

IN DEFERENCE TO REFERENCE

by

Gabriel Mark Aurel Segal

B.A., London University, 1981
B.Phil., Oxford University, 1984

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND PHILOSOPHY IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1987

© Gabriel Mark Aurel Segal 1987

The author hereby grants to M.I.T. permission to reproduce and to
distribute publicly copies of this thesis document in whole or in part.

Signature of Author _____

Department of Linguistics and Philosophy
6/1/87

Certified by _____

Ned Block
Thesis Supervisor

Certified by _____

Noam Chomsky
Thesis Supervisor

Accepted by _____

George Boolos
Chair, Departmental Graduate Committee

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUN 26 1987

LIBRARIES

Archives

Abstract

This dissertation consists in three separate papers.

The topic of 'In Deference to Referents' is the semantic structure of belief reports of the form 'A believes that p'. I argue that no existing theory of these sentences satisfactorily accounts for anaphoric relations among expressions inside and outside of the embedded complement sentences. I propose a new account of belief reports, which assigns to embedded expressions their normal semantic values, but which also exploits Frege's idea of using senses to explain the apparent failures of extensionality in the reports.

In 'Who's Afraid of Narrow Minds?' I defend the thesis that the propositional attitudes ascribed by folk psychology possess a level of content ('narrow content') that is independent of anything outside the subject. I examine various objections to that conclusion, and show that they fail.

In 'Seeing What is not There' I examine the nature of the representational contents attributed to perceptual states by the computational theory of vision. I discuss Tyler Burge's claim that such contents are not narrow, but depend essentially on aspects of the external environment. I show that the claim is false, and that the explanatory power of the theory of vision depends upon its employment of a notion of narrow content.

Acknowledgments

A large number of people have helped a great deal in the preparation of this dissertation. It is difficult to express gratitude in amounts that are duly vast, without sounding smarmy. If I fail in my attempt to do so, too bad.

I would like to thank:

- The members of my thesis committee, Ned Block and Noam Chomsky. Both have been generous in the extreme with their time and energy. Without their guidance the dissertation would have been considerably more feeble than it is, had it been at all.

- George Boolos, who has helped me with every aspect of my philosophical thought, and in numerous other ways. His ability to make accessible to me conceptually and technically difficult matters is unsurpassed, as is his kindness.

- Jim Higginbotham, in large measure, who seemed always to be ready to offer explanations, advice and help.

- Jerry Fodor for much comment, criticism and discussion, during the infancy of this work, and for being a source of inspiration.

- Richard Larson for discussions frequent, varied and often useful, and for substantive help with matters of semantics and syntax.

- Peggy Speas for being ever available as a sounding board, for explanations of linguistics and for the main ideas in the section of 'In Deference to Referents' that is drawn from a paper that we co-authored.

- Rob Cummins who read draft after draft of 'Seeing What is not There', and gave clear, lengthy, detailed and invaluable comments on each one. Without his generosity a substantial part of this thesis would scarcely be.

- James Blevins, Ned Block, George Boolos, Sylvain Bromberger, Richard Cartwright, Noam Chomsky, Alex George, Jim Harris, Jim Higginbotham, Paul Horwich, Richard Larson, Dan Leary, Jay Lebed, Sarah Patterson, Thomas Ricketts, Jorge Rodriguez, Marc Ryser, Peggy Speas, Carol Tenny, Judith Thomson, Tom Uebel and Crispin Wright for searching criticisms, helpful comments and/or valuable discussions relating to 'In Deference to Referents'

- Ned Block, George Boolos, Sylvain Bromberger, John Carriero, Noam Chomsky, Adrian Cussins, Jim Higginbotham, Jay Lebed, John McDowell, James Page, Sarah Patterson, Mike Smith and Thomas Uebel for contributions of the same sort to 'Who's Afraid of Narrow Minds?'

- Mike Antony, Ned Block, George Boolos, Noam Chomsky, David Kirsh, Jay Lebed, Sarah Patterson and Paul Pietroski for such contributions to 'Seeing What is not There'.

- Ellen Hildreth, of the M. I. T. Vision Laboratory for patiently informing me about the computational theory of vision.

- My teachers at Oxford: John McDowell, Chris Peacocke, David Pears and Peter Strawson. What philosophical wherewithal I now possess is largely owed to them.

- Colin McGinn, my first teacher, who taught me how to think.

- Gottlob Frege for sense and reference.

- Tyler Burge, to whose work I seem to turn at every turn.

- Concepta Siembab, Marilyn Silva and Debra Stephenson for generous and friendly help with various of the practical details involved in getting a dissertation done.

- Dan Leary for keeping word-processors, printers and disks in working order, and making them write the dissertation for me, as well as for philosophical discussion and comment.

- My parents for having me and keeping me.

I echo Michael Dummett in pointing out that although only I can take responsibility for the contents of the thesis, I may not in fact hold sole responsibility for any errors that appear in it.

TABLE OF CONTENTS

PAPER ONE: IN DEFERENCE TO REFERENTS	8
1 Introduction	9
1.1 Methodological Remarks	10
2 Frege's Problem	13
2.1 Constraints on a Solution	14
3 The Problem of Anaphora	18
4 Dummett's Solution	19
4.1 Criticism of Dummett's Solution	21
5 The Kaplanesque Proposal	26
5.1 Criticism of the Kaplanesque Proposal	28
6 Davidson's Account	32
6.1 Criticism of Davidson's Account	33
7 Higginbotham's Account	37
7.1 Criticism of Higginbotham's Account	41
8 Frege Vindicated?	45

PAPER TWO: WHO'S AFRAID OF NARROW MINDS?	49
1 Introduction	50
1.1 Background	51
1.2 Conceptions of Content	52
2 First Argument	59
2.1 Reply to First Argument	59
3 Second Argument	62
3.1 Reply to Second Argument	64
4 Third Argument	65
4.1 Reply to Third Argument	67
4.2 A Modified Burgian Treatment of Proper Names	71
5 Fourth Argument	76
5.1 First Case: Perception	76
5.2 Reply to First Case	78
6 Second Case: Reference	88
6.1 Reply to Second Case	89
6.2 Points of View	91

PAPER THREE: SEEING WHAT IS NOT THERE	96
1 Introduction	97
1.1 Burge's Claim	97
1.2 Preview	99
2 Marr's Theory	99
2.1 First Point	103
2.2 Second Point	104
2.3 Third Point	106
3 Burge's Argument	109
4 Against Burge's Strong Claim	114
4.1 The First Case	114
4.2 The Straightforward Interpretation	114
4.3 The Devious Interpretation	117
4.4 Recapitulation and Remark	121
5 The Weaker Claim	123
5.1 Second Case	123
5.2 Undermining of the Devious Interpretation	125
5.3 Defence of the Straightforward Interpretation	128
6 A Specific Example	131
7 Exegesis	136

PAPER ONE
IN DEFERENCE TO REFERENTS

The wheel's still in spin
And there's no tellin' who that it's namin'
- Bob Dylan

In Deference to Referents

1 Introduction

The chief purpose of a semantic theory for a language, or a set of sentences, is to explain how the semantic features of whole sentences derive from those of their parts. Such explanations require a set of rules that allow for the calculation of the semantic value of a complex expression on the basis of its semantic structure, or logical form, and the values of its semantically simple parts. The topic of this paper is the semantics of propositional attitude attributions, particularly 'believes that' sentences, like (1) and (2) below. I want to know the logical form of these sentences, and the conditions that determine the semantic values of their constant parts, 'believes' and perhaps 'that'.

Two enterprises might lead a philosopher to seek for a semantics for 'believes that' sentences or indeed for any other sort of sentence. One such enterprise, associated with Frege, and these days with Quine, is the development of a new language more suitable for scientific purposes than the natural language which is the initial object of study. For one with such a concern the choice of a semantics for 'believes that' sentences will be guided by the suitability of the resulting canonical

idiom - that is, the idiom that perspicuously displays logical form - for use in a psychological theory. The enterprise that I shall be concerned with in this paper, however, is the other one.

The second enterprise is the one Davidson distinguishes with the words: 'the task of a theory of meaning as I conceive it, is not to change, improve or reform a language, but to describe and understand it'.¹ In seeking for an account of the semantics of attitude attributions I seek to understand the workings of the actual English sentences.

The main purpose of this paper is to examine a number of alternative accounts of the semantics of English 'believes that' sentences. My aim is partly to promote the attitude expressed in the title of this paper, a deference to referents, and tentatively to offer a proposal in keeping with that attitude. But my aim is also to enhance our grasp of what a good explanation in the theory of meaning is, by looking at the kinds of evidence and arguments that may be relevant to it.

1.1 Methodological Remarks

Before launching into an examination of concrete cases, I want briefly to discuss in the abstract the question of what kinds of evidence will be relevant to the evaluation of specific theories of the logical form of natural language sentences.

What is essential to a semantic theory, in the first place, is that it

1. Davidson, D., 'Truth and Meaning', Synthese xvii, 1967.

succeed in correctly predicting the observable semantic features of whole sentences. The form that such predictions take will depend upon what kind of theory one is working with. So for example, a theorist working within a Davidsonian theory of meaning - and I take myself to be such a theorist - will try to prove the familiar T-sentences: 'S is true iff p', where S is a structure-revealing description of an object language sentence, and p is a statement, in the theorist's metalanguage, of exactly the conditions under which that object-language sentence is true. What is crucial here is to get the T-sentences right. When we are in the fortunate position of constructing a theory of meaning for our own home language, the evidence is not hard to collect, for it is not hard to tell when a T-sentence is correct. The trick is to construct a theory that proves it.

What further evidential constraints one wants the theory to be answerable to, will depend upon exactly what is to count as a good description and explanation of the linguistic phenomena under scrutiny. A methodological conservative might seek only to articulate a body of knowledge possession of which would suffice to interpret utterances made by speakers of the language. For the methodological conservative two semantic theories that prove exactly the same T-sentences, but employ different axioms and mechanisms, are likely to be as good as each other. They both get the right results. Each achieves what needs to be achieved, but they do so in different ways. There is then no need to seek for further evidence that would count for one theory and

against the other.²

From the point of view of the methodological liberal, on the other hand, there is in principle no end to the kind of evidence that might be relevant to the evaluation of a semantic theory. There are also all sorts of general constraints on good explanation, that can be applied in semantics just as they can in any other domain.

I believe that there is much to be seen from the liberal perspective, that might otherwise remain hidden. Since I see no way of saying, in advance of a search, what might count for or against a given semantic theory, I see no reason to think that such a search would be a waste of time. On the contrary, I think that a detailed examination of what may count for or against particular proposals can only bring illumination, both of the natural language itself, and of the exigencies of understanding how that language works.³

The liberal position will stand or fall by the quality and amount of the results yielded by research that follows its guidance. This paper as a whole may therefore be taken as a partial defence of liberalism.

2. The archetypical methodological conservative is Davidson. See in particular his 'Belief and the Basis of Meaning', Synthese 27, 1974. See also Quine, W. V. O., 'Methodological Reflections on Current Linguistic Theory', in Harman, G., and Davidson, D., eds. The Semantics of Natural Language, Humanities Press, New York, 1972, for a different version.

3. For a defence of a liberal position, see Chomsky, N., Reflections on Language, Pantheon, New York, 1975, and my 'The Topiary Theory of Language', in Fukui, N., Rapoport, T., and Sagey, E., eds., M. I. T. Working Papers in Linguistics Vol. 8.

2 Frege's Problem

We turn now to semantics.

(1) Ralph believes that Tanya is a terrorist

(2) Ralph believes that Patty is a terrorist

Patty and Tanya are the same person. Let us suppose that Ralph is ignorant of this, and that he has heard, from what he considers to be a reliable source, that Tanya is a terrorist. On the other hand, he firmly believes that Patty is no terrorist. In some circumstances it would be wrong for me to report Ralph's belief as the belief that Patty is a terrorist. Let us suppose that the wrongness of the report amounts to its falsity, rather than, say, its being merely misleading, or lacking in explanatory value.

It should be noted that such a supposition is not mandatory. One could hold that, contrary to first appearances, intersubstitutions of codenotational⁴ expressions in the content sentences of propositional attitude reports do preserve truth value. Appearances to the contrary would then be explained by pragmatic, rather than semantic, considerations. It would be worth exploring the consequences of such a

4. I shall use 'denotes' and 'has as semantic value' interchangeably, to describe the relationship that holds, in extensional contexts, on any roughly Fregean theory, between an expression and its semantic value. My use of 'denotes' is not standard, but is easily understood.

view.⁵ But that is not my present purpose. I take it that the view is prima facie less plausible than the alternatives and it is those that I wish to explore in this paper.⁶

So (1) and (2) may have different truth values, in spite of the fact that they are composed of expressions with the same denotations in the same order. Frege's response to this was to claim that expressions embedded inside content sentences (that is the 'that' clauses of propositional attitude reports) do not there denote their normal denotations, but rather denote their senses. That is, Frege claimed two things: one; expressions inside content sentences do not denote their normal denotations, and two; expressions inside content sentences do denote their normal senses.

2.1 Constraints on the Solution

Let us follow Frege in supposing that embedded expressions like 'Tanya' and 'Patty' have some feature, "sense", that differs from their normal denotation, and that explains how (1) and (2) can have different truth values. The first question I want to ask is this: How do embedded expressions contribute their senses to attitude attributions? In particular I want to know if it is necessary that senses play the role of denotations of the embedded expressions. Or is it possible that the

5. For some such exploration see Salmon, N., Frege's Puzzle, M. I. T. Press, Cambridge, 1986.

6. I owe the idea of this paragraph to Crispin Wright.

embedded expressions contribute their senses in some other way?

It might look as though the answer is mandated by the roles that sense and denotation are designed to play.

What role does an expression's denotation play in the semantics? From the point of view of a semantic theory, it just is what the expression contributes to the determination of the truth value of sentences in which it appears.⁷ and So the truth value of a sentence is a function just of the denotations of its component expressions. Suppose this is true of (1) and (2) in particular. Then whatever the expressions in (1) and (2) contribute to the determination of the sentences' truth values will just be what the expressions denote in those sentences. This will be true for, inter alia, the embedded expressions, like 'Tanya' in (1). What is a sense? It is what an embedded expression contributes to the determination of the truth value of the attitude attribution in which it appears. It thus looks as though the only way the embedded expressions could contribute their senses is by denoting them.

So it appears that there is only one possible answer to my question 'How do embedded expressions contribute their senses to propositional attitude attributions?'. They denote them. I think it is important to have another look at the argument, because it seems to me that something like it provides the underlying motivation for many modern

7. C. f. Evans, G., The Varieties of Reference, Oxford University Press, 1982, second chapter, and Davidson, D., 'Reality Without Reference', Dialectica 31, 1977. In my view Davidson makes somewhat too much of the point.

Fregean treatments of attitude attributions: that is treatments that have words inside content sentences denoting their normal senses, rather than their normal denotations. It seems simple enough: the truth value of an attribution is a function of, among other things, the senses of embedded expressions. Since it is also true that the truth value of a sentence is a function only of the denotations of its components, the senses of the embedded expressions must, in those contexts, serve as their denotations.

But the argument is invalid. Consider (3) and (4).

(3) '...e...' (4) '...e'...'

(3) and (4) are schematic representations of opaque attitude attributions. e and e' have different senses but the same normal denotations, so (3) may have a different truth value from (4). We may infer from this, by the Fregean principles, that the denotation of at least one component expression in (3) differs from that of one component expression of (4). But it is not entailed that the expressions that have different denotations must be e and e'. Perhaps, for example, something else in the complex denotes the sense of e or e'. That would account for the opacity effects without our having to deny that e and e' there denote their normal denotations. Or perhaps there is some other, more devious and complicated manner in which the denotation of some part of the whole attribution is affected by the senses of the embedded expressions.

So perhaps the right answer to our question is just this. An expression in a content sentence contributes its sense to the truth

value of the attitude attribution just by having it, and the denotation of something else in the attribution is affected by it.

There is thus the possibility of using senses to explain the opacity effects in a manner slightly different from Frege's. But does this idea get us anywhere? Is there anything interesting to be done with it? That depends upon whether there is in fact some way of implementing the proposal, that is, some way of actually doing the semantics. We would have to find some way of allowing the sense of one expression to be, or somehow to determine, the denotation of another, a way that successfully predicts the opacity effects.

One reason why one might feel inclined to search for an alternative account along those lines is this. There are various ways in which it appears that expressions even as they are embedded inside a content sentence, are denoting their normal denotations. If we could allow the sense of an embedded expression to affect the truth value of the attitude report without being the expression's denotation in that context, then we could leave the expression free to denote its normal denotation, without losing our account of the failures of substitutivity.

In the next section I shall present some linguistic evidence that suggests that embedded expressions do denote their normal denotations. Then I shall investigate how well a Fregean theory might assimilate the evidence, and whether some alternatives fare better.

3 The Problem of Anaphora

The linguistic evidence that indicates that embedded expressions denote their normal denotations comes from cases of cross-denotation among expressions inside and outside content sentences. The clearest examples are provided by what appear to be straightforwardly anaphoric pronouns, like that in (5)

(5) Betty believes that Barbara_i is bright,
and she_i is bright⁸

In (5) it looks as though 'she' is straightforwardly anaphoric with 'Barbara'. That is to say it looks as though the denotation of 'she' is determined to be whatever is denoted by its antecedent 'Barbara'. I have marked this with the subscripts 'i'. On the Fregean theory the denotation of 'Barbara' in (5) is a sense. But it certainly looks as though the denotation of 'she' is not a sense, but the person, Barbara. Assuming that a person cannot be a sense, we have a problem: 'she' and 'Barbara' cannot denote the same thing, if the one denotes a person, and the other denotes a sense.

8. 'She is bright' is meant to be outside the scope of 'believes'. In future examples possible ambiguities are always to be resolved in that direction, so that the conjunction takes one of its conjuncts outside the scope of the doxastic operator.

4 Dummett's Solution

Dummett,⁹ notices the problems of apparent coreference among expressions inside and outside content sentences, like that illustrated in (5). He offers the following solution on Frege's behalf.

Since the account of opacity effects depends upon assigning a sense to 'Barbara', that is what we must do. And Dummett wants to preserve the intuition that (5) contains the semantically monadic predicate (6):

(6) Betty believes that x is bright and x is bright

a predicate that can only be satisfied by a single thing. If that is so, then the expressions 'Barbara' and 'she', standing in for the place holders 'x', must have the same denotation. What Dummett denies, therefore, is that 'she' denotes the person Barbara. Instead it must denote the sense of the expression 'Barbara', which is the latter expression's denotation in (5).

The little sentence 'she is bright' in (5) now seems be attributing the property of being bright to a sense. That cannot be right. Dummett tries to rectify this by adjusting the semantic values of other things outside the content sentence, so that overall it ends up with the right truth conditions.

9. Dummett, M., Frege: Philosophy of Language, Duckworth, London, 1973, ch.9.

Dummett's first suggestion is that we assign to the predicate, 'x is bright' as it occurs outside the content sentence, a special semantic value. This special semantic value is a function that maps onto the True just those senses that are senses of expressions the denotations of which are mapped to the True by 'x is bright' in normal contexts.

