Lewis, D. 1981 'What puzzling Pierre believes' *Australasian Journal of Philosophy* 59: 283 -289.

Marcus, R B. 1983 'Rationality and Believing the Impossible' *Journal of Philosophy* 80, no. 6: 321 -338.

Marcus, R.B. 1990 'Some revisionary proposals about belief and believing' *Philosophy and Phenomenological Research* 50 : 133 -153

Marti, G. 2007 'Identity, Substitution and the Subject -Predicate Structure' in m. O'Rourke and C. Washington (eds.) *Situating Semantics*. (MIT: in press)

Moore, J. 1999 'Misdisquotation and substitutivity: when not to infer belief from assent' *Mind* , April 1999.

Nelson, M. 2005 'The problem of puzzling pairs' *Linguistics and Philosophy* 28: 319 -350.

Over, D.E. 1983 'On Kripke's Puzzle' *Mind* vol. 9, no.366: 253 -256

Perry, J. 1979 'The problem of the essential indexical' *Nous* 13: 3 -21.

Richard, M. 1988 'Taking the Fregean seriously' in D.Austin (ed.) *Philosophical Analysis*. (Dordrecht: Kluwer, 1988), 219 -239.

Richard, M. 1989 'How I say what you think'. *Midwest Studies in Philosophy* 14:317 -337.

- Richard, M. 1990 *Propositional Attitudes: An Essay on Thoughts and How We Ascribe Them*. (Cambridge University Press).
- Richard, M. 1997 'Propositional Attitudes' in B. Hale and C. Wright (eds.), *A Companion to Philosophy of Language*. (Oxford:Blackwell), 197 -226.
- Ryle, G. 1949 *The Concept of Mind* (Chicago: the University of Chicago Press).
- Salmon, N. 1986 *Frege's Puzzle*. (Cambridge, Massachusetts:MIT Press).
- Saul, J. 1999 'The road to hell: Intentions and propositional attitude ascription' *Mind and Language* 14: 356 -375
- Schiffer, S. 1977. 'Naming and Knowing' *Midwest Studies in Philosophy* 2: 28 -41
- Schwitzgebel, E. 2002 'A phenomenal, dispositional account of belief'. *Nous* 36, no.2: 249 -275
- Sider, T. 1995 'Three problems for Richard's theory of belief ascriptions' *Canadian Journal of Philosophy* 25: 487 -513.
- Soames, S. 2002 *Beyond Rigidity: the unfinished semantic agenda of Naming and Necessity* (New York: OUP)
- Sosa, D. 1996. 'The import of the puzzle about belief' *Philosophical Review* 105: 373 -434.
- Steward, H. 1997 *The ontology of mind: events, processes and states*. (Oxford:Clarendon).
- Taschek, W. 1988. 'Would a Fregean be puzzled by Pierre?' *Mind* 97: 99 -104
- Taylor, K. A. 2002 'De Re and De Dicto: against the conventional wisdom'. *Philosophical Perspectives*, 16: 225 -265.

Wettstein, H. 1995 *Has semantics rested on a mistake? and other essays*
(Palo Alto, Cal. : Stanford University Press)

Wettstein, H. 2004 *The Magic Prism* (Oxford: OUP)

Vendler, Z. 1957 'Verbs and Times' *Philosophical Review* vol. 66, no.2
:143 -162