In general, then, predicates occurring in contexts wherein they appear outside a content sentence, but are predicated of an expression the value of which is determined within a content sentence, get special semantic values. These values are functions mapping onto the True just those senses that are senses of expressions the denotations of which are mapped onto the True by the predicate in normal contexts. Now the utterer of a sentence like (5) will be saying something that is true, just in case, if I might be so bold as so to put it, what she thought she was saying was true. /

But Dummett finds himself unable to rest with this view. The view required the assignment of special semantic values to predicates outside content sentences, in the awkward cases we are considering. But predicates, like pronouns, bound variables, and subject expressions, can have their semantic values determined from within a content sentence.

Suppose that 'John is here and I did not expect that John would be here' is true. Dummett writes (ibid. p. 276) that from this

we can ... infer 'John is something that I did not expect he would be' (namely, here), which is to be analyzed as ' $(\exists F)(F(\text{John}) \text{ and I did not expect that } F(\text{John}))$ '.

And now the argument continues just as before. Predicates outside content sentences will have to be codenotational with those inside. Those inside denote their senses, so those outside must denote their senses as well.

We must conclude that predicates like 'x is a bright', in (5) denote, in both their occurrences, their senses. But the sense of a predicate is not its special semantic value. The proposed treatment which depended upon the assignment of special semantic values must, then, be abandoned. So, as Dummett says (p. 277) of a context like the second conjunct of (5):

we have to see the entire context, apparently transparent but occurring within a sentence in which an opaque context also occurs, as in reality opaque, with all expressions contained in it having their indirect reference [=denoting their senses]

How can the utterer of (5) now say anything at all? Dummett suggests that we see the whole sentence as prefaced by an implicit operator such as 'It is true that'. 'That she is bright' just names a thought, and 'It is true' works as a predicate the semantic value of which is a function that maps true thoughts onto the True.

What are we to make of this emended Fregean account? I have a general expression of skepticism, a technical objection, and a methodological criticism to offer.

4.1 Criticism of Dummett's Solution

Here is the general expression of skepticism. It is very hard to see

what mechanism could allow for the retroactive effects of future utterances upon present ones, of the kind Dummett is proposing. Suppose I say, at t , 'Fred is the smartest student', then, after pausing to consider the wisdom of my telling this to Fred, I go on to say, at t' , 'And he knows that he is the smartest'. Dummett cannot hold off the conclusion that my initial utterance of 'Fred' denotes its normal sense. For he is committed to my second utterance being opaque from the second word onwards. Since the first utterance is linked to the second one in just the way that spreads opacity, on Dummett's theory, it will be opaque as well. Suppose, on the other hand, that I had uttered 'Fred is the smartest student', at t , paused to consider the wisdom of my telling this to Fred, and then had not gone on to say anything about Fred's psychological states at t' . Well, then of course my utterance would have been transparent. We have one initial utterance, at t , compatible with two developments at t' . It is overwhelmingly natural to think that the semantics of 'Fred is the smartest student' is completely settled at t . And that is incompatible with the consequences of Dummett's theory.

There are, of course, many ways in which I can utter an expression e , ambiguously, then later say something to pin down a denotation for it. This happens with such mundane sentences as 'Before he spoke the King cleared his throat'. But it is plausible that the denotation of e is really determinate at the moment of utterance. All that can happen later is that the epistemic state of the audience may be improved. Even if I utter e with the intention of denoting one thing, a , then change my mind, and disambiguate so as to make it look as though I meant to

denote another, b, all this could be accounted for in terms of the epistemic state of the audience. I really denoted a, but brought it about that my audience thought I was denoting b.

The difficulty, moreover, becomes aggravated: consider the dialogue:

Me: 'Fred is the smartest student'

Other: 'And he knows that he is the smartest'

On Dummett's theory, both tokens of 'he' must denote a sense. But the first 'he' looks to be anaphoric to 'Fred'. Is there anything which prevents our familiar argument from implying that my utterance of 'Fred' denotes a sense? It appears not. For there is no visible feature of this case that distinguishes it semantically from the case where I say it all myself. It is the serial utterance of two parts of the sequence, by different speakers, which brings about the final predication of Fred. But I do not see a way of exploiting this so as to allow that 'Fred' and 'he' do not share a denotation. For what Other says stands in exactly the semantic relation to my utterance, as it would have if I had said it.

But surely the denotation of my utterance of 'Fred' cannot depend upon what Other goes on to say. Such a consequence must make a reductio ad absurdum of any view from which it issues forth.

A defender of Dummett might respond by suggesting that all sentences are really composed of an opaque component that names a thought, and the implicit 'it is true that' operator. If that were so, then there would be no question of whether a given utterance were opaque or transparent, which question would have to be settled by what happened later.

But the reply gives up too much. We can hardly suppose that expressions never denote ordinary external objects, without losing our already precarious grip on how we manage to say things the truth of which depends upon how things are in the world outside. And Dummett has something very like this point in mind, when he writes (op. cit. p. 198)

The denotation of an expression is its extra-linguistic correlate in the real world: it is precisely because the expressions we use have such extra-linguistic correlates that we succeed in talking about the real world.

Let us move on to my second objection to the Dummettian extension of Frege. Let

(7) Sandra believed she was going to be late

be uttered by someone other than Sandra. On Dummett's account 'Sandra', in virtue of its relation to an opaquely embedded 'she', must denote a sense. But which sense? In (7) 'she was going to be late' denotes the thought that Sandra believed. Moreover it does so, one supposes, because each of its components denotes a sense, and the senses denoted are identical to the senses that compose the thought to which Sandra is related. So 'she' should denote the sense that stands in subject position in the thought that Sandra believed. This will not be the sense of the word 'she', but the sense of the word 'I' in Sandra's mouth. For Sandra had a first person belief about herself. But 'Sandra' in someone else's mouth cannot denote that sense.

Dummett might perhaps respond by weakening the theory of anaphora. The apparent anaphoric relation in (7) and the like would be

reconstructed so that the expressions in question would not denote the very same senses, but would denote senses that present the same denotation. In (7) 'Sandra' would denote the sense that enables the speaker to think of Sandra, and 'she' would denote the sense that featured in Sandra's own thought about herself. But to concede this would be to undermine the whole Dummettian approach. On this proposal '...x... A believes ---x---' and its fellows would no longer be bona fide monadic predicates. And once one is prepared to give up anaphora in these cases, they can be accounted for without the problematic expedient of assigning indirect denotations to expressions outside content sentences (see the Kaplanesque proposal, explained below, next section).

Perhaps neither of my two objections to Dummett are unanswerable. But even so, the theory looks poor: it is complicated, and the complications (wholesale shifts of denotation, suppressed operators appearing) seem highly ad hoc. They lack independent motivation. Contexts outside clauses governed by opacity inducing verbs do not exhibit the features we would expect of them if they were opaque. Intersubstitutions of codenotational expressions with different senses, inside them, preserve truth (try it). The hypothesized wholesale shifts of denotation thus have no detectible properties.

So let us put aside Dummett's version of the Fregean proposal and look for something better.

5 The Kaplanesque Proposal

Let us backtrack to the beginning of Dummett's story, and see if the Fregean has other possibilities. The problem was that in (5) it appears that while 'Barbara' and 'she' are codenotational, one denotes a sense and the other denotes a person. Dummett tried to deny that 'she' denotes a person. But perhaps it is better to deny instead that the two expressions are really codenotational. Could it not be that, rather, they denote things that stand in some special, and specifiable relationship to each other?

The idea would be to have expressions inside content sentences denoting senses, but use some relation that holds between those senses and the expressions' normal denotations, to account for cross-referring. How would this go?

Let us try (8):

(8) $R('Tanya', Tanya)$

understanding it to mean something like 'the sense of "Tanya" represents Tanya'.

We can exploit (8) in a Fregean version of (1):

(9) $R('Tanya', Tanya)$ and Ralph believes_d
 $\ulcorner Tanya \text{ is a terrorist} \urcorner$

The corner quotes create an oblique context within which each expression takes a sense as its semantic value. The corner quoted 'Tanya is a terrorist' now denotes the sense of the sentence 'Tanya is a terrorist'. 'Believes_d' expresses a two place relation that holds between believers and the senses or "dicta" of their beliefs.

Anyone familiar with David Kaplan's paper 'Quantifying In'¹⁰ will recognize that I have constructed (9) under its influence. I will call the kind of treatment illustrated in (9) 'The Kaplanesque proposal'. But note that Kaplan himself made no such proposal, but used the materials in another way, for another purpose.

With all this apparatus in play it is easy enough to cope with anaphora, thus

(10) Ralph believes that Tanya_i is a terrorist,
but she_i is too honest

becomes

(11) R('Tanya', Tanya_i) & Ralph believes_d
['Tanya is a terrorist' but she_i is too honest]

with 'she_i' codenotational with 'Tanya_i'. One occurrence of 'Tanya' is within the content sentence, and there serves the Fregean purpose of denoting a sense, and another occurrence is outside, and allows for the

10. Kaplan, D., 'Quantifying In', in Davidson, D., and Hintikka, J., eds., Words and Objections, Reidel, Dordrecht, 1969.

straightforward treatment of anaphora.

I do not think, however, that (11) is a serious candidate for the logical form of (10). It surely cannot be necessary to have the expression 'Tanya' occurring no less than three times in a correct rendition of a sentence in which it occurs only once.

The underlying difficulty with (11) is that it includes the clause (8), 'R('Tanya', Tanya)', and therefore actually says that the sense of the expression 'Tanya' represents Tanya. This does not seem to be something that is literally said by the sentence (10). Rather it is the fact that the sense of 'Tanya' represents Tanya that allows a hearer to ascertain the correct denotation of the pronoun 'she'. So (10) functions because (8) is true, but it does not say that it is true.

The sensible course at this stage would be to pull (8) out of (11) and place it somewhere else in the theory of meaning. As long as (8) appears in the theory, it can easily be exploited to account for the anaphora in (10). We would simply take the denotation of 'she' to be, not the denotation of its antecedent 'Tanya', but the object that is represented by that denotation.

5.1 Criticism of the Kaplanesque Proposal

But how, exactly, is (8) going to figure in the theory of meaning? It must either be an axiom, or derivable from some axioms. Indeed, if the representation relation R is to be usable in general, then we need a method of deriving, for every embedded expression that determines an

extension, what is represented by the sense of that expression. That is, we need a general method of obtaining clauses of the form (12):

(12) $R(e, b)$

where e is an embedded expression and b is the object that the sense of e represents.

But it is not easy to arrive at these clauses. Consider for example (13):

(13) John said that he loves [his wife_{*i*}],
but he is rude to her_{*i*}

How are we going to get a clause that allows us to recover the denotation of 'her'? The simple disquotational axiom suggested by (8) will clearly not work here. What would we do with (14)?

(14) $R(\text{'his wife'}, \text{his wife})$

'His wife' unlike 'Tanya', does not have a fixed denotation, but will have its denotation determined by the context. The theory of meaning thus cannot simply state which object is represented by the sense of 'his wife'. Rather it will have to contain a rule that specifies, for each context, which object the sense of 'his wife' represents in that context.

And new complexities are presented by quantifier/pronoun constructions like those illustrated in (15).

(15)(i) John said he has [some sheep_{*i*}].
Olga will vaccinate them_{*i*}.
(ii) John said he has [a new wife_{*i*}].
You will meet her_{*i*}.

Which relation holds between the sense of 'some sheep' in (15)(i), and the sheep that 'them' denotes? And which relation holds between the sense of 'a new wife' in (15)(ii), and the denotation of 'her'?

I have no doubt that with ingenuity and hard work the dedicated Fregean could construct the appropriate mechanisms for associating extensions with senses. But I think that if we reflect upon precisely what information is required to get us to the right extensions, a non-Fregean line will appear much more promising.

In each of our examples the appropriate extension for the anaphoric pronoun is recoverable from the semantic interpretation that the content sentence as a whole would have had, had it been treated as transparent, and given its normal semantic interpretation. Thus 'she' in (10) denotes what 'Tanya' would have denoted had (10) been denotationally transparent. And 'her' in (13) denotes what 'his wife' would have denoted had (13) been transparent.

Similarly in (15)(i) and (ii) the appropriate extensions for the pronouns would be recoverable from their antecedents, if we gave the entire content sentences their normal transparent interpretations. For the theory of meaning will have the resources to recover the right extensions from transparent treatments of the content sentences, as is apparent from (16):¹¹

11. For an account of these resources see Evans, G., 'Pronouns, Quantifiers and Relative Clauses (1)', Canadian Journal of Philosophy, vii, 1977.

- (16)(i) John has [some sheep_i]. Olga will vaccinate them_i
(ii) John has [a new wife_i]. You will meet her_i.

Knowledge of what the interpretation of the content sentences would have been, had they been transparent, is thus sufficient to determine the extension of the anaphoric pronouns.

It is clear also that in some cases, nothing less than that will do. In the examples (15) you have to know what the transparent interpretation of the entire content sentences would have been, before you can figure out the denotation of the unembedded pronouns. It is impossible to figure out the extension of 'them' in (15)(i) without first ascertaining that 'them' denotes some sheep. And you will only figure that out if you give a transparent interpretation to the expression 'some sheep'. Moreover you will only figure out that the precise sheep in question are the ones that John owns if you interpret the two expressions 'he' and 'owns' as if they had occurred transparently.

Similar remarks apply to (15)(ii). You will only be able to ascertain the denotation of 'her' if you interpret each element in the content sentence as if it occurred transparently.

To account for the anaphora in the sentences (15) what we need to do is thus precisely equivalent to treating the whole contexts transparently. Nothing more and nothing less than this is required.

It seems then that rather than attempt to formalize the Kaplanesque representation relation, R, we should look for some way of accounting for the failures of substitutivity while yet allowing the embedded

expressions to denote their normal denotations.

6 Davidson's Account

The point which started off this whole discussion of alternative proposals, was that if we could allow the denotation of one expression to be sensitive to the sense of another, then we might be able to use senses to block the undesirable substitutions, while leaving embedded expressions free to denote their usual denotations. But how could we do this?

Donald Davidson's account of the logical form of 'says that' offers a very attractive answer to this question. Suppose that the logical form of (1) is really (17):

(17) Ralph believes that. Tanya is a terrorist.

with 'that' as a demonstrative. Davidson himself suggests an understanding of this that has no use for the notion of sense. In fact that is one of the main motivations for the proposal. But since we already have the notion of sense in play, we might as well just have the 'that' referring to the sense of the embedded sentence 'Tanya is a terrorist'. Now the opacity effects are easily explained. Since the sense of 'Tanya is a terrorist' differs from that of 'Patty is a terrorist', Ralph may believe that

Tanya is a terrorist

but not believe that:

Patty is a terrorist

Since the content sentence is also a sentence in its own right, we may assign to its component expressions their normal denotations, and thereby solve our problems with codenotation among expressions inside and outside content sentences.

Davidson's proposal, even as interpreted in the unDavidsonian manner, is, as I said, attractive. The reason it is attractive is that it seems to read the semantics explicitly off the surface structure, or very nearly so. On Davidson's account it is easy to see exactly what each familiar English word is doing, how each contributes to the whole.

But the whole story could not be so simple and elegant. That would be too much to hope for. And, indeed, Davidson's idea does have various problems.

6.1 Criticism of Davidson's Account

Davidson's proposal perhaps derives a modicum of prima facie plausibility from the fact that the indirect discourse 'that' and the demonstrative 'that' are spelled similarly. But this is quickly dispelled by noting that in many languages (e. g. the Romance languages) the complementizer of indirect discourse and the demonstrative have completely different orthographical and phonetic forms. In some languages (Japanese, Hebrew, Korean) the complementizer, unlike the demonstrative, does not even have the superficial form of a separate

word, but rather of a mere affix.¹²

In English also, in spite of the orthographical coincidence, the demonstrative 'that' has a cluster of properties that the 'that' of 'believes that' does not. Notice first that they are pronounced differently. One usually says

(18) Ralph believes *ðət* Ortcutt is a spy

shortening the 'that' to '*ðət*'. The demonstrative 'that', however, seems never to undergo this phonetic reduction, or destressing. One cannot say 'Hey, look at *ðət*', or:

(19) *Ralph believes *ðət* . Ortcutt is a spy.

which interprets (18) in Davidson's way.

Second, the 'that' of 'believes that' and 'says that' is deletable. One can say:

(20) Lemmy said he needed money

with the 'that' dropped. But the demonstrative 'that' cannot be dropped.

You cannot say (21)

(21) *Yes ... is what I said

with the 'that' deleted, instead of (22)

(22) Yes that is what I said

12. German and Finnish, on the other hand, are more like English in the relevant respect.

Now there is not, to my knowledge, any well-understood and generally accepted explanation of what determines the possibilities of deletion and destressing. But it is very plausible that, in both cases, the explanation will invoke the semantic role of the expressions involved. It looks as though the reason why one cannot delete the demonstrative pronoun 'that' is precisely because it has an important referring function. And I believe that similar remarks apply to the phonological characteristics. Stress often goes with semantic role; and it is hard to see what else could account for the stressing rules in this case.¹³

Although these objections, taken by themselves, are far from conclusive, they are very well worth noticing, for they provide an example of how semantics can interact with other parts of linguistics. The 'that' of 'says that' can be deleted and destressed, the English demonstrative pronoun cannot. This is simply empirical data. If it is true that the 'that' of 'says that' functions semantically like a demonstrative, as Davidson holds, then it must be false that what explains the phonological and syntactic facts is semantic role. Whether this is so or not is an empirical question at the intersection of the various parts of linguistics. The examples provide a nice demonstration of methodological liberalism at work, and point towards a rich field of evidence that may help us choose between competing semantic theories,

13. For a discussion of these and other syntactic difficulties with Davidson's proposal, see Speas, M., and Segal, G., 'On Saying *That*', in Mind and Language, 1986.

should we have such aspirations.

Another, and more serious difficulty with Davidson's proposal is that the English sentences do not, in important respects, behave semantically as they would if they were paratactic. Thus

(23) Everyone said that they loved their mothers

is open to two interpretations. It could mean that each person said something like 'They loved their mothers', where 'they' is not anaphoric with (or bound by) 'Everyone', but refers to some other group of individuals. But it can also be readily understood as

(24) Everyone_i said that they_i loved their mothers

(24) means that each person claimed of herself that she loved her mother (saying something like 'I love my mother!'). But the Davidsonian rendition of (23)

(25) Everyone said that. They loved their mothers.

is only interpretable in the first way. (Imagine someone pointing to an inscription of 'They loved their mothers' on a blackboard, and saying 'Everyone said that!'. How would you interpret this?)¹⁴

The point of the example is not that Davidson could not define a samsaying relation that would allow (25) to mean the same as (24). It is that (23) is not, as a matter of empirical fact, understood by English

14. This kind of example is due to Higginbotham, J., 'Linguistic Theory and Davidson's Program in Semantics', in LePore, E., ed., The Philosophy of Donald Davidson: Inquiries into Truth and Interpretation, Oxford, Basil Blackwell, 1985.

speakers as we would expect if Davidson's analysis were correct.

The empirical evidence against Davidson's proposal is not conclusive. But it is good enough evidence to warrant a continuation of our search for a satisfactory account of our problematic pieces of language.¹⁵

7 Higginbotham's Account

What we were looking for was some way of making the denotation or extension of some part of the attitude attribution depend upon the senses of the embedded expressions, without having to deny that those expressions denote their normal denotations. We found that the suggested simple solution of just having the 'that' denoting the senses was not entirely satisfactory. But another way of doing it is suggested by Higginbotham (op. cit.).

The proposal is that we assign values to expressions only relative to the larger expressions of which they form a part. So, in general, if we have an expression *e* inside a larger expression, say a sentence, *S*, instead of assigning to *e* a value all by itself, we assign to *e* a value relative to *S*. The point of this is that since an expression may simultaneously appear in several different embedding expressions, we

15. There are also objections to Davidson that concern the logical properties of 'says that' sentences. These sentences appear to have implications that they would not, if their logical form were paratactic. For a recent example see Burge, T., 'On Davidson's "Saying That"', in LePore ed. op. cit.

may assign to it a value relative to one that differs from the value, if any, that we assign to it relative to another.

This idea of assigning values only relative to embedding expressions is useful not just for attitude attributions, but can solve problems with sentences of a completely different kind.¹⁶ This is important. If it were not so, then we would be introducing sweeping changes in the semantic theory, changes that affected what we said about every sentence, just to cope with the special problems with attitude attributions. If that happened, one might suspect that the cure was more radical than the disease required.

The success of the proposal may ultimately depend upon just this. If the trick of making assignments of values relative to embedding expressions proves fruitful in a range of different cases, that will vindicate its use here. If on the other hand it idles, and simply makes other areas needlessly cumbersome, that may cast doubt upon it.

Let us see how the idea applies to belief attributions. Consider (1) 'Ralph believes that Tanya is a terrorist', carved up as in (26):

(26) [_{S₁}Ralph believes [_{S'}that [_{S₀}Tanya is a
terrorist]]]

(26) is articulated into the three phrases, S₀, embedded in S', which is in turn embedded in S₁.

16. Specifically, certain sentence connectives, such as 'if' and 'unless', express different truth functions in different linguistic environments. See Higginbotham op. cit.

In order to cope with cross-denotation we want expressions inside content sentences to have their normal denotations. So we treat the parts of the content sentence S_0 , 'Tanya is a terrorist' relative to S_0 itself as if it were unembedded, and appeared in isolation. This means that the expressions inside the content sentence just get their normal denotations as desired. This is expressed in (27). (I have significantly altered Higginbotham's notation, for ease of assimilation).

- (27) (i) $\langle \text{'Tanya'}, S_0 \rangle$ denotes Tanya
 (ii) $(\forall a)(a \text{ satisfies } \langle \text{'x is a terrorist'}, S_0 \rangle \text{ iff } a \text{ is a terrorist})$
 from which it is easy to derive
 (iii) $\langle S_0, S_0 \rangle$ is true iff Tanya is a terrorist

The angle brackets are to be taken literally, so that we are technically assigning values to ordered pairs of expressions. But intuitively we can just understand $\langle \text{'Tanya'}, S_0 \rangle$ to mean, "'Tanya" considered relative to S_0 '.

Now we look at S' , 'that Tanya is a terrorist' and consider its role relative to the large sentence S_1 , 'Ralph believes that Tanya is a terrorist'. S' , considered relative to S_1 , appears as the object of the verb 'believes'. Higginbotham (following Harman¹⁷) takes the objects of propositional attitudes to be or to be objects "similar to" interpreted logical forms. Since, for Higginbotham, the objects to which semantic

 17. Harman, G., 'Deep Structure as Logical form', in Harman and Davidson eds., op. cit.

values get assigned are themselves logical forms, the 'that' clause is taken to be true of things that are similar to itself. Similarity is used in conscious imitation of Davidson's samesaying relation: two things will be similar to each other if they have the same or similar semantic features. The role of the content sentence may be expressed by (27)(iv):

(27)(iv) $(\forall a)(a \text{ satisfies } \langle S', S_1 \rangle \text{ iff } a \text{ is similar to } S_0)$

(iv) together with (v):

(v) $\langle \text{'Ralph'}, S_1 \rangle$ denotes Ralph

and

(vi) $(\forall a)(\forall b)(\langle a, b \rangle \text{ sats } \langle \text{'x believes y'}, S_1 \rangle \text{ iff } a \text{ believes } b)$

along with simple interpretation rules give us (vii)

(vii) $\langle S_1, S_1 \rangle$ is true iff $(\exists a)(a \text{ is similar to } S_0 \text{ and Ralph believes } a)$

(vii) is a T-sentence that tells us, more or less, that S_1 , the sentence 'Ralph believes that Tanya is a terrorist' is true iff Ralph believes something similar to the sentence 'Tanya is a terrorist'.

Higginbotham's treatment has embedded expressions denoting their normal denotations. This provides a simple solution to the problem of anaphora that we have been looking at.

Let us return to our example (10), 'Ralph believes that Tanya_i is a terrorist, but she_i is too honest', now constructed as (28):

(28) [S_1 Ralph believes that [S_0 Tanya_i is a terrorist]]
 but [S_2 she_i is too honest]

The challenge was to find some way of getting the unembedded 'she_i' in S_2 to denote Tanya, so that S_2 can get its correct truth conditions. Thus all we need to do is to find some pair $\langle x, y \rangle$ such that the denotation of x considered relative to y , is Tanya; as expressed in (29):

(29) $\text{Den}\langle x, y \rangle = \text{Den}\langle \text{'she}_i\text{'}, S_2 \rangle = \text{Tanya}$

But this is now no problem. 'Tanya' considered relative to the content sentence, will just get its normal denotation. So we have (30):

(30)(i) $\text{Den}\langle \text{'Tanya}_i\text{'}, S_0 \rangle = \text{Tanya}$
 (ii) $\text{Den}\langle \text{'she}_i\text{'}, S_2 \rangle = \text{Den}\langle \text{'Tanya}_i\text{'}, S_0 \rangle$
 So (iii) $\text{Den}\langle \text{'she}_i\text{'}, S_2 \rangle = \text{Tanya}$

The aspect of the theory that provides embedded expressions with their normal denotations seems to be satisfactory. However the other aspect of the theory, the one that is supposed to account for the failures of substitutivity, is less so. How are these failures to be explained?

7.1 Criticism of Higginbotham's Account

'Ralph believes that Tanya is a terrorist' is true, but 'Ralph believes that Patty is a terrorist' is false. Higginbotham's account allows for this easily enough. The content sentence, considered relative to its embedding attitude report, is treated as a predicate true of things

similar to itself. The content sentences 'Tanya is a terrorist' and 'Patty is a terrorist' are not similar to each other and obviously might fail to be similar to the same things. Since each is true of things similar to itself, they will then be true of different things. One may be similar to, and so true of, Ralph's belief content, while the other is not.

It seems to me that the explanation of the failures of substitutivity is not quite satisfactory. The two content sentences 'Tanya is a terrorist' and 'Patty is a terrorist' are treated, relative to their embedding attitude report, as predicates. They are predicates with different extensions: that is what accounted for the opacity effects. But the predicates are composed of expressions that have, on Higginbotham's theory, the same denotations. For the counterpart expressions inside the two content sentences will always get the same values. Relative to the content sentences, they are assigned their normal values, so codenotational expressions continue to share their denotation. And relative to the belief report as a whole they are not assigned values at all: only the whole content sentence gets a value relative to the larger sentence. Thus the extension of the content sentences considered as predicates does not depend upon the values of their parts.

Now, the departure from compositionality taken by itself, need not be a failing. There is no a priori reason to believe that semantics for a real natural language will everywhere be compositional.¹⁸

18. C. f. Hintikka, J., 'A Note on Anaphoric Pronouns and Information Processing by Humans', Linguistic Inquiry, 1987, p. 116, and the reference there cited.

Compositionality is one kind of explanation of how it is that we may understand a new sentence that is built up, in known ways, from known parts. There may be other ways in which to explain our knowledge of how the extension of new complex predicates is built up from old simple parts. But some such account needs to be offered for the particular case of content sentences, as treated by Higginbotham, if the violation of compositionality is not to matter.

At this juncture it will be well to reflect upon why the failure of compositionality does not afflict Davidson.

Davidson's samesaying relation is not invoked within the theory of meaning to explain the extension of any expression. For him every element in the whole attitude attribution has its extension determined by some antecedently available part of the theory of meaning. The content sentence is treated only as a sentence in its own right. It is not a predicate true of other things that samesay with it. Consequently compositionality is not violated.

Moreover Davidson has an acceptable explanation of how samesaying is supposed to work.¹⁹ Davidson uses samesaying only as a relation among utterances and inscriptions. If I say that Galileo said that the Earth moves, then samesaying relates my utterance of the content

19. The best source is his 1976 paper, 'Reply to Foster', reprinted in Inquiries into Truth and Interpretation, Oxford University Press, 1984.

sentence to a similar sort of thing.²⁰ To arrive at something that samesays with Galileo's utterance one proceeds as follows. Take Galileo's utterance. Construe it as an articulated linguistic structure of Italian. Apply to it a theory of meaning (for Italian, in English). This theory will deliver English sentences that have the same truth conditions as the Italian one. Now, there will be various different ways of stating the conditions under which Galileo's utterance is true, differences among which will be too subtle for the theory of meaning to note. But among these options will be some that are better suited to capturing the flavour of Galileo's utterance; more 17th century Italy, less 20th century America. Those suitably flavoured ways of stating the truth conditions of Galileo's utterance samesay with it.²¹

Higginbotham thus departs from Davidson in two ways. First, he writes samesaying into the theory of meaning itself, as a relation that gives the extension of certain complex predicates. And second, he treats samesaying as a relation among abstracta, not Fregean thoughts exactly, but "interpreted logical forms".

The second departure creates two lacunae that need filling. First, we know what it is to utter an utterance, but we lack an account of what

20. It is thus no accident that Davidson has never offered an account of the logical form of 'believes' sentences. Samesaying needs relata, and Davidson refuses to recognize any analogue of an utterance, in the case of belief. For an explicit avowal of this refusal see his 'Knowing One's Own Mind', Presidential Address delivered before the Sixtieth Annual Pacific Division Meeting of the American Philosophical Association, 1986.

21. I am indebted to Bruce Vermazen for my understanding of samesaying.

it is to believe a logical form. So when similarity is used to account for the extension of a content sentence, we lack, even in principle, a way of determining which specific objects it relates. One logical form is provided by the content sentence itself, as the theory treats it, but where is the other one? And, second, we know what it is for two utterances to samesay, but we lack an account of what it is for two logical forms to samesay.

The first departure from Davidson means that until the lacunae are filled, the semantics is radically incomplete. For 'samesaying' appears as a technical expression in the theory of meaning. It is an expression that is crucial to the account of how the semantic features of complex expressions depends upon their structure and the semantic features of their parts. The theorist of meaning thus cannot leave it unexplicated.

I do not claim that the required fillers of the lacunae are not be found. But I recommend a different course.

8 Frege Vindicated?

Let us collapse the number of logical forms in play down to one. Instead of trying to relate, as separate things, the logical form that Galileo believes, to the logical form of the content sentence 'The Earth moves', let us suppose that Galileo believes that very logical form.²² We

22. This was Harman's original idea, as I understand it.

may thus take a second step back towards Frege (the first, taken by Higginbotham, was mentioned four paragraphs back) and treat the content sentence (still relative to the embedding report, of course) as a singular term, rather than a predicate.

And, since we are in the mood for a retreat towards Frege, we might as well try a third step. Let us assign to each semantically simple expression in the content sentence a sense as its denotation. Embedded complex expressions, including the whole content sentences, will then denote logical forms (their own) in which inhere senses. The extension of a content sentence, considered as a singular term, will thus be a function of the denotations of its component expressions (senses), and its structure: compositionality regained.

We are still left with a problem similar to that of saying what it is to believe an interpreted logical form. We now have to say what it is to believe a thought, a logical structure of senses.

But we need not hurry to answer the question, since it no longer plays an important role in the semantics. There is just a relation that is denoted by the expression 'believes', a primitive and unanalyzed part of the content of the familiar idiom of belief reporting. Our problem was to explain how the semantic features of complex expressions depended upon those of their parts. We never promised to give philosophical theories of every primitive semantic value that we came across.²³

23. These remarks echo Davidson's defence of his use of the samesaying relation. See 'On Saying That' p. 104, of Inquiries into Truth and

So my suggestion is that instead of the unhelpful (27)(iv) we write (31):

- (31) (i) <'Tanya', S_1 > denotes \ulcorner Tanya \urcorner
(ii) <'is a terrorist', S_1 > denotes \ulcorner is a terrorist \urcorner

I assume that S' just inherits the value of S_0 considered relative to S_1 . This gives us (iii):

- (iii) < S' , S_1 > denotes \ulcorner Tanya is a terrorist \urcorner

the corner quotes once again signalling that the expressions inside them denote senses, rather than their normal denotations. (31)(iii) together with the original (27)(v) and (vi) suffice to yield (29)

- (32) < S_1 , S_1 > is true iff Ralph believes \ulcorner Tanya is a terrorist \urcorner

That is to say 'Ralph believes that Tanya is a terrorist' is true iff Ralph believes the thought that Tanya is a terrorist. So the opacity effects are explained in Frege's original way, by each embedded expression denoting a sense. But since we also have those very same expressions denoting their normal denotations, as expressed in (27)(i) and (ii), we can easily account for the appearance that embedded expressions denote their normal denotations. They appear to because that is just what they do.

Of course it would be rash to put a great deal of faith in the sketch

Interpretation, and the references cited thereat.

of the account I have just offered. We need to see how it extends to more complicated cases, how well or badly it captures desirable inferences, and so on and so forth. And we need to spell out the mechanics of the proof of the T-sentences: a task that quickly leads into difficulties with Frege's hierarchy. And perhaps most importantly, we had better be prepared to investigate whether anything could play the theoretical role that we have assigned to senses, and if so, what sort of thing this would be.

I confess to a moderate degree of optimism about each problem. Optimism about all the problems, however, would be supererogatory.

PAPER TWO

WHO'S AFRAID OF NARROW MINDS?

The mind is its own place
- John Milton

Who's Afraid of Narrow Minds?

1 Introduction

The purpose of this paper is to defend the view that all of the propositional attitudes that we ascribe to one another in our everyday folk psychology are possessed of what has come to be called 'narrow content'. The defence will by no means be complete, since I shall neither be attempting to meet all varieties of objection to the view, nor offering positive arguments in its favour. Rather I shall be concerned to fend off one major source of opposition to it. This opposition comes from Gareth Evans¹, and John McDowell.²

In this section I shall give a brief general introduction to the dispute. In the remaining sections I shall deal with Evans' and McDowell's objections to my view.

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

2. McDowell, J., 'The Sense and Reference of a Proper Name' ('SRPN') in Platts, M., ed. Reference, Truth and Reality, Routledge and Kegan Paul, 1980; 'De Re Senses' ('DRS'), in Synthese, 1986; 'Criteria, Defeasibility and Knowledge' ('CDK'), Proceedings of the British Academy, 68, 1982. Evans and McDowell appear to hold the same view, down to a detailed level. The leading idea originates with McDowell.

1.1 Background

In the last fifteen years or so a certain thesis about reference, largely owed to Kripke³ and Putnam,⁴ has become popular.

The view, which I call the 'direct reference' view, is that certain expressions are "extension-dependent".⁵ An expression is extension-dependent if its meaning depends upon its extension. So if 'Sandy' and 'Mandy' are extension-dependent, and refer to numerically different (even if type-identical) individuals, they have different meanings.⁶ And if 'Sandy' had referred to Mandy, it would have meant something other than what it actually does mean.

A similar thesis is applied to natural kind terms. But here the meaning of the expression depends not upon which actual things happen to fall under it, but upon the type to which those things belong.

There are two attitudes that the proponent of the direct reference view might take towards empty extension-dependent expressions.

3. Kripke, S., Naming and Necessity, Harvard University Press, 1979.

4. See e.g., Putnam, H., 'The Meaning of "Meaning"' in his Mind Language and Reality, Cambridge University Press, 1975.

5. I think the term 'extension-dependent' is owed to Wiggins.

6. Different extensions are sufficient for different meanings, but may not be necessary. It is open to the direct reference theorist to hold that two extension-dependent expressions with the same extensions may yet have different meanings. McDowell and Evans do hold this, see below section 4).

'Mandy' when empty certainly does not have the same extension as 'Mandy' when it refers to a genuine Mandy. This entails, on the direct reference view, at least that empty extension-dependent expressions do not have the same meaning as do non-empty ones. One might hold that an empty 'Mandy' has a different meaning from the name of a Mandy, properly so-called. McDowell and Evans, however, hold the stronger thesis that empty extension-dependent expressions have no meaning.

1.2 Conceptions of Content

It seems natural to take the direct reference view to apply to concepts as well as to expressions of language. Doing so, we arrive at the idea that which content a given concept has will depend upon what it refers to.⁷

The extension of the direct reference view from words to concepts is encouraged by the tight connections that we find between the contents of propositional attitudes and the sentences that we use to express and report them. If 'Sandy' and 'Mandy' have different meanings, then we would expect the beliefs they may be used to express, or report, to have different contents. (Indeed, on some views an expression's

7. Terminological matters. A concept is a thought component. A thought is a psychological correlate of an interpreted sentence: it is a syntactically structured object and it has a content (or some contents) which may or may not be essential to it. Thoughts are the objects of propositional attitudes; they may be believed, doubted, desired-true etc. 'Content' has to be understood more or less intuitively. It covers at least semantic content and phenomenal content, but claims that there are other sorts of content are to be regarded as at least intelligible.

meaning is to be identified with its role in expressing beliefs; in which case the extension of the direct reference view from language to thought would be automatic.) So the belief that Sandy is sullen would have a different content from the belief that Mandy is sullen. And if 'Sandy' has no referent, and so no meaning, then 'Sandy is sullen' cannot be used to express, or report, a belief. And this suggests that the belief that Sandy is sullen depends for its existence on its object, Sandy.

We might hypothesize that the mental states have two kinds of characteristics. One of these kinds of characteristic would be determined, at least in part, by things outside the head. The second kind of characteristic would be logically independent of anything outside the subject. Properties of this second kind would be insensitive to variations of the subject's surrounding environment: to them it would not matter whether there were a world outside the subject, nor, if there were, what it was like.

Characteristics of the first kind, if there are any, endow psychological states with what has come to be called 'wide content'; those of the second kind, if there are any, provide narrow content. I call McDowell's and Evans' view the 'Wide Content Only' ('WCO') view. On the WCO view some psychological states have only wide contents. If the belief that Mandy is morose is such a state, then that belief has no narrow content. The only content it has is of the kind that requires the existence of its object, Mandy. No Mandy, no wide content; no wide content, no content at all.

For McDowell and Evans it is thus no accident that our specifications of psychological states are bound up as they are with the objects of those states. When the best description of a psychological state makes mention of the state's object, this is because what it is for a subject to be in the state essentially involves her being related to that object. So if one is to believe that Mandy is morose, one has to stand in a certain relation to Mandy herself.

The view that I shall defend, the "Narrow Content Also" (NCA) view, is that all propositional attitudes have a narrow content. So the belief that Mandy is morose has a content that is not Mandy-dependent. The belief that Mandy is morose perhaps has a wide content as well as a narrow one. Removal of Mandy would eliminate that. But one would be left with a belief endowed with a definite content nevertheless.

The existence of narrow content would be the existence of something that simultaneously performs two separate tasks. The first accounts for its narrowness. If Zippy thinks Zonker is zany, then the narrow content of his mental state must be that much of it that is within Zippy.

Suppose that Zippy's concept, Zonker,⁸ refers to Zonker. This would be the upshot of a combination of factors. Some of these will lie outside Zippy: for example, certain relations crossing the gap in space and time

8. My use of the underlining device is to be understood as an attempt to do something that is as much like quoting concepts and thoughts as is possible.

between thinker and object of thought, that make it the case that when Zippy thinks some thought of the form Zonker is F, his thought will be true iff Zonker, in particular, is F. But these external relations cannot by themselves be all that is required to let Zippy think of Zonker. Presumably there are also important contributions made at Zippy's end of the chain. It is a task of the concept's narrow content to make such contributions. The concept's narrow content is thus determined by those of its properties that are entirely independent of any links of the chain that are outside Zippy.

The second task for narrow content is more nebulous in nature. Narrow content must deserve its name, and be recognizably **content**. It is not obvious what performance of this task entails. But for me narrow content is, at least, what gives a subject's perspective, or subjective point of view. The narrow content of a belief is the way in which the world is presented by the belief to its subject. Sandy's belief that Mandy is trendy makes the world appear a certain way from Sandy's perspective; which way that is depends only upon the belief's narrow content. Narrow content alone must suffice to provide something that should be counted as a subjective point of view.

The espousal of a theory that makes room for narrow content is partly motivated, at least in my case, by the intuition that the mind is, to put it suggestively, essentially inner. A subject's point of view of the world, I believe, is properly thought of as intrinsically independent of the world of which it is a point of view. It should depend only upon how things are with the subject, considered in abstraction from her

environment. For the friend of narrow content, then, specifications of points of view that make mention of external objects, must be specifications exploiting merely accidental features of what they specify.

The WCO theorist is skeptical that the two tasks set for narrow content are always performed, or even performable, by a single thing. The WCO theorist holds that the concepts that are ascribed by extension-dependent terms do not have narrow content. For what determines the reference of such concepts is mostly outside the head and independent of the mind.⁹

On the WCO view, what there is in the head that corresponds to these concepts does not, considered in isolation from its surroundings, have anything that should be called 'content'. It cannot by itself, in the absence of its relations to an external referent, count as a component of a point of view.

Although I think that the main underlying motivation for the WCO view is the idea that the internal component of a putative extension-dependent concept is too exiguous a thing to be endowed with

9. C. f. Burge, ('Belief De Re', The Journal of Philosophy LXXIV, 1977), who writes 'features of the mental entity itself ... are not always sufficient to pick out the relevant object' in these cases '[f]actors which determine the objects about which the believer holds his beliefs ... are appropriately counted nonconceptual - indeed, in most cases, noncognitive.' On his view what determines reference is independent of the mind in the strong sense of not being part of its cognitive domain. McDowell (DRS, see also CDK) takes issue with this. For him the mind's cognitive reach extends to include those factors that are literally (but not metaphorically) outside the head. For him, the independence of these factors thus consists only in the fact that they would exist, even if the mind within which they were located, did not.

content, I shall not have a great deal to say about the idea explicitly. The worry is rarely articulated very fully, and so it is hard to know what, exactly, would be required to alleviate it. I shall take up one statement of the concern in section 6, below, and argue that it is misconceived. I shall enter some intuitive considerations that indicate that what is ensconced entirely within the head is really content. Any more than that would require a theory of narrow content. Once such a theory was provided then it might be possible more seriously to take up the question of whether or not what is narrow is also content.

I think that rather than launching into the development and defence of such a theory, it will prove more fruitful to fend off a priori (and a posteriori, for which see my 'Seeing What is not There') objections to the plausible and natural view that there is something there to have a theory of.

With that end in mind, I shall be focussing on a particular expression of the WCO view, that McDowell gives (in SRPN). We are to suppose that a researcher is interested in the ways of a jungle tribe. He learns that the locals use the word 'Mumbo-Jumbo' as the name of one of their Gods. Not being of similar faith, he concludes, rightly we shall suppose that the name has no bearer.

If a native says something that intuitively has the English translation 'Mumbo-Jumbo brings thunder', then, McDowell holds, the interpreter should not use attributions of the belief that Mumbo-Jumbo brings thunder to help explain the native's actions - including those actions that are the utterings of sentences containing 'Mumbo-Jumbo'. For,

according to McDowell (ibid. p. 153) 'no belief is expressed by his words'. There just is no such belief as the belief that Mumbo-Jumbo brings thunder. And, for the most part, the same is view is taken of other empty proper names: they cannot be used in content sentences that express or ascribe beliefs.¹⁰ So the existence of thoughts ascribed by content sentences in which proper names feature, depends upon the existence of a bearer for the name: no bearer, no thought.

I shall be dealing with what I take to be four distinct arguments that are offered in support of the WCO view.¹¹ These arguments are not explicitly distinguished, but appear all run together in a dense couple . of pages of SRPN, and the last, and most interesting, is supplemented in detail elsewhere. Evans' contributions are gathered from different places in his book. Interpretation is therefore difficult, and it is possible that they would not wish to be associated with all of the arguments I attribute to them. We shall consider the arguments in ascending order of depth and plausibility.

10. There may be exceptions, but these have special features that set them apart from typical proper names. See section 4) below.

11. Two additional arguments are to be found in Evans' book. I have not dealt with these because I have relatively little to add to Blackburn's response to them; for which see Blackburn, S., Spreading the Word, chapter 9. I do not endorse Blackburn's account of narrow content.

2 First Argument

The native says something that intuitively has the translation 'Mumbo-Jumbo brings thunder'.

Here is what appears to be a brief, almost self-contained argument (McDowell *ibid.*):

No belief is expressed by his words: they purport to express a belief which could be described in the transparent style, but since no appropriate belief could be thus described, there is no such belief as the belief they purport to express.

A belief describable in the transparent style is a belief concerning some object, that it satisfies some property. On McDowell's view it is only such beliefs that are attributed by content sentences with proper names.

I am not sure that McDowell really intends the quoted words to express an argument. But, in case he does, I respond.

2.1 Reply to First Argument

The argument would seem to proceed as follows. First premise: the words purport to express a belief that could be described in the transparent style. Second premise: no appropriate belief could be so described. Therefore, subconclusion: there is no such belief as the belief the words purport to express. Therefore, conclusion: there is no

belief that they express.

I have two objections. The first is to the first premise. What does it mean to say that the words 'purport to express a belief which could be described in the transparent style'? I think the most that can reasonably be purportedly purported is that there is a belief, the belief is expressed by the words, and the belief is describable in the transparent style.

McDowell, however, thinks otherwise:

When one sincerely and assertively utters a sentence containing a proper name ... one does not mean to be expressing a belief whose availability to be expressed is indifferent to the existence or non-existence of a bearer for the name.

Were this true, then the words of the native's utterance would purport to express a belief that not only was describable in the transparent style, but that was so essentially. But McDowell has no right to claim this.

Surely enough, if one utters a sentence containing a proper name, one is, in the normal case anyway, committing oneself to the existence of its bearer. But that is explained by the truth of one's utterance being dependent on this ontological fact. There is no warrant for the claim that it is written into lay beliefs, or somehow implicit in lay practice that one takes not just the truth, but the very existence, of one's singular beliefs to depend upon the existence of referents for words used to express them. Those less concerned with philosophical theories than we are, including perhaps McDowell's native, are not up

to their necks, knowingly or otherwise, in the WCO view.¹²

The second premise, that no appropriate belief could be described in the transparent style, is true.

The subconclusion, that there is no such belief as the belief the words purport to express would follow, perhaps, if the first premise were interpreted in the strong form that McDowell seems to intend. Then, perhaps, we could think of the native as committed to the principle: either I express a transparently describable belief, or I express no belief. If, on the other hand, we interpret it as I suggested, thus allowing it to be true, then what follows is merely: there is no belief that has the property that the purportedly expressed belief is purported to have. It does not follow that there is no belief purportedly expressed by the words.

My second objection is to the move from the subconclusion to the conclusion. What McDowell really needs to show is that no belief is actually expressed by the words. But this in no way follows from what he has so far argued. It is not in general true that what is expressed must also be purportedly expressed. Why should it be true in this case?

12. Since McDowell's claim is undefended and prima facie implausible, I do not undertake a defence of my opposed opinion. Were such to be offered, it would be built around the frequent and unabashed uses of empty proper names in descriptions of others' beliefs, that we indulge in. Normal people do not have the sense that it is literally impossible to believe that Santa Claus has a white beard, whether or not he exists.

Even if McDowell did intend the rather feeble argument we have been discussing, he certainly would not want his WCO view to rest on it. So let us move on.

3 Second Argument

Here is an official defence of the WCO view (ibid.):

The non-Fregean [=WCO] view¹³ can be defended on these lines. An interpreter's ascription of propositional attitudes to his subject is in general constrained by the facts (as the interpreter sees them). This is partly because intelligibility, in ascriptions of belief at least, requires conformity to reasonable principles about how beliefs can be acquired under the impact of the environment; and partly because the point of ascribing propositional attitudes is to bring out the reasonableness, from a strategic standpoint constituted by possession of the attitudes, of the subject's dealings with the environment. Now, whether a name has a bearer or not (in an interpreter's view) makes a difference to the way in which the interpreter can use beliefs he can ascribe to the subject in making sense of the subject's behaviour. A sincere assertive utterance containing a name with a bearer can be understood as expressing a belief correctly describable as belief, concerning the bearer that it satisfies some specified condition. If the name has no bearer (in the interpreter's view), he cannot describe any suitably related belief in that transparent style.

What follows the above quotation is a comment about what, perhaps, an interpreter could glean from the utterance of a sentence with an empty proper name, given that it has no sense. Then comes the brief

13. When McDowell calls the view 'Non-Fregean' he has in mind Frege's famous claim in 'On Sense and Reference' that the thought that Odysseus was set ashore at Ithaca while sound asleep 'remains the same whether "Odysseus" has reference or not'.

argument I considered in the last subsection. Neither of these sequentia adds to the argument begun by the words in the long quote. Can we, then, reconstruct, from those words, an argument for the non-Fregean view? Perhaps the following.

The argument begins with a statement of the principle of charity. The ascription of attitudes is constrained by what the external facts are taken to be, in general, so that we can ascribe beliefs and behaviours that make sense in the light of those facts, to whatever extent is possible (possible, perhaps, relative to certain other constraints). When a name has a bearer, we can attribute beliefs expressed by utterances containing the name, and use these attributions to make sense of the actions in a special way. The special way involves taking the the belief to be describable¹⁴ transparently.

Perhaps McDowell would say that in these cases we can see the subject's truck with the environment as directly guided by that environment. We can interpret his actions by ascribing attitudes that involve special cognitive rapports with those parts of the environment with which he interacts. Such rapports are logically dependent, for their existence, upon their objects.

But then, the conclusion would be, these attitudes, which are necessarily describable in the transparent mode, must be distinct from

14. 'Describable', not 'described'. The point is that the belief is taken to be about a particular object. It may be so taken even when described by a content sentence in which intersubstitutions are not truth-preserving.

any attitudes not so describable. Such attitudes are simply not available in the absence of an object.

3.1 Reply to Second Argument

If that is the argument, then it is vulnerable to the following objection. There is no valid move from the premise that the existence of an object for the attitude is required for the availability of a certain sort of description of the attitude, to the conclusion that the object is required for the availability of the attitude itself. It might be true that the existence of an object allows us to describe an attitude by relating it to an object with which the subject is intimate, without it being true that the very same attitude could not be described in another way, a way that made no mention of the object. To show otherwise McDowell would have to argue that the attitude itself was dependent upon the object: and that is exactly the point of dispute.

The existence of descriptions of a subject's mind that make mention of particular objects provides no evidence for the existence of a special class of object-involving attitudes. This is true no matter how useful these descriptions might be, for the interpretation of the subject's actions. The question is whether what makes true a transparent description may correctly be seen as composed of an extension-independent attitude of the subject's and an external object to which it stands in some particular relation. Nothing in McDowell's argument suggests a negative answer to this question.

4 Third Argument

The third argument to be found in Evans and McDowell results from their conception of the proper form for a theory of sense.

Their idea, one which I applaud, is that a Davidsonian theory of meaning may serve as a theory of sense. Such a theory will have machinery to derive, for each sentence of the object language, a T-sentence of the familiar form 'S is true iff p': 'S' would be replaced by a structure-revealing description of an object-language sentence, and 'p' by a statement, in the meta-language, of that sentence's truth conditions. The machinery will include clauses that assign semantic values to simple expressions of the object language, and clauses that allow for the computation of the semantic values of complex expressions, on the basis of those of their parts. What the theory actually explicitly states, then, is what expressions denote. But the theory may yet serve as a theory of sense, if in saying what the denotation of an expression is, it does so in a way that displays, or reveals, the expression's sense.

McDowell and Evans hold that proper names should be dealt with by clauses of the form:

(1) 'a' stands for A

where 'A' has the same sense as 'a'. Such a clause, McDowell says (ibid. p. 143, c.f. also Evans, op. cit. p. 35) 'gives - or more

strictly, in that context, as good as gives - the sense of the name'. That context is a theory of meaning that delivers T-sentences capable of doing their bit in making sense of the linguistic actions of object-language speakers. The clause explicitly says only that 'a' stands for A. That is something that might equally be expressed by the clause:

(2) 'a' stands for B

where 'B' has the same referent as 'a', but a quite different sense. But only the first clause correctly shows the sense of 'a'. In this way a theory that includes the first clause may suffice, if understood correctly, as a theory of sense. It so suffices because of what it shows, not what it says.

On a theory of meaning that gives the senses of proper names in McDowell's suggested manner, there is no giving of the senses of proper names that lack a referent. For the honest interpreter cannot enter into her theory of meaning the clause:

(3) 'Mumbo-Jumbo' stands for Mumbo-Jumbo

nor any other clause of the suggested form, for an empty proper name. McDowell does not say why the interpreter cannot accept (3), but the reason is obvious. (3) entails: ($\exists y$)('Mumbo-Jumbo' stands for y), which the interpreter must take to be false.

If this is the right way to go about things, then empty proper names can have no senses; the moral that McDowell draws.

Evans (ibid. pps. 25-7) appears to make much the same argument. He points out that a clause of the form of (1)

identifies the semantic value of the name in a way which **shows**, or **displays**, its sense. When a theory of semantic value meets this condition quite generally we may say that it can **serve** as a theory of sense ... This conception of the relation between Fregean theories of sense and of semantic value is attractive ... but it is clear that if we suppose that expressions without semantic value may still have a sense, we cannot avail ourselves of this conception.

And, if a proper name has no sense, then nothing is expressed by an utterance that contains it. Hence McDowell's conclusion that no belief is expressed by the native's utterance 'Mumbo-Jumbo brings thunder'.

The form of the argument appears to be this. We begin with the premise that the best way to go about framing a theory of sense, is to construct a theory of denotation that may do duty for one. Performance of this duty consists in employment of clauses that state what needs to be stated for the theory of denotation, but that do so in a special way. The clauses must state what they state in a way that shows the senses of the expressions they deal with. For McDowell and Evans following this advice seems to require the incorporation into the theory of meaning of clauses of the form (1). Then they note that empty proper names cannot be honestly handled by such clauses. The conclusion is that these names have no senses: and this is, or leads directly to, the WCO view.

4.1 Reply to Third Argument

The argument takes a false step. The step from the constraints upon

the clauses of the theory of denotation, to the recommendation of clauses of form (1), is nowhere justified. It is true that such clauses provide an admirable way of implementing the general proposal, for the special case of proper names with referents. But McDowell and Evans nowhere argue (that I am aware of) that the clauses provide the only way of following the general program in the particular case. So it will be possible to short-circuit the argument, if we can find clauses of some form other than (1), that will do for the theory of sense, but that can be applied to empty proper names.

Evidently some clause like the following is our best bet:

(4) (x)(N stands for x iff ...x...)

But what is to fill in the blanks?

The question appears to be embarrassing. What fills the blanks must suffice to specify the referent of the name, if it has one, and at the same time must display the name's sense. But there is no reason to suppose that for every name, a description that does both of these things will be abroad. Appropriate descriptions, it seems, will only be forthcoming where there is some specific descriptive condition, associated with the name, and generally held, by those who understand the name, to fix its reference. (Some empty proper names may have that property, which is why I said earlier that McDowell and Evans might admit exceptions to their general thesis). If a name lacks this property, then, though we might easily enough find a description that specifies the name's referent, we will have no right to suppose that it gives the name's sense (c.f. Evans, *op. cit.* p. 48).

Let us not panic. If the name has a bearer, then an appropriate description can always be found. We just have to combine the standard disquotation trick with an '=' sign. So we could employ a clause like:

(5) $(x)(\text{'Mumbo-Jumbo' stands for } x \text{ iff } x=\text{Mumbo-Jumbo})$

If our metalanguage is regimented with a standard logic, then (5) will not do as it stands. The second occurrence of 'Mumbo-Jumbo' is in a quantifiable position, and will permit existential generalization.

But there are two manoeuvres available to remedy this, both of which I find independently advisable.

Intuitively it seems as though (5) is true rather than false or lacking in truth value. If that is right then the metalanguage that best captures intuition will be based on a free logic, a logic that does not permit existential generalization in (5). An appropriate and elegant regimentation of a metalanguage suitable for formulating a version of (5) in such a way that it comes out true, has been provided by Burge.¹⁵

McDowell, of course, would not countenance the intuition that (5) should be counted true. But that is not the issue. His claim was that a theory of meaning that employed axioms for proper names that showed the names' senses as it stated conditions upon their reference could not handle empty proper names. But (5), ensconced in the appropriate language, will serve just as well as (1) for non-empty proper names,

15. Burge, T., 'Truth and Singular Terms', somewhat ironically in Platts, ed., *ibid.*

and is not falsified by empty ones.

The second manoeuvre for getting out of ontological trouble would be to treat names as predicates. There are two ways of doing this. The first would be to take a leaf from Quine's book¹⁶ and collapse the semantically complex '= Mumbo-Jumbo' into the semantically simple predicate 'is Mumbo-Jumbo', with 'is' copular.

We would then avail ourselves of the (relatively) ontologically neutral clause:

(6) $(x)(\text{'Mumbo-Jumbo' stands for } x \text{ iff } x \text{ is Mumbo-Jumbo})$

Here the unquoted occurrence of 'Mumbo-Jumbo' is not a singular term, and its position is not (first-order) quantifiable. The predicate 'is Mumbo-Jumbo' would be uniquely true of Mumbo-Jumbo, if he existed, and otherwise would have empty extension.¹⁷

The import of (6) would be that the object-language expression 'Mumbo-Jumbo' would function as a descriptive phrase true of just Mumbo-Jumbo, or of nothing. So 'Mumbo-Jumbo brings thunder' would be elliptical for 'The Mumbo-Jumbo brings thunder'.

Employing the Quinean predicates in an account of natural language proper names is relatively plausible for those names which have a

16. The book is Quine, W. V. O., Word and Object, M. I. T. press, 1960, the leaf is 174.

17. Dummett makes just this response to Evans and McDowell, in Dummett, M., The Interpretation of Frege's Philosophy, Cambridge, Mass., Harvard University Press, 1981, p. 136.

unique, generally accepted purported referent in the community. 'Mumbo-Jumbo' would probably be an example, along with 'Pegasus' and some others. It does not seem to work very well for common or garden names like 'Ike', 'Mike' and 'Spike'. If a theorist of meaning were to tell us that

(7) (x)('Mike' stands for x iff x is Mike)

we would not learn very much. If there is one main Mumbo-Jumbo, then (6) does specify the conditions that determine the value of 'Mumbo-Jumbo'. If there are many Mikes, then (7) does no such thing for 'Mike': which thing does x have to be, if it is to be the value of the word?

4.2 A Modified Burgean Treatment of Proper Names

Most ordinary proper names are best treated as predicates that are capable of being true of many different individuals, rather than being true of just a unique individual or nothing. This treatment of names is due to Burge.¹⁸

Burge notes that names may without impropriety function as common nouns. One can say, for example, 'If there is one main Mumbo-Jumbo, then (6) does specify the conditions that determine the value of "Mumbo-Jumbo"', 'If there are many Mikes, then (7) does no such thing for "Mike"' and 'An empty "Mandy" has a different meaning from the

18. Burge, T., 'Reference and Proper Names', The Journal of Philosophy LXX, 1973.

name of a Mandy, properly so-called'.

The meaning of proper names is apparent in such sentences. To be a Mandy is to be someone (perhaps even something) who is properly called 'Mandy', whose name (or one of whose names) is 'Mandy'. What it is to be properly called 'Mandy' is a question that lies outside the theory of meaning proper. The theorist of meaning does not have to give an account of the conditions under which one is the genuine owner of a proper name. She is committed only to there being such an account to be given.¹⁹

Since there genuinely is such a thing as being a Mike, and being a Mike is necessary and sufficient for 'Mike' to be true of one, we may enter (8) into our theory of meaning:

(8) (x)(x satisfies 'Mike' iff x is a Mike)

(8), unlike (7), does specify the conditions that determine whether or not something is a semantic value of 'Mike'. Where (7) left us wondering which Mike was in question, (8) tells us that any old Mike will do, and so leaves nothing undetermined.

Burge (ibid p. 432) recommends that when a proper name functions in a sentence as a singular term, the sentence should be formalized as an open sentence: 'Aristotle is human' becomes

19. For a creditable theory of how names get hooked up to their bearers see Evans, G., 'The Causal Theory of Names', Aristotelian Society Supplementary Volume vol. XLVII, 1973, and chapter 11 of Evans op. cit.

(9) Is-Human($\{x_1\}$ Aristotle(x_1))

The free variable may be quantified from without, but, more standardly, will receive its value from a contextual act of demonstration. In saying 'Aristotle is human' one is referring to a particular Aristotle. So, on Burge's view, it is as if one said 'That Aristotle is human', indicating the Aristotle that one has in mind.

I would, however, prefer 'The contextually definite NN is F' to 'That NN is F' for capturing the flavour of uses of proper names as singular terms. For uses of names are very rarely accompanied by anything resembling an act of demonstration on the speaker's part. It is rather that a speaker will, in using a proper name, succeed in identifying a particular individual for his audience, if in the context of utterance there is a most salient bearer of that name. When the context is thus propitious the speaker has to do nothing more than utter the name in order successfully to refer. No demonstration is required. And if the context is not thus propitious, then the audience will not know to whom the speaker is trying to refer. (C. f. Evans *ibid.* p. 374, for some considerations in a similar but not identical vein).

Burge offers two arguments for his demonstrative-ridden formulation, neither of which are satisfactory.

His first argument is that 'Jim is 6 feet tall', like 'that book is green' are (*ibid.* p. 432):

incompletely interpreted - they lack truth value. The user of the sentences must pick out a particular (e. g. a particular Jim or book) if the sentences are to be judged

true or false. It is this conventional reliance on extra-sentential action or context to pick out a particular which signals the demonstrative element in both sentences.

But the behaviour of 'The contextually definite F' does not seem relevantly different from demonstratives in this respect. If I write 'The contextually definite dog is hungry' you are no more capable of assessing the truth value of my claim than if I write 'that book is green' or 'Jim is 6 feet tall'. If one is attributing to the object language a bivalent logic, then all three sentences come out false. If one is attributing to it truth value gaps, then all three will be seen to lack a truth value.

Burge's second argument is that proper names, like demonstratives, always seem to take the widest possible scope. But this also seems true of 'the contextually definite' (compare 'the actual' as an operator that ensures that a definite description be read rigidly). The modifier 'contextually' singles out the actual context, and so makes the definite description act similarly to a demonstrative.

So I would recommend that (8) be employed in the proof of T-sentences that tell us this sort of thing: 'Mike is F' uttered by s at t is true iff in s's context at t, there is exactly one contextually definite Mike, and he is F.

Our problem was to find way to construct clauses that allow us to show the sense of a name that has a bearer, by specifying its referent, but that also allow us, without undesirable ontological commitment, to convey the senses of bearerless names.

We have found three ways of meeting the requirements. We can either use clauses like (5) with a free logic or use Quinean predicates, as in (6); either of those two options would be suitable for names that have a primary purported referent. Or we could follow Burge (with suggested amendment) and employ clauses like (8), with appropriate supplementary mechanics for proving the right T-sentences.

The possibility of exploiting these methods does not suffice to show that empty proper names do have senses, where this is to mean that they can be used to express or report genuine thoughts. But the argument that we were considering is met. That argument depended on the assumption that only clauses of the form (1) could serve the duplex function of giving a name's referent and showing its sense. That assumption was rash.

5 Fourth Argument

5.1 First Case: Perception

For McDowell the countenancing of narrow content as a general feature of propositional attitudes carries a commitment to 'the idea that thought relates to objects with an essential indirectness'. There are various ways in which this indirectness might come up, for example a description theory of reference for names and demonstratives, or what resembles a 'sense data' account of perception. As he puts it in DRS (p. 293), crediting demonstratives with 'senses that determine objects in such a way that that the senses are expressible whether the objects exist or not' ... generates a falsification of demonstrative thought

akin to the falsification of perceptual experience that is induced by representative realism. Representative realism postulates items that are "before the mind" in experience whether objects are perceived or not, with the effect that when an object is perceived, it is conceived as "present to the mind" only by proxy.

And Evans (op. cit. p. 199) suggests that resistance to a WCO (he calls it 'Russellian') theory of demonstrative thoughts might arise from a picture of mind that

is already rejected when one acknowledges that if in perception anything is before the mind, it is the public objects themselves, not some internal representation of them. The thesis we have been considering [=the WCO thesis] is really no more than a corollary of this realism.

McDowell and Evans are claiming that the supporter of narrow content

is committed, willy nilly, to a faulty picture of reference, knowledge, and perception. This picture suffers the defects of the old 'sense data' theory of perception, and has, as its least endearing feature, a Berkeleyan idealism, or solipsistic skepticism, as a consequence.

How does the argument go? Realism (not "representative realism", a view that McDowell thinks cannot but collapse into skepticism) depends upon holding that when one perceives an object, what is 'before the mind' (as Evans puts it) really is the object. Suppose it were not so, but that what was before the mind was some mental representative of the object. Then we should have to deduce the existence and nature of an external object, from the representative itself. But no such deduction could be sound, since the representative could exist, exactly as it is, in the absence of any such object. And it becomes a mystery how we can even form a conception of any occupant of a reality that lies thus for ever beyond our ken.

What goes for perception goes for cognition generally. So, the fear is, the consequences of an NCA view are dire. If a sense or concept always comes between us and the external objects of our thoughts, and from these senses or concepts there is no sound deduction of the existence of the external objects, how can we ever know the object is there, or even understand what it would be for an object to be there? We will be, as Dummett puts it 'inescapably locked into a third realm inhabited by thoughts and their components ... unable to reach the

outer world of actual physical things'.²⁰

5.2 Reply to First Case

I am disturbed by the threatened conclusion. However, I cannot find anything that is essential to the NCA view, that leads into the mire. So I shall try to state the minimal requirements of the view, and argue against Evans and McDowell that the paths I need tread are hardly more perilous than those open only to the WCO theorist.

Embracing narrow content involves no more than the claim that to think (have an experience) of x involves having in one's mind some distinct thing, y , that both refers to (is of) x , and that has genuine content of a sort that is independent of x . Such a claim is indeed a common component of the dangerous views that Evans and McDowell, quite rightly, turn their backs on. But to say just that would provide the materials for only the most pitiful of inductive arguments. And that is not what McDowell and Evans intend.

Certainly there are ways of interpreting the relations that hold among the various elements included in the NCA picture, that will invite trouble. In particular if we thought of the narrow content of a concept as the object of the concept, then we would have to explain how we could perceive, or even conceive of, something lying beyond the narrow content, out there in the real world. Evans and McDowell frequently

20. Dummett, *ibid.* As I implied, Dummett does not accept the McDowell/Evans argument.

saddle the NCA view with this error. The misconstrual of the NCA position is apparent in the quotations provided above, and quite explicit in other places. McDowell for example (CDK) considers an argument that begins with the premise that a subject may not be able to tell whether his experience is deceptive or not, and proceeds towards the conclusion that experience provides 'at best a defeasible ground' for knowledge. We shall be looking at a similar argument shortly, but now I want to draw attention to how McDowell elaborates on it.

He reformulates the argument:

In a deceptive case, what is embraced within the scope of experience is an appearance that such and such is the case falling short of the fact: a mere appearance. So what is experienced in a non-deceptive case is a mere appearance too.

Then, in the next paragraph he refers back to a part of the argument with the words 'As before the object of the experience in the deceptive case is a mere appearance.' And he goes on, understandably, to claim that in the non-deceptive case the object of experience is not a mere appearance, but some part of the real outside world.

Surely enough some famous works have contained an argument that began with the premise that one sometimes cannot distinguish deceptive experiences from non-deceptive ones, and ended up with the conclusion that what is experienced, the object of experience, is the same in the two cases. The next natural stopping place is some sort of idealism or scepticism. But that is an importantly different argument from the one that McDowell appeared to start out with, and ought to have been considering.

The latter argument moves simply from the stated premise to the conclusion that the experiences, considered in and of themselves, are the same in the two cases. Such a claim does not entail that the objects of experience are the same. And the sensible move at this stage is simply to deny that they are. In the sense in which what I experience, the object of my experience, is, in the non-deceptive case, some real thing in the world, there is nothing that is experienced, no object of experience, in the deceptive case. Yet the experiences themselves may be the same in their non-relational characteristics; and that is all that is required by the NCA theorist.

Availing myself of this essential distinction between the non-relational character of an experience, and the object of which it is an experience, clears me of the crude version of McDowell's and Evans' objection. But underlying the crude form is a more subtle and interesting line of thought, that proceeds without the needless slide.

To keep my feet on firm ground, I must show that cognitive relations to the physical world can constitute knowledge, even if they consist in having in one's mind something that could exist, even if the rest of the world did not. So the following sort of story must be intelligible. I know that there is a keyboard in front of me, because I see it there. My seeing it involves my having a visual presentation that could be just as it is even in the absence of the keyboard (I could be hallucinating). Yet, nevertheless, I see the keyboard and I know that it is there.

To McDowell stories like the little one told by my last three sentences seem highly problematic. Objecting to a 'criterial' view, he writes

(ibid.) that it

envisage[s] ascribing knowledge on the strength of something compatible with the falsity of what is supposedly known. And it is a serious question whether we can understand how it can be knowledge that is properly so ascribed.

The question is serious for the following reason. If I admit that the basis for my claim to knowledge that p is the experience, and I also claim that my experience is compatible with not- p , then I appear to be committed to a pair of claims that sit uncomfortably on the same knee. My experience legitimates my claim to know that p . But, my experience being what it is, I must also concede that for all I know, not- p . The discomfort may be heightened by the observation that if for all I know, not- p , then it follows that none among my items of knowledge rules out not- p . And that entails, by elementary logic, that the knowledge that p is not one of those items.

I require it to be possible for an experience to be both legitimate ground for a claim to knowledge that p , and such that it is compatible with the failure of p to obtain. I must therefore deny that the compatibility of the experience with not- p forces the concession that for all I know, not- p . And this seems right independently. For that denial is tantamount to the quite harmless assertion that one may be in a position to claim to know that p on the basis of evidence from which p does not follow logically.

McDowell (ibid.) does not see how it is possible to 'drive a wedge between accepting that everything that one has is compatible with things not being so, on the one hand, and admitting that one does not

know that things are so, on the other.' In particular, he holds, no appeal to conventions governing the use of the word 'know' can help us out here, since we would have to show that the conventions were 'well-founded'. But, McDowell says, the 'criterial' view and its fellows (among which, on this particular matter if few others, the NCA view is to be counted) concedes 'that the sceptic's complaints are substantially correct'. And that concession, he says, cannot leave room for any appeal to the well-foundedness of the conventions.

It is perhaps unsurprising that each side thinks that it is the other who concedes too much to the sceptic. All that I am conceding to the sceptic is that the experiences that are available to justify my claim to know that p may be expected to be compatible with not-p. I cannot see how this undermines the well-foundedness of the "conventions" governing the use of the word 'know'. Or, to make a stronger claim in the same vein: given what the word 'know' means, and hence, what knowledge is, the sceptic is right at least in saying that what grounds our claims to knowledge does not always establish it beyond all possible doubt. It is just a fact about knowledge, that one's possession of it may be based upon something that does not logically entail the truth of what is known. The contrary supposition must stem from a need to believe that what distinguishes a subject who knows that p, from one who does not, must lie in the characteristics of the subjects' experiences, considered in and of themselves, rather than in, say, some external thing to which the experience may be suitably related. But it is exactly that which the sceptic holds, and that we have no need to concede.

Consider two subjects whose experiences are the same, in both cases warranting a claim to knowledge. Suppose that in one case the experience reveals the truth, and in the other, it misleads. In one case the subject really does know what he claims to know, in the other, the claim is in fact false, but still justified, in the sense of being adequately grounded, so that the subject was quite right to make it. McDowell (ibid.) goes so far as ask.

How can a difference in respect of something considered as cognitively inaccessible to both subjects ... make it the case that one of them knows how things are in that inaccessible region while the other does not -- rather than leaving them both, strictly speaking, ignorant of the matter?

It is a fear that this sceptical question has no answer, that leads McDowell to hold that the experiences of the two subjects cannot, considered by themselves, be the same. I shall argue later that the question is one that McDowell also must face, in which case its fearsomeness or otherwise counts neither way. But I think the fear is groundless. If we take a closer look at the question we can see that it lacks potency.

The 'difference in respect of something cognitively inaccessible to both subjects' is simply a difference that does not effect the intrinsic nature of the subjects' experiences, considered in abstraction from the outside world. McDowell's formulation in terms of 'cognitively inaccessible' regions quite unnecessarily invites the sceptic's challenges. For if we take cognition to include knowledge, as we should, then of course even the proponent of narrow content can hold that the difference in the surroundings of the two subjects is 'cognitively

accessible'. No skeptical conclusions will then follow.

If we adopt a restrictive use of 'cognition', then the region in which hides the distinguishing feature of the two cases, is indeed cognitively inaccessible. But then we have no difficulty in answering McDowell's question. The difference in the 'cognitively inaccessible' region is the difference between the obtaining of what the subjects claim to know, and its not obtaining. The very boring fact that 'S knows that p' entails that p, thus ensures that one subject may know that p, while the other does not.

The proper way to respond to the sceptic is not to search for some feature of our experiences that guarantees their accuracy, but to make it plausible that even if "everything that one has is compatible with things not being so, on the one hand, one may yet know that things are so, on the other". I do not think that the task would be very hard. Let me offer the briefest of caricatures of how the tale would unfold.

'Everything that one has' means everything that one may legitimately draw upon in ascertaining how things are. To say that everything that one has is compatible with, say, its not raining is just to say that it is logically possible that everything that one has could be just as it is, and it not be raining. Suppose that everything that you have includes the visual, auditory, olfactory and tactile sensations typically associated with being out in the rain. Everything else is, as far as you can tell, just as it should be; there is no evidence that your senses are out of order, that you are being tricked, or that anything is disturbing the

status quo. You have, in short, overwhelming evidence that it is raining. But that is all you have.

If you have overwhelming evidence that it is raining, and this evidence is caused, in the usual manner, by its actually raining, then you know that it is raining. What more could possibly be required to legitimate your claim to knowledge? But still, the conditions that legitimate the claim to knowledge do not logically guarantee that it is raining. To suppose that what grounds a legitimate claim to knowledge must logically guarantee that what is claimed to be known actually obtains, is simply to confuse knowledge with a kind of certainty that is a philosopher's invention. One might expect a sceptic to make this confusion, but it is hardly prudent for the rest of us.

The position that McDowell attempts to adopt is supposed to leave adequate room for the required distinction between knowledge and certainty. It seems to me, however, that the manoeuvres required to make this room, lead McDowell into exactly the admissions he strives to avoid.

On his articulation of the WCO position the difference between our two subjects whose epistemic states differed only in what went on outside of their heads, must be written into their experiential states. For one subject, his experience would be held not to 'fall short of the facts': rather, his experience would just be the fact itself 'made perceptually manifest' to him. For the other, what is experienced would be 'mere appearance'. This poses a risk to our fallibility. If the experience of its raining were truly different when it was raining, from

when it was not raining, how could we fail to distinguish the two? And if we could always distinguish the two, how could we be fooled by the deceptive experiences?

McDowell's answer is obvious enough. The content of a subject's experience is not fully accessible to introspection: at least, one is not always in a position to tell which sort of experience one is having. One may not be able to tell whether one's experience is the fact that it is raining, within one's cognitive reach, or a mere appearance that this is so. Thus he concedes that 'there may be deceptive cases experientially indistinguishable from non-deceptive cases' (ibid.). This indistinguishability of course cannot reside in the experiences themselves, but only in the fact that a subject may not know which of the two he is enjoying. So it is not conceded that in the two cases there exist two experiences, which experiences are identical. It is only conceded that the mere appearance of its raining may seem the same (not be the same) as the reality of its raining, when that is experienced.

But even this meager concession is fatal. McDowell concedes that the subject cannot tell whether he has an accurate experience or a deceptive one. But if a subject cannot tell the difference between the two experiences, then that difference is not something he may exploit in ascertaining which of the two he has. So the difference that McDowell has written into the subjects' experiences cannot be epistemically relevant.

McDowell (ibid.) attempts to answer this criticism.

One could hardly countenance the idea of having a fact made manifest within the reach of one's experience, without supposing that that would make knowledge of the fact available to one.

The root idea [behind the criticism] is that one's epistemic standing on some question cannot intelligibly be constituted, even in part, by matters blankly external to how it is with one subjectively. For how could such matters be other than beyond one's ken?

When someone has a fact made manifest to him, the obtaining of the fact contributes to his epistemic standing on the question. But the obtaining of the fact is precisely not blankly external to his subjectivity

The criticism is not answered by a restatement of the contested conclusion (the first and penultimate sentences in the quoted sections), and a reformulation of the objector's point in a vocabulary that obscures it (the second paragraph). Like 'cognitively inaccessible', 'blankly external to how it is with one subjectively' is a slippery phrase. McDowell can certainly respond to its use by saying that if an experience is a fact within one's cognitive reach, then the fact is not blankly external to how it is with one subjectively. But the objection proceeds simply from the concession that a subject cannot distinguish between the two experiences. It does not matter whether the difference between the experiences is counted as being part of how it is with the subject subjectively, or not. If he cannot distinguish between the two things, then that distinction is not something that he has to go on when it comes to deciding how it is in the world around him.

The place at which we end up, then, after winding our way through McDowell's view and its consequences, is exactly the one where the NCA theorist is already at rest. We are left with the task of explaining how it can be that when "everything that one has is compatible with things

not being so, on the one hand, one may yet know that things are so, on the other". And that of course, is just facing the question about how the contents of a 'cognitively inaccessible' region can make the difference between knowledge and lack of it.

6 Second Case: Reference

The argument goes a little differently in the case of reference. McDowell (SRPN) says that opposition to his view involves a 'suspect conception of how thought relates to reality', a conception which 'would have to seek its support in the idea that' reference is mediated 'by way of a blueprint or specification, which, if formulated, would be expressed in purely general terms.' McDowell and Evans see this conception as one that, like the faulty theory of perception and knowledge we discussed, must keep the mind locked up within itself, unable, in the end, to conceive of anything outside.

Perhaps that last claim is correct. How would reference be accounted for on the suspect conception of how thought relates to reality? Concepts would have to contain fully individuating descriptive elements to take up the burden of reference. The concept would then refer to whatever object uniquely satisfied the description. The likely difficulty would be to find descriptive elements that could do the trick. These elements themselves would have to be free of any indexical or object-involving components, or the original conception of how thought relates to reality would be forfeit. And then, surely enough, the theory

would never allow the mind to break outside of itself. The argument would take some spelling out, but it looks fair.

6.1 Reply to Second Case

However, none of this need concern the proponent of narrow content. It is no part of the general NCA view that the narrow content of a concept is a 'blueprint or specification which, if formulated, would be expressed in purely general terms.'²¹ The NCA position requires only that the narrow content of a concept be that in the subject which allows him to refer to the object, and that it be recognizably content. It does not have to take up the full burden of reference on either of these counts. The first constraint requires only that narrow content be suitable to take up part of the burden of reference. The rest may be supplied by external relations between subject and object. The second constraint requires that referring concepts have some sort of mental content that is reference-independent, but does not require that this content be semantic content expressible in form of an individuating description. Evans and McDowell do not accept this last; and here, I think, we find one of the deepest points of disagreement between the NCA and the WCO theorists.

21. This is not violated by any of the treatments of names that I recommended for the theory of meaning. True enough, 'is Mumbo-Jumbo' may be used in a specification of necessary and sufficient conditions for being the referent of 'Mumbo-Jumbo', and it shows the name's sense. But nothing is implied about what it is that determines what the name refers to.

Both Evans and McDowell take it as an essential part of the NCA view, that on it narrow contents must by themselves fully individuate their extensions. Thus Evans (ibid. p. 202) writes 'The methodological solipsist [=NCA theorist] wants M-thinking [=narrow content] to be recognizably **thinking** - which requires at least that it be a representational state.' And 'It is of the essence of a representational state that it be capable of assessment as true or false'. It does follow from this thesis that the narrow content of a referring concept would have to be fully individuating of its referent. The reason for this is as follows.

A thought of the form a is F where a is a referring component, will be true iff the referent of a satisfies F. Now suppose that the thought is assessable as true or false purely in virtue of its narrow content. Then the narrow content of a must determine which thing would have to satisfy F, for the thought to be true. But, since narrow content is reference-independent, it could only do this by description.

But it is not, Evans and McDowell notwithstanding, any part of the NCA view that the narrow content of a thought must be capable of assessment as true or false. The condition that I placed upon narrow content, and that corresponds to Evans' condition that it be recognizably thinking, was only that it be recognizably mental, and deserving of the second part of its name. Narrow content can have that status without its being the case that the narrow content of any thought must be that in virtue of which it is assessable as true or false.

Evans moves too quickly from the requirement that narrow content be recognizably thinking to the requirement that it be assessable, taken by itself, for truth value. Why shouldn't thinking just be what goes on in the subject when he undergoes states which, when hooked up appropriately to a surrounding context, become evaluable for truth? There is nothing in the notion of a thought (the lay, not the Fregean notion, of course) that dictates that its possession of truth conditions, or truth value, is intrinsic to it, rather than something it has only given a context.²²

6.2 Points of View

Thought considered purely as the possessor of narrow content may be content, and recognizably so. Consider, if you will, a Twin Earth example involving McDowell's native. One case is as McDowell originally specified, with no Mumbo-Jumbo and a deluded native. In the other case there is a Mumbo-Jumbo, godlike in nature, and the native's concept Mumbo-Jumbo refers to him. But the native's individualistic properties, properties the existence and nature of which depend only upon the subject himself, viewed in abstraction from any particular surrounding context, are the same in the two cases.

Whether or not the native's thoughts differ in truth conditions across

22. I think that any suggestion that the issue is becoming merely terminological, would be misguided. Questions about what should be called 'thinking' or 'content' are better seen as questions about what thinking and content are. There is a genuine issue here, and an important one.

the two cases, the way the world appears to him will be the same. If we were to transplant the native from one context to the other, he would not notice. The two cases are therefore the same from the subject's point of view.²³ Whatever differences there are between the two cases, are ones that do not affect the way the world appears to the native. If the subject's point of view fails to determine the extension of his concepts and the truth conditions of his thoughts, then this shows that there is a level of content that is not truth-conditional, a kind of thought that is not, taken by itself, assessable for truth value. It does not show that the subject's subjective point of view has no content.

When we want to understand, from a sympathetic human point of view, what motivates someone, what guides his thoughts and actions, precisely what we try to find out is how things are from that person's point of view. Suppose, for example, that you are interested in understanding some aspect of Freud's thought that involved his concept of the Unconscious. I am imagining trying to understand Freud, the man, and his picture of the world, not trying to learn about the world by studying Freud. Suppose that to begin with you have little conception of what the Unconscious is supposed to be. You learn an expression 'the Unconscious', with which you associate a concept that is little more than a place-holder in your thought.

To understand Freud's concept better you must try to develop a

23. A point of contact here with Blackburn's (Blackburn op. cit.) defence of narrow content.

Freudian picture of the world. You must try to see the world through Freud's eyes. You ask yourself 'How would Freud say the Unconscious was manifest in this series of actions?', 'How did Freud think that unconscious thoughts might become conscious?' and so on. The whole enterprise would be spoiled if in constructing the picture you paid attention to whether or not the Unconscious existed, or to whether anything else in it was part of the objective world, rather than a figment of Freud's imagination. If Freud believed in something that you do not believe in, you still include the thing in your conception of the world according to Freud. If Freud did not believe in something that you do believe in, you do not include the thing in your picture.²⁴

Analogously if you want to understand how things are from the native's point of view, why he thinks what he thinks, and does what he does, you must pay no attention to whether or not Mumbo-Jumbo exists. Your sympathetic understanding of the workings of the native's mind will suffer if differences between the two cases are built into your pictures of the world according to the native. If your models of how things appear from the native's point of view make it seem as though things appear differently to him across the two cases, you have made a mistake.

If that much of the contents of the native's mind as appears in your

24. Here there is some harmony with Dennett, D., 'Beyond Belief', in Woodfield, A., ed., Thought and Object, Oxford University press, 1982, though we have different ideas about which world best represents the subject's point of view.

model is insufficient, taken by itself, to be assessed for truth value, this shows that to be recognizably thinking requires less than Evans supposed. For the understanding that is given to you by the picture is an understanding of the native's conception of the world. A conception of the world, thus solipsistically viewed, is recognizably thinking. Indeed thought is more easily recognized in this way than in any other; for one recognizes the thoughts from the inside, from the perspective of the thinker.

Understanding another's point of view, in this sympathetic, everyday, sort of way, involves adopting that point of view in the first person. One does not have to believe, or even try to believe, or feign to believe, that the world is as it appears from that point of view. One merely has to imagine how the world would appear to someone whose point of view it was.

To be sure, what one envisages while trying to grasp how things appear from another's point of view, is a world of extra-mental, extra-cranial things. But this does not render the content that one strives to understand wide rather than narrow. The content is understood in terms of how the world appears when conceived through it, not upon what there is, by of a world, to be conceived of.

I, in my world, can understand how the world appears to the residents of Putnam's Twin Earth, by thinking of water. On Twin Earth there is no water, and the extension of their concept is some other substance, XYZ. When I think of water, in order to understand the thoughts of a Twin Earthperson, I do not need to attribute to that

person a concept that has the same extension as the one I am employing myself. The other's concept need not have any extension at all. I take note only of how my concept presents its actual extension, not of what is thereby presented. The content of the concept that I thereby come to grasp is therefore narrow; it remains the same whatever its extension. The actual extension of my concept is, one might say, a ladder to the narrow content (one must, so to speak, throw it away after one has climbed up on it).

McDowell (SRPN) says this:

In practice an interpreter might say things like 'This man is saying that Mumbo-Jumbo brings thunder', and might explain an utterance which he described in this way as expressing the belief that Mumbo-Jumbo brings thunder. That is no real objection. Such an interpreter is simply playing along with his deluded subject - putting things his way.

Understanding how things are from another's point of view requires playing along with her, putting things her way. I see no reason to think that the kind of understanding of a subject that we derive by putting things her way, is inferior or defective in any way. Nor do I see any reason to think that what is thereby made accessible to us is something less than the contents of a person's mind.

PAPER THREE

SEEING WHAT IS NOT THERE

The Eye altering alters all
- William Blake

Seeing What is not There

1 Introduction

1.1 Burge's Claim

Tyler Burge¹ sets out to show that the computational theory of vision² is not individualistic. He tells us that

According to individualism about the mind, the mental natures of all of a person's or animal's mental states (and events) are such that there is no necessary or deep individuating relation between the individual's being in states of those kinds and the nature of the individual's physical or social environments. (Burge, pps. 3-4).

When Burge claims that Marr's theory is not individualistic, what he means is that there is some necessary or deep individuating relation between some of the representational contents³ attributed by the theory

1. Burge, T., 'Individualism and Psychology', The Philosophical Review, Jan. 1986.

2. I shall follow Burge in referring to it as Marr's theory. As Burge notes, very substantial contributions have been made to the computational theory by various others. The most complete and accessible account of the theory is to be found in Marr's book, Vision, San Francisco: W. H. Freeman and Co., 1982. All references to Marr are from the book, unless otherwise specified.

3. By 'representational content' I intend something that is individuated by its semantic, or intentional properties. Two identical syntactic

to the visual system, and aspects of the extra-cranial environment.

The point of the qualifier 'necessary or deep' is to make it clear that the kind of individuation in question is grounded in essential properties of what is individuated. The essentialism need not be Aristotelian; it is just that if any state or event lacked the individuating properties, it would, ipso facto, be debarred from belonging to the individuated kind. Burge's claim is that, on Marr's theory, representational contents are individuated by reference to certain features of the world outside the subject, in such a way that the contents could not exist in environments that lacked those features. Consequently the representational contents a subject is capable of having are not determined solely by what we might call her individualistic properties, properties that depend only upon how things are within her physical boundaries, solipsistically regarded. According to Burge, Marr's theory entails that two subjects whose physical states were type identical, down to the last quark, could have representations with different contents, if they inhabited suitably different physical surroundings.

I think that Burge is mistaken. Marr's theory, at least on a straightforward and reasonable interpretation, is perfectly individualistic.

tokens may instantiate (or have, if you prefer) different representational contents. A representation is a syntactic object, though typically it will have semantic properties.

1.2 Preview

In the next section I shall explain relevant aspects of Marr's theory, as I understand it, pausing to note points that will be important to the subsequent argument. In section 3 I shall recount Burge's argument that Marr's theory is not individualistic, and explain more precisely what is at issue. The issue will turn out to depend upon the description we should expect from Marr's theory, of a certain type of counterfactual conditional. Each conditional concerns, as is not surprising, which representational contents a subject would have enjoyed, had she graced a suitably different environment. Burge argues that according to Marr's theory, her representational contents would vary across environments, and that therefore the theory is not individualistic. In sections 4, 5 and 6 I shall dispute this, giving my own account of the proper reaction to the counterfactuals. The final section will be concerned with exegesis. Burge cites some passages from Marr that he (Burge) takes to support his position. I shall show that the passages cohere perfectly with my own understanding of Marr's theory.

2 Marr's Theory

Marr (p. 99) formulates his task thus:

Our overall goal is to understand vision completely, that is, to understand how descriptions of the world may efficiently and reliably be obtained from [retinal] images of it. The

human system is a working example of a machine that can make such descriptions ... one of our aims is to understand it thoroughly, at all levels: What kind of information does the human visual system represent, what kind of computations does it perform to obtain this information, and why? How does it represent this information and how are the computations performed and with what algorithms? ... How are these specific representations implemented in neural machinery?

As Burge notes, the enterprise has three separate parts (see Marr p. 24-27).⁴ i) 'The computational theory (Marr p. 24) 'in which the performance of the device is characterized as a mapping from one kind of information to another, the abstract properties of this mapping are defined precisely, and its appropriateness and adequacy for the task at hand are demonstrated.' ii) The theory of representation and algorithm: the information must be represented in some representational system, with specific primitives that make explicit specific information, and the computations must be performed by some particular algorithms. The second part of the theory describes the representations and algorithms. iii) The theory of hardware implementation: this describes how the system is physically realized. In what follows we shall be concerned with the representational aspects of the theory.

The computational theory of vision is a theory that treats the human visual system as a machine that constructs descriptions of the world in terms of three dimensional objects, from retinal images. The retinas are given computational descriptions, called 'gray arrays', that represent light intensity values in a two-dimensional coordinate system. The

4. For a more sustained treatment see 'Artificial Intelligence: a Personal View' in Haugeland, J., ed. Mind Design, Cambridge, M. I. T. Press, 1981.

computational problem is: given gray arrays as input, how does the system produce the complex descriptions of 3D objects as output?

The process breaks up into three stages (see Marr p. 37, p. 330). The first has two parts. Initially the system must construct a description of the patterns of light and shade on the retinas; (Marr p. 42) 'representations are obtained of the changes and structures in the image'. These representations make explicit certain features that will be useful for the construction of higher representations. Next, certain computations are performed to construct a first sketch of the surface outside the viewer, that represents reflectance changes on it. This is called the 'primal sketch' (for a qualification of this view of the primal sketch, see section 7, below). At the second stage, on the basis of the primal sketch, the system constructs a 2.5D sketch of the viewed surface, explicitly representing information about contour, and rough depth, in a viewer-centered coordinate frame. Finally the system constructs from the 2.5D sketch a 3D model representation, that describes shapes and their orientations relative to an object-centered coordinate frame.

The passage from each stage to the next is regarded as the making of an inference: Marr writes (p. 68) 'the true heart of visual perception is the inference from the structure of an image about the structure of the real world outside.' At each stage what is inferred from the information explicit in the representations that are operated upon goes beyond anything available from that information alone. There is thus a poverty of stimulus typical of problems in cognitive science. 'The

theorist needs to discover what a priori assumptions⁵ are brought to bear by the system, that allow it to make the inferences. (Marr p. 219, c. f. Burge p. 33)

The question to ask is, What assumptions are reasonable to make - that we unconsciously employ - when we interpret silhouettes ... as three-dimensional shapes?

If the visual system failed to make such assumptions, it would not be able to infer the nature of the distal causes of its surface irritations from those irritations. Since the visual system provides accurate representations under normal conditions, the actual structure of the world is, in practice, very important to the discovery of the assumptions built into the system.

The guiding principle is that when the representations constructed are correct under a given set of conditions, this is evidence that the inferences that yield them are based upon assumptions that are true in such conditions.⁶ The kind of assumptions involved are that, for example, (Marr p. 44) 'the visible world can be regarded as being composed of smooth surfaces having reflectance functions whose spatial structure may be elaborate', (Marr p. 51) 'if direction of motion is ever

5. 'A priori' is Marr's expression. The a priori assumptions are not necessarily true. The point is that they are innate, not learned.

6. The assumptions are not expected to be explicitly represented in the system. So it may be that the mere fact that a system succeeds under certain conditions is sufficient to vindicate the attribution of assumptions that are true in those conditions. Difficult philosophical questions arise here, into which I shall not enter, and for a discussion of which see Pylyshyn, Z., Computation And Cognition: Toward A Foundation For Cognitive Science, Cambridge, M. I. T. Press, 1984, chapter 5.

discontinuous at more than one point - along a line, for example, - then an object boundary is present'. When the assumed conditions hold, inferences made from the gray arrays come out correct - at least, within reasonable idealizations. When the assumed constraints do not hold, illusions tend to result.

I now turn to three points that will be important to our argument.

2.1 First Point

The theory describes how the visual system infers, from the structure of the gray arrays, representations of three dimensional objects in objective space. The nature of the lower descriptions is determined primarily by the exigencies of solving the overall computational problem. That is to say, they must a) be recoverable from the level below, and b) be used in the level above.

Marr (p. 36) explains:

we arrived at the idea of a sequence of representations, starting with descriptions that could be obtained straight from an image but that are carefully designed to facilitate the subsequent recovery of gradually more objective, physical properties about an object's shape.

The process of constructing representations is regarded as inferential, and the theory of vision is the theory of the inferences. Each inference requires an input (from the level below), probably additional assumptions (innate), and an algorithm to compute an output representation. Each of these must be described precisely, and it must be shown that the algorithm works, that the required representations

are actually computable in the manner described: (Marr p. 75) 'we have to analyze exactly what will work, and prove it'. The first point I emphasise, then, is that if a representation is attributed to the system, it must be shown exactly how the representation is arrived at, by a sequence of inferences, from the gray arrays. Representations require a bottom-up account.⁷

2.2 Second Point

Since representations require a bottom-up account, they do not come cheap. The more specific⁸ the informational content, the more computational activity required to derive it from the gray arrays. Given this constraint one needs a solid motivation for positing a representation. Such motivation is provided by some need that the representation will satisfy, some purpose to which it must be put. In the case of lower representations the purpose will be provided by what is required of the computations at the next level:

The critical question is, What spatial relations are important to make explicit now, and why? The answer to this, of course, depends on the purpose for which the representation is to be used. For us, the purpose is to infer the geometry of the underlying surfaces (Marr p. 80).

7. A small qualification: the gray arrays themselves are representations of light intensity values on the retinas, but we are not offered an account of how these representations are constructed. Ultimately, I suppose, that account should be forthcoming.

8. More specific in the sense of being compatible with a narrower range of distal causes: the information that there is a crack on the viewed surface is more specific than the information that a dark line appears on it that could be either a crack or something else.

The information that is made explicit at each stage is carefully chosen to suit subsequent computations.

The story is much the same for the highest representations that Marr's theory derives, the 3D model representations. Here again:

Stated baldly, the strong constraints come from what the representation is to be used for. (P. 326).

The representations, according to Marr, are used for the recognition of shapes. Such a purpose requires that the representation:

(1) use an object-centered coordinate system, (2) include volumetric primitives of various sizes, and (3) have a modular organization. A representation based on a shape's natural axes (for example, the axes identified by a stick figure) follows directly from these choices.

What is represented at a given stage is tightly constrained by what is exploited at the next stage. All the information explicit in the representations resulting from the early processes is (Marr p. 268) 'announced' in the 2.5D sketch. What appears in the 2.5D sketch is used to account for our ability to discern surface features, and as input to the algorithm that computes the 3D model representation. The elements of the 3D model representation itself are carefully motivated by what is required for object recognition. In general representations require a top-down motivation.⁹

9. This must not be taken to mean that the theory ascribes top-down visual processing. Marr (p. 100) insists that, on the whole, it does not. Processing is bottom-up, motivations are top-down.

2.3 Third Point

The theoretical role of a representation is given by its place in the sequences of inferences that begin with the gray arrays, and culminate eventually in hypotheses about the shapes that give rise to them. But it must, of course, also be shown that the visual system actually employs the particular representations attributed to it, in the particular processes attributed to it. A variety of sorts of empirical evidence are available to constrain the attributions of representations and provide clues that allow for their discovery. The main source of evidence is introspective: we know a great deal about what is revealed to us in perception. The experimental context in which this sort of evidence is mostly tapped concerns illusions.

So, for example, Chapter 3, section 3 of Marr's book, which is 48 pages long, and explains how we compute stereoscopic information from the flat gray arrays, contains no less than 15 random dot stereograms. These are pairs of square dot-patterns; the two squares in each stereogram are the same except for a region in their centers, which is displaced slightly in one square relative to the other. If one views them stereoscopically (exactly one square visible to each eye) (Marr p. 102) 'one vividly and unmistakably perceives a square floating in space above the background.' By varying the density of the dots, and the amount of displacement, one uncovers a great deal of information about exactly what is required for the computation of depth.

Evidence is fairly abundant at lower levels, but tends to peter out at

the level of the 3D model representation. There are a couple of global considerations (Marr p. 326):

First, in order to manipulate objects and avoid bumping into them, organisms must be able to perceive and represent the disposition of the objects' surfaces in space. This gives a minimal requirement of something like the 2.5D sketch. Second, in order to recognize an object by its shape, allowing one then to evaluate its significance for action, some kind of three-dimensional representation must be built from the image and matched in some way to a stored three-dimensional representation with which other knowledge is already associated.

And three additional pieces of evidence are cited. The first is that stick figures (p. 327) 'are usually recognized easily despite the limited amount of shape information they portray.' This suggests that the type of information present in stick figures is important in human vision. The second comes from certain illusions that have to do with the way we perceive elements in a pattern. Whether or not a given shape at a given orientation is seen as a square or as a diamond may depend upon axes that we assign to a larger pattern in which it appears. The third comes from deficits. Certain brain lesions result in losses of recognitional capacities that seem to require recovery of a shape's canonical coordinate system (i. e. its stick figure).

All these forms of empirical evidence that particular representations and algorithms are being used are collected from behavioural data. The broader considerations concern general requirements upon the sorts of actions that we visual perceivers are actually capable of; circumnavigation and manipulation of objects, and the like. Narrower and more specific data is collected from discrimination experiments: subjects are asked whether or not they perceive a stereogram as a

single raised square, whether they can identify a three-dimensional shape (a bucket, for example) from unusual two-dimensional representations of it or whether they see something as a square or as a diamond. Although these tests typically involve verbal behaviour, this is incidental. A dog could be trained to salivate when it saw a raised square or a bucket.

What is really being tested for is a mere discriminative capacity. Intentional contents are inferred from discriminative behaviour. Such behaviour by itself does not, to be sure, tell you what the contents are. But it is assumed that a subject cannot discriminate between two distal stimuli unless they cause representations with different contents; so the discrimination tests reveal at least when different contents arise from different arrays.¹⁰ And much of the computational theory can get by without the theorist knowing more than which differences in stimulations yield different contents. If a subject can discriminate between two stimuli, then there must be some features of the gray arrays and some innate assumptions, that allow her to do so, and the theorist must search for an account of what these are. The distinction between the two contents will then be exploited in subsequent computations. So, for example, the fact that a highlight causes a

10. The claim that different contents, rather than merely different representations, are necessary for a discriminatory capacity is, of course, controversial. But I think the claim is mandatory for anyone who thinks that contents have an explanatory role to play. What Burge would dispute is that different contents are sufficient for a discriminatory capacity. For the no-content view see Stich, S., From Folk Psychology To Cognitive Science: The Case Against Belief, Cambridge, M. I. T. Press, 1983. If Stich is right then there is a much shorter route to individualism than the one I am taking.

different representation from an object boundary will be used in the explanation of how we discern larger scaled features of the viewed shape.

My third point, then, is that attributions of representations are checked against behavioural evidence.

3 Burge's Argument

We are now in a position to consider Burge's argument that Marr's theory is not individualistic. The argument is stated in six separate steps, and a conclusion, as follows:

1) The theory is intentional. 2) The intentional primitives of the theory and the information they carry are individuated by reference to contingently existing physical items or conditions by which they are normally caused and to which they normally apply. 3) So if these physical conditions, and possibly attendant physical laws, were regularly different, the information conveyed to the subject and the intentional content of his or her visual representations would be different. 4) It is not incoherent to conceive of relevantly different physical conditions and perhaps relevantly different (say, optical) laws regularly causing the same non-intentionally, individualistically individuated physical regularities in the subject's eyes and nervous system. ... 5) In such a case, (by 3)) the individual's visual representations would carry different information and have different representational content, though the person's whole non-intentional physical history ... might remain the same. 6) Assuming that some perceptual states are identified in the theory in terms of their informational or intentional content, it follows that individualism is not true for the the theory of vision.

I confess to being somewhat perplexed by the structure of the argument. Individualism about the mind entails that mental states are

not individuated (in a necessary or deep manner), by reference to the external physical items by which they are normally caused. If the individuating relation mentioned in step 2) of the argument is of the necessary or deep variety, then little more is needed to refute individualism. Step 2) attributes to the intentional primitives what anti-individualism demands of mental states; individuation by reference to external things. So all that would be required for Burge's desired conclusion would be that a) if the intentional primitives are not individualistically individuated, then nor are the intentional contents that they make up (which is obvious), and that b) some perceptual states are identified by their contents (which is stated in step 6)). So, if 2) adverts to deep or necessary individuation, steps 3), 4) and 5) do not contribute to the argument.

On the other hand if the individuating relation of 2) is not necessary or deep, then surely step 3) would not follow from it, as Burge claims. For only if the individuating conditions in 2) were essential (in the sense explained in section 1) above) to the individuated primitives, would it be true that if the conditions differed, then so would the contents.¹¹

I think it is best to assume that the individuating relation of 2) is meant to be necessary and deep. Step 2) can therefore be seen as the crucial step of the argument. Steps 3), 4) and 5) can then be treated as an elaboration of what the anti-individualistic claim of 2) comes to.

11. Burge is normally such a careful and precise thinker that one wonders why the argument is so disorganized.

So I shall treat 3), 4) and 5) as an explication of anti-individualism, rather than as part of the argument for it.

Steps 3)-5) invite us to consider what representational contents a given earthly subject would have had, had she inhabited an alien environment in which the very same gray arrays as are here caused by one sort of distal cause, there result from a distal cause of some different sort.

Let us give our subject a name, 'Visua'. When we consider how things would have been with Visua, had matters been otherwise, we shall refer to her as Visua*. Visua* is thus Visua's counterfactual counterpart, Visua as she would have been, had she inhabited some other environment. Visua and Visua* are individualistically identical. This means that their individualistic histories are the same. They undergo the same sequence of gray arrays, their brains, individualistically considered, undergo the same machinations. But things outside their three dimensional physical boundaries may be varied.

The features that are to be varied are some of those contingently existing items or conditions that normally cause Visua's representational contents, and to which the contents normally apply. Burge claims that Marr uses such items to individuate contents. In my view, he merely uses them as a guide to discovering what the contents are. So, to pick a neutral expression: the items and conditions to be varied are those in terms of which Marr identifies contents.

In the counterfactual environment these mundane items and conditions

are systematically replaced by different ones. But these replacement items and conditions in the counterfactual environment cause the very same gray arrays as do their actual counterparts. So there are gray array types the typical distal causes of which will differ across twins. From now on when I refer to the counterfactuals, without further qualification, I mean counterfactuals that fit the broad outline just given: same bodies, environments that are systematically different in the way described.

There are two claims that Burge might make about Marr's theory and the counterfactuals. If we take him at what seems to be his word, his view is that it is a consequence of Marr's theory that in any counterfactual environment of the envisaged sort, Visua*'s representational contents would have differed from her actual ones. For he says that if the relevant items were different, the contents of the subject's representations would have been different. But there is a weaker claim that would suffice to satisfy Burge's anti-individualistic wishes, and perhaps is what he really intends. This would be that there is at least one counterfactual environment in which Visua* would have had different representational contents. If the items had been different, the contents of the subject's representations might have been different.

The truth of the weaker claim would show that Marr's theory attributes representational contents in such a way that they do not depend solely on a subject's individualistic properties. Different environments would be capable of endowing a given subject with

different representational contents, even if her individualistic properties were held constant.

I shall argue against both claims. The argument will proceed as follows. First I shall make a general point about what we may legitimately suppose to vary across environments. Then I shall outline a particular test case and consider what Marr's theory should say about it. I will offer what I call a 'straightforward' interpretation of the theory, upon which it would predict that Visua and Visua* compute representations with the same intentional contents, even though the standard distal causes of their gray arrays are quite different. I shall then explain Burge's "devious" interpretation of Marr, which contradicts my straightforward one. I shall argue that the predictions yielded by Marr's theory, deviously interpreted by Burge, would be unmotivated and incorrect. This will undermine the devious interpretation, and refute Burge's strong claim that in any of the counterfactual habitats Visua* will have different contents from Visua's. The background will then be in place to consider the more plausible, but still false, weaker claim.

The general point is this. Since the patterns of light on the twins' retinas are identical, and the very earliest representations (up to and including the zero-crossings, at a minimum) describe those light patterns, the content of those earliest representations must be the same. But the computations that derive the higher representations exploit both the input descriptions and additional, innate assumptions about the structure of the world. The representational content derived

by a given computation thus depends, in part, upon the innate assumptions. A given computation of a representation, formally (syntactically) described, may be given different semantic interpretations. So, even if we know what the earliest descriptions are, and what formal processes operate upon them, we cannot yet infer the content of the representations thereby constructed. How we interpret the resulting representations must depend upon which assumptions we attribute to the system. If we regard Visua*'s system as constructing different representational contents from Visua's, we must also regard it as making different assumptions.

4 Against Burge's Strong Claim

4.1 The First Case

Suppose that Visua*'s environment is one in which we had artfully attached a stereoscope to her face at birth. We have done this in such a way that she is not aware of it. Her eyes receive only images from the stereoscope's two screens, which we are capable of adjusting by remote control. We arrange things so that her stimulations will be identical to Visua's. The rest of her distal environment, however, will be different from Visua's.

4.2 The Straightforward Interpretation

Discussing the assumptions that allow us to infer a 3D shape from a

silhouette Marr writes (P. 219):

if a surface violates these implicit assumptions, then we should see it wrongly. Our senses should deceive us in the sense that the shape we assign to the contours will differ from the shape that actually caused them.

Visua*'s environment is one in which the viewed surfaces violate the implicit assumptions. Specifically, the system will assume that the two retinal images arise from a single surface, rather than two distinct surfaces. On a straightforward interpretation of Marr we should predict - exactly as commonsense dictates - that Visua*'s senses deceive her. Moreover, we should be able to determine what the contents of her illusions are (Marr p. 100):

The brain is capable of measuring disparity and using it to create the sensation of depth. For purposes of demonstration, a stereoscope from a souvenir shop will do: when individual views are seen with just one eye they look flat. However if you have good stereo vision and look with both eyes the situation is quite different. The view is no longer flat: the landscape jumps sharply into relief, and your perceptions are clearly and vividly three-dimensional.

Visua*'s brain is regularly confronted with stereograms, and so regularly delivers vivid, illusory, representations of 3D scenes. It does this because those are the representations it constructs from the relevant arrays. If the gray arrays that in Visua result from 3D objects were caused in Visua* by our stereoscopic facsimiles, the latter would derive just the representational contents that the former does. Visua* would misrepresent flat distal causes as 3D scenes.

Burge writes (p. 35)

If the properties and relations that normally caused visual impressions were different from what they are, the individual would obtain different information and have visual

experiences with different intentional content. If the regular, law-like relations between perception and the environment were different, the visual system would be solving different information-processing problems.

But how would it obtain the different information? How would it solve the different information processing problems? On Marr's theory the contributions of environment and organism, to the business of forming veridical representations, are very carefully separated. The environment contributes the gray arrays and conforms, usually, to the assumptions. There is no way that the environment can get the correct information into the visual system other than by bombarding its peripheries and conforming to the assumptions. The organism is given only the gray arrays and its innate assumptions to work with. It must infer, from the gray arrays, what caused the bombardment.

Marr's theory is designed to tell us how the information processing problem for this environment is solved. It is solved by the brain's construction of a representation: the representation is constructed, with the aid of innate assumptions, from the gray arrays. The representation will be correct if and only if the conditions assumed by the system obtain (barring flukes, in which case the description is arrived at by an unsound inference). When the conditions do obtain, then the representation will typically be correct, and the information processing problem will typically be correctly solved. When the conditions do not obtain, a representation is constructed that fails correctly to depict its distal cause.

The construction of representations is thus only the extraction of information about the environment when the assumed conditions hold. If

the conditions do not hold, then the system will fail to deliver correct representations. It won't magically solve another problem. It will screw up.

If the informational content of a representation is defined by what it actually regularly covaries with, then, of course, the twins' representations will carry different information. The question is whether that sort of informational content is what the visual system manipulates according to Marr's theory. I am about it to argue that it is not.

4.3 The Devious Interpretation

The straightforward interpretation requires supposing that the predictions the theory delivers about the representations constructed from specified gray arrays are not limited in applicability to a range of contexts that excludes the stereoscope world.

Burge would argue that the predictions are thus limited. He points out that on Marr's theory (p. 33):

the information carried by the representations - their intentional content - is individuated in terms of the specific distal causal antecedents in the physical world that the information is about and that the representations normally apply to.

And concludes:

Thus the individuation of intentional content of representational types presupposes the veridicality of perception.

And (p. 35):

The methods of individuation and explanation are governed by the assumption that the subject has adapted to his or her environment sufficiently to obtain veridical information from it under certain normal conditions.

And (p. 37)

the intentional content of one's visual state (or representation) is individuated against a background in which the relevant state is normally veridical.

So the devious interpretation of Marr runs like this. We are to assume that in any environment where representations have standard causes under normal conditions, the representations will be correct under those conditions. So, (the interpretation proceeds) if a creature's visual system is such that it constructs a representation of a certain type, R, from arrays that are standardly caused, under normal conditions, by a thing of a certain type T, then the representation will have an intentional content that represents those causes as things of type T.

The devious interpretation is thus the analogue for visual representation of a causal theory of reference of the kind that Putnam¹² has applied to natural kind terms of natural language, and that Burge himself¹³ has extended to the content of the propositional attitudes attributed in folk psychology.

On the devious interpretation, Visua* is assumed to see correctly

12. Putnam, H., 'The Meaning of 'Meaning'', in Gunderson, K., ed., Minnesota Studies in the Philosophy of Science, Minneapolis, University of Minnesota Press, 1975.

13. See e. g. 'Other Bodies', in Woodfield, A., ed., Thought And Object, Oxford University Press, 1982.

under normal conditions in her environment. Her representations will represent their typical normal causes correctly. It follows, then, that she will not be representing the two surfaces of the stereogram as a single 3D scene, but as two flat surfaces.

But Burge's argument for the devious interpretation is invalid. It is certainly true that Marr, when studying humans in earthly context, typically supposes that what a representation represents is what causes it under normal conditions. But it does not begin to follow from this that if some other type of distal cause had been the typical one, the representation would have represented that.

What the theorist needs to find out is what representations are constructed under what circumstances. One good way to do this is to find conditions under which the subject appears to see correctly and then to infer that what is seen is what is there. Under those conditions the representations will represent their distal causes. For us, here, it is roughly true that the normal circumstances are the ones in which our sight works well. Since that is so Marr may attribute to our representations contents that typically are true of their normal distal causes. In an environment where the normal circumstances were not propitious for vision, the representational contents would not typically be true of their normal causes; they would typically misrepresent them.

There is not the slightest reason to believe that in any environment where our representations had normal causes, we would be able to see those causes correctly. Such a supposition lacks independent plausibility, and is certainly not supported by the computational theory

of vision.

I argued above (in section 2.3) that, on Marr's theory, attributions of representations are checked against behavioural evidence. So, for example, Marr uses the observation that we successfully circumnavigate and manipulate objects, to justify attributing representations in the form of 3D models. Attributions of representations that would be in line with the devious interpretation could not survive even this crudest sort of observational test. For Visua* would not act in any manner appropriate for one who was correctly perceiving images on two flat screens just in front of her eyes. She would not, for example, try to remove the screens, in order to see what was going on behind them. Behavioural evidence of the kind actually demanded by Marr would therefore count against the devious interpretation.

The evidence would, however, be quite consistent with the straightforward interpretation. That interpretation predicts that Visua* would be illuded, that she would be representing the stereoscopic images as a single 3D scene. That is just the prediction that fits the data.

The straightforward interpretation is also in line with intuition. We know what the phenomenology of 3D vision is like. We know what it is like to view a stereoscopic picture, and what it like to see things as flat. We know what it would be like for Visua* to have the experiences she has, phenomenologically construed. It is very hard to believe that experiences like that could have representational contents that were

true of flat surfaces.¹⁴

4.4 Recapitulation and Remark

Let me recapitulate the argument so far. Burge correctly perceived that Marr attributes representations that are correct under normal circumstances here. Burge inferred, incorrectly, that the underlying principle was that vision is always correct under normal circumstances for the seer. A better reading of Marr is that since we have independent good reason for supposing that we see correctly under normal circumstances here, the contents the theorist attributes are ones that in fact represent their typical distal causes. We have no reason for supposing that Visua* sees correctly in her environment, and so no reason to suppose that her representations depict their typical normal causes.

The content of higher representations interacts substantially with behaviour. When behaviour is appropriate in the light of distal stimuli, and this fact is legitimately explained in terms of perception, we may infer that perception is yielding correct representations. When behaviour is not thus appropriate, we should predict that perception is awry.¹⁵

14. Perhaps a supporter of Burge would hold that the phenomenal properties of the twins' perceptual experiences would vary. That would be an interesting extension of anti-individualism, and one that would stand in radical need of an argument.

15. Inappropriate behaviour is not, to be sure, conclusive evidence of misperception. Peculiar desires or false non-perceptual beliefs may be what is responsible for the wacky behaviour. But all that is required to

One hypothesis that would be quite reasonable in the light of the behavioural evidence, would be that Visua*'s representations have exactly the contents that Visua's do. Since Visua* is Visua's twin, all of Visua*'s movements would be ones that someone would make, if she were seeing what Visua actually sees, and responding rationally to her perceptions. When Visua sees and walks round a sleeping rhinoceros, Visua* will travel a similarly shaped route. It will be as if Visua* herself is subject to the illusion that she, too, sees a sleeping rhinoceros. It would thus be quite safe to attribute to Visua* mental states, including perceptual states, that are just like Visua's. Nothing that Visua* does, non-intentionally described, would count against the hypothesis.

The right way to make sense of what Visua* was doing would seem to be to ascertain in what sort of environment her motions would be appropriate. The obvious candidate for such an environment would be Visua's, and so attributing to Visua* contents that would be true, were she confronted with the distal causes of Visua's representations, would be sensible. But Visua's environment might not be the only one in which Visua* could reasonably be supposed to be seeing successfully. What if there were another one? That is a topic for the next section.

undermine the devious interpretation is that systematically wacky behaviour sometimes be explicable in terms of misperception, rather than other attitudes.

5 The Weaker Claim

My argument in the First Case depended on the point that successful sight under normal circumstances cannot be presupposed for just any old brain in any old environment. Where the subject's behaviour is inappropriate to her environment, or appropriate in a way that can be explained without attributing correct visual representations, we should not automatically attribute correct representations. There is thus nothing to count against the straightforward interpretation.

5.1 Second Case

Suppose, however, that Visua* is perfectly adapted to her environment and that the success of her meanderings through it is attributable not to fluke or artful contrivance, but to her visual perception. In such a case Burge's argument could proceed without the false premise that veridicality is automatically assumed to result from the existence of normal causes.

We shall assume, then, that Visua* wears no stereoscopic glasses and can see successfully in her environment. The typical distal causes of some of her gray arrays must still differ from the causes of Visua's counterpart arrays. But we shall suppose that these differences, whatever they are, would not give the lie to the view that each protagonist could correctly perceive the distal causes of her own

irritations. This could happen if, say, the differences between the environments were small and subtle, or the items that differed across the two cases played no significant role in the subjects' lives, or the items were out of range of senses other than sight.

Burge's argument could then proceed as before. The theorist would discover that Visua* could see correctly under conditions that were normal for her. The representations to be attributed would be ones that, under those conditions, depict their normal causes. By hypothesis, constraints that are true here, and assumed by Visua's system, are false in Visua*'s world. So, Burge's argument would go, the theorist should attribute assumptions to Visua*'s system that were correct for her world. And, just as Visua's assumptions allow her to construct correct representations of her distal stimulations on Earth, so would Visua*'s assumptions allow her to construct representations that correctly depict her distal stimulations on her world. Since the stimulations and constraints differ across worlds, Visua and Visua* will construct representations with different contents.

The case under dispute may be schematically described thus. On Earth (E) we have a constraint, C_e , that is replaced on Twin Earth (T) by a different constraint C_t . On E the normal cause of gray array type G is some item O_e ; on T the normal cause of gray arrays of type G is some other sort of thing, O_t . The constraints are such that if the visual system assumes C_e it will infer from arrays G the presence of an O_e , and so form a representation of intentional type R_e , that is satisfied by, and only by, objects of type O_e . But if it assumes C_t it

will infer from arrays G the presence of an O_t , and form a representation R_t that is satisfied by and only by O_t s. (See Key, part (a)).

KEY

(a)

E: Earth T: Twin Earth G: A type of gray array
 O_e : the type of object that normally causes Gs on E
 O_t : the type of object that normally causes Gs on T
 C_e : A constraint that is true on E, but false on T
 C_t : A constraint that is true on T, but false on E
 R_e : A representation that is true of, and only of, O_e s
 R_t : A representation that is true of, and only of, O_t s
A visual system assuming C_e will construct R_e s from Gs
A visual system assuming C_t will construct R_t s from Gs

(b)

C_n : A constraint that satisfied if either C_t or C_e are
 R_n : A representation that is true of both O_t s and O_e s
A visual system assuming C_n will construct R_n s from Gs

5.2 Undermining of the Devious Interpretation

On the devious interpretation, we should argue as follows. Visua* sees correctly on T. On T, Gs are typically caused by O_t s. So seeing correctly requires the construction of R_t s from Gs. So Visua*'s system must be assuming C_t , and constructing representations R_t from Gs. Visua's system, on the other hand, assumes C_e to construct R_e s from

Gs. Since R_e s have different extensions from R_t s, Visua and Visua* undergo representations with different contents.

But this argument is invalid. It does not follow from the fact that Gs are caused by O_t s that correct representations constructed from Gs must be representations of O_t s, hence of intentional type R_t . Any representation that includes O_t s in its extension will be correct on T. In particular a representation that is satisfied by both O_e s and O_t s will be correct on T. If the representations constructed by Visua and Visua* are of some neutral type R_n that is satisfied by both O_e s and O_t s, then they may see correctly while yet having the same representations.

Ascribing to both Visua and Visua* representations that would be correct in either environment is exactly what would be required by the straightforward interpretation. The straightforward interpretation thus requires that both protagonists construct representations of the neutral intentional type R_n . Correlatively, the assumptions we would attribute would be broad enough to be satisfied in both environments. The systems would neither assume C_t nor C_e , but some neutral assumptions C_n that would be true in both environments. (See Key, part (b)).

Burge's argument that the computational theory is not individualistic depends upon reading into it a particular form of a causal theory of reference. The Putnamian theory might be true of some parts of natural language and even of some parts of folk psychology, but one cannot automatically assume that it applies to a computational theory of a perceptual module. Scientific psychology will perforce depart from folk

psychology in various ways. One such departure might well be a failure to adhere to a causal theory of reference of the folk variety.¹⁶

Burge must attribute to the computational theory of vision a particular way of choosing, from among the various ways of describing the typical causes of a correct representation, the one that will give the content of the representation. In our example the recommended selection for Visua* would have to be R_t rather than R_n . I find nothing in Marr's theory to support this interpretation. Both the causal theory itself and the particular Putnamian version of it are things that Burge himself has, without provocation, added into the computational theory.

However, even if the theorist of vision espouses no Putnamian causal theory, she might still attribute contents in the way predicted by the devious interpretation. If that were the case, it would be because the balance of various sorts of evidence and argument happened to motivate attributing to Visua and Visua* the specific representations R_e and R_t , respectively, rather than R_n .

I shall argue that the balance would never tip that way. The attributions predicted on the straightforward interpretation are more in line with Marr's actual methods, and with the dictates of good explanation.

16. My own view is that folk psychology comes unstuck on twin earth cases, dictating both that the twins do have the same psychological states, and that they do not. I reserve that argument for another paper.

5.3 Defence of the Straightforward Interpretation

Let us elaborate the counterfactual. Let us bring Visua* to Earth, and place her and Visua in a laboratory together. Visua* is still assumed to be an evolutionary product of and to have grown up in one of the counterfactual habitats, her visit to Earth is merely three-dimensional.¹⁷ Let us subject them both to two experiments. In the first experiment we arrange an experiment in which conditions C_e are violated, and conditions C_t obtain, and subject the subjects to objects of type O_t , to cause arrays of type G . In the second experiment we arrange matters so that conditions C_e obtain, and G s are caused by O_e s.

The devious interpretation would have us say that in the first experiment Visua* correctly sees O_t s, but Visua is subject to the illusion that she is seeing O_e s. In the second experiment it is Visua who suffers the illusions and Visua* who is seeing correctly. On the straightforward interpretation they would both be seeing all the distal causes as O_n s.

Representations require a top-down motivation. Top-down motivations come from finding some use or purpose to which the content is put. So to motivate the attribution of O_t s and O_e s rather than mere O_n s, we would have to find some purpose that the more specific contents could

17. If the strain of imagining that Visua* is Visua herself, counterfactually described, is now unbearable, you may suppose that the twins are merely physically type identical.

serve, but that the vaguer contents could not. We would need to find some higher computational process into which the contents would feed, and it would have to make some difference to this process whether its input was on the one hand R_e or R_t , or, on the other, R_n . What would this be? Nothing that issued in any discriminative ability, or recognitional capacity. For the twins are twins, and will be the same in every testable respect.

The twins' performances in all the tests we could devise would be the same. They will, for example, register 3D effects from all the same stereograms, be able to identify exactly the same 3D shapes from silhouettes, make the same distinctions among surface markings (for example distinguish shadow boundaries from highlights from texture changes from object boundaries), make the same groupings in patterns and so on.

But it is tests of exactly this kind that the theorist actually uses to determine how many distinct contents she needs to attribute. Attributions of representations are checked against just this sort of behavioural evidence. And there is good reason for this. It would be a violation of Occam's razor to invoke different contents where the invocation would be doing no work. There would just be no point to invoking the two contents, where one would do. For there would be no theoretical purpose served by distinguishing between the contents. Consequently the straightforward interpretation has economy on its side. The devious one scratches where there is no itch.

Moreover, the global considerations that help to justify the

attribution of contents would be the same in the two cases. For successes and failures, behaviourally construed, would come at exactly the same points. The twins would make the same circumnavigations and manipulations. Visua* would duck to avoid flying objects, duck to avoid images of such in a 3D movie, read scrawly handwriting, sew buttons on a shirt, make drawings of what she sees, and so on and on, just as Visua does. It would seem odd to suppose that throughout all this Visua* would be systematically misperceiving something that Visua was representing correctly. There would be no motive for attributing different contents, no explanatory gain to be achieved by doing so.

On the straightforward interpretation we would simply reason that both twins could solve the information processing problems set by both Earth and Twin Earth. We would infer that their representations were of type R_n , and that they were assuming neither C_t nor C_e but the more conservative C_n - all that was required to account for their representations. And that is exactly what the evidence warrants.

To attribute contents deviously rather than straightforwardly would not only be out of line with Marr's actual practice, it would run directly counter to basic canons of good explanation.

The general message is that we can never suppose that a mere difference in distal cause, however regular, however ubiquitous in evolutionary background must make a difference to representational content. To show that the differences between the twins' environments showed up in perception, we would have to show that their representations were specific enough to distinguish the two sorts of

distal cause. Given the identity of discriminative abilities, this could never be shown. The best theoretical description will always be one in which the representations fail to specify their extensions at a level that distinguishes the two sorts of distal cause. It will always be better to suppose that the extension includes both sorts of thing.¹⁸

Being tickled by something is not tantamount to knowing what that thing is.

6 A Specific Example

Burge (p. 42-3) himself provides an example that nicely illustrates what is wrong with the devious interpretation, and right with the straightforward one. It will be worth going over the argument again with his specific example.

He asks us to imagine a subject, P, among whose early visual representational types is one regularly caused by thin shadows. In such a case, Burge argues, the perception would represent shadows. If the same perceptual type were on occasion caused by a small crack, Burge says, then P would misperceive the crack as a shadow. So, even under normal circumstances P may be fooled by a crack. But suppose that,

18. Note that refusing to proliferate contents beyond what is needed to explain discriminatory capacities is not special to the computational theory of vision. The reason why we do not attribute, for example, a concept of bank-managers to dogs is that dogs cannot discriminate bank-managers (or banks) from other things.

counterfactually, the optical laws were different. In the counterfactual world there are no instances of the right sort of shadow. Instead the very gray arrays that are factually caused by the shadows are, counterfactually, caused by small cracks. In such a case, Burge says, P would be visually representing cracks as cracks.¹⁹

We face a choice. Do we attribute to P in one environment representations of shadows, in another, representations of cracks? Or do we follow the straightforward route and attribute to P, in both environments, representations of crackdows (thin, dark, marks that could be either shadows or cracks)?

What would be required for positing the different representational contents would be a top-down motivation. We would need a reason for supposing that higher processes exploited the information that it was a shadow (crack) that marred the visible surface, rather than a crackdow. If the representation of a crackdow would account for all the higher processing just as well as would the representation of a shadow (crack), then it is a crackdow representation that we should attribute. Anything more discerning would be unmotivated.

But, by hypothesis, P cannot discriminate between shadows and cracks. We could therefore never locate a capacity of P's the explanation of which would require the attribution of the more precise

19. Note the analogy with water and XYZ: if there were a small amount of XYZ on Earth we would probably misidentify it as water. We would call it 'water', but we would be wrong, as chemical investigation might show. But on Twin Earth the word 'water' is true of XYZ, which predominates there.

content. No top-down motivation would be available.

We do form representations, during early vision, that have contents of about the specificity of crackdow, and ones that have about the specificity of shadow or of crack. It will be instructive to consider how these are arrived at, and what motivates their attribution.

The first representations that reliably covary with external phenomena are formed when the outputs of the zero-crossing filters are conjoined (c. f. Burge p. 33). Marr attributes to these representations the content edge.

It is not abundantly clear whether "edge" representations are meant to represent features of the retinal images themselves, or features of the viewed surface. In three places Marr states that representations at this level are still of the retinal images (p. 71, p. 91, p. 366); and in others (p. 41, quoted above, p. 93) strongly implies this. But he also tends to talk as if they can represent elements of the surface (p. 52, p. 68, p. 91); and this is how Burge seems to interpret him. Another possibility would be to think of these representations as being of the outputs of the zero-crossing filters. In any case, however, representations that really do refer to external things must arise here or soon after.²⁰

It is interesting that the theory survives unclarity on this point. It

20. My own view, unlike Burge's, is that we should avoid attributing to the system representations of external things for as long as possible, probably until the 2.5D sketch itself. At lower levels we do not need to attribute such contents to explain what the system is doing.

does so precisely because nothing is at stake except the inferential role of the representations, their place in the sequence of steps that mediate the passage from gray array to 2.5D sketch and the 3D model representation. It makes no difference whether one treats the edge segment representations as already referring to marks on the visible surface, or treats them as referring to previous representations and sees the later processes as inferring from them the presence of such a mark.

Let us, in any event, follow Burge in supposing that edge representations do represent elements of the viewed surface. It is very important to realize that 'edge' has a very broad meaning. As far as the early representation goes, an edge could be the boundary of a shadow, a highlight on a surface, a property of texture, a real surface reflectance change (for example, one caused by a patch of ink), or an object boundary.²¹ 'Edge' thus functions like 'crackdow' to give the content of an unspecific representation the extension of which is not matched by that of any simple English expression.

The higher processes perform computations to decide what sort of "edge" is really present, whether it is a shadow boundary, or a real physical edge, or whatever. These computations require exploitation of further information, for example about depth or motion.

21. It would therefore be impossible to construct a twin case for the content edge. Since any reflectance change on a surface counts as an "edge", any twin environment will either provide "edges" or be a world of many illusions. If the distal stimuli were not even marks on surfaces, then seeing successfully would be out of the question for creatures with brains like ours.

What motivates the attribution to the system of a representation that is really of, say, a shadow boundary, rather than of an "edge", is the fact that we can tell, under normal circumstances, what sort of "edge" we are seeing: we can tell, that is, whether we are seeing a shadow boundary, or a highlight, or a real physical edge. The representational content is manifest to us subjectively, and clearly detectable in discriminatory behaviour. Representations with different contents are needed to explain the discriminative capacity, and this provides the top-down motivation for positing them.

It is only because we can normally tell which sort of "edge" we are seeing, that Marr attributes the different representational contents, and seeks for an account of how they are inferred from the gray arrays.

Similarly for small cracks and shadows. Only if the subject could visually discriminate between them, under normal circumstances, would it be correct to attribute to her crack or shadow representations, rather than crackdown representations. The system will represent crackdowns as shadows, or as cracks, rather than merely as crackdowns, only if it extracts this information from the retinal array, with the aid of additional cues, assumptions and computations. Whether or not the system is doing this depends upon it, not upon what normally causes the gray arrays.

It is no accident that Burge had to try to make up his own example, rather than take one from the theory itself. The reason is simply that every content that the theory actually does attribute has a top-down motivation and a bottom-up account. The result is that it is too tightly

constrained by the individualistic facts, the gray arrays and the testable dispositions, to construct a twin case. Burge claims that his general argument, which includes the shadow and crack example 'is independent of the theory of vision that we have been discussing [i. e. Marr's theory]. It supports and is further supported by that theory.' It seems to me, rather, that his argument flies in the face of the theory.

7 Exegesis

Burge's devious interpretation of Marr is defended by two lengthy . examples of specific theoretical claims, and by a couple of additional quotes. In this final section I will show that the straightforward interpretation is compatible with all Burge's evidence.

Here are the essentials of one of the specific claims (Burge p. 33):

In building up informational or representational primitives in the primal sketch, Marr states six general physical assumptions that constrain the choice of primitives ... [for example:] a) the physical world is composed of smooth surfaces having reflectance functions whose spatial structure may be complex; b) markings generated on a surface by a single physical process are often arranged in continuous spatial structures - curves, lines etc.... These assumptions are used to identify the physical significance of - the objective information given by - certain types of patterns in the image. The computational theory states conditions under which these primitives form to carry information about items in the physical world. ... [C]onditions are laid down under which certain patterns may be taken as representing an objective physical condition; as being edge, boundary, bar or blob detectors.

I accept this as a fair account of something that Marr does. But how does this support the claim that 'the intentional primitives of the theory

... are individuated by reference to physical items and conditions by which they are normally caused'? Certainly there is reference to external contingencies, but the reference has, even in Burge's own words the form of 'assumptions ... used to identify' informational content. I take this to mean that the theorist uses the assumptions about the world in order to find out what the representations deployed by the visual system are. When Marr seeks, as Burge says, to constrain the choice of primitives, he is trying to limit hypotheses about which primitives there are in the visual system. He is not laying down conceptual constraints on what it is for there to be a particular primitive in a system.

In his second example Burge expresses the relation between Marr, environment and representation thus (ibid.): 'Marr motivates a central representational primitive by stating physical constraints'. Again this is true, and again this fails to support Burge's interpretation. The physical constraints motivate the representational primitive by providing evidence about what is being represented. Nothing is entailed about individuation.

The conflation of what is taken to justify the ascription of a given content with what makes it the case that the content is there, is apparent at numerous points in Burge's exposition. He (p. 27) quotes Marr (p. 43) as saying

The purpose of the representations is to provide useful descriptions of aspects of the real world. The structure of the real world therefore plays an important role in determining both the nature of the representations that are used and the nature of the processes that derive and maintain them. An important part of the theoretical analysis

is to make explicit the physical constraints and assumptions that have been used in the design of the representations and processes.

Burge says that this is 'tantamount to the chief point about representation or reference that generates our non-individualistic thought experiments'. The chief point is 'what entities in the objective world one intentionally interacts with ... affects the semantical properties of ... representational types ... and how we individuate them'. So Burge takes 'determining the nature' to mean something like 'partly constituting the essence'. Marr's words here are compatible with such a reading, but they may more naturally be understood as making a much more innocent claim.

It is likely that by 'purpose' and 'design' Marr intends evolutionary purpose and design. If that is right, then the kind of determination in question would be evolutionary determination. Marr says (Marr p. 75) 'nature seems to have been very careful and exact in evolving our visual systems'. For this reason we should expect what goes on in the visual system to be highly sensitive to features of the real world. The sensitivity shows up in the representational primitives and the assumptions exploited by the system. These have been 'carefully and exactly' selected by evolution. The relation between the structure of real world and representational content is that the former is an evolutionary cause of the latter. Marr's point is that if we are to discover the nature of the representations and processes we must constantly look to the outside world for evidence. Nothing is said, or implied, about individuation in the strong Burgean sense.

A quote from Marr that might seem to lend to support to Burge is this (Marr p. 44, Burge p. 27)

It is of critical importance that the tokens [representational particulars] one obtains [in the theoretical analysis] correspond to real physical changes on the viewed surface: the blobs, lines, edges, groups and so forth that we shall use must not be artifacts of the imaging process, or else inferences made from their structure backwards to the structure of the surface will be meaningless.

At a casual perusal this passage might seem to indicate adherence to the view that if a representation is to have meaning, it must correspond to something in the real world. Lack of such a correspondence would render the token meaningless. But this is nothing like what Marr is actually saying. The comment is made in the course of a description of the task of defining (ibid.)

a representation of the image of reflectance changes on a surface that is suitable for detecting changes in the image's geometrical organization that are due to changes in the reflectance of the surface itself or to changes in the surface's orientation or distance from the viewer.

[The representation] should include some type of "tokens" that can be derived reliably and repeatedly from images and to which can be assigned values of attributes like orientation, brightness, size ...

The point is that if one selects tokens that do not correspond to real physical changes then when one tries to construct from them a complex sketch of what is going on outside, one will end up with gibberish. They will simply not be suited to the task of building up coherent complex representations. Marr is making an empirical claim, not an a priori stipulation. What else would an empirical scientist be doing? There is no support for Burge's anti-individualism to be found in these passages.