1988

Formal Semantics And Pragmatics Of Belief

Ronald Romane Clark

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

FORMAL SEMANTICS AND PRAGMATICS OF BELIEF

by

Ron Clark

Department of Philosophy

Submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Graduate Studies

The University of Western Ontario

London, Ontario

March 1988

ISBN 0-315-40761-1

THE UNIVERSITY OF WESTERN ONTARIO

FACULTY OF GRADUATE STUDIES

CERTIFICATE OF EXAMINATION

Chief Advisor

William J. Harper

Advisory Committee

R. Pearce

Examining Board

Anil Gupta

Ausonio Marras

R. Pearce

[illegible]

The thesis by

Ron Clark

entitled

Formal Semantics and Pragmatics of Belief

is accepted in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy

Date 21 April 1988

Nigel B Crowther

Chairman of Examining Board

ABSTRACT

It is argued that an adequate model-theoretic treatment of belief requires the devices of a "two-dimensional" intensional logic; a logic, that is, that can represent relevant features of the epistemic situation of an agent in something like the way a logic of indexicals represents possible contexts of discourse. The general approach is illustrated informally by means of several examples, Kripke's "Puzzle About Belief" among them. A model theory is developed for a formal language containing a belief operator, quantifiers, description operator and identity predicate, and a truth predicate and reference operator. Certain properties of the model theory are demonstrated, and the application to "indexical belief" is sketched.

ACKNOWLEDGEMENTS

I am grateful to many members of, and visitors to, the Philosophy Department at U.W.O., not only for the opportunity to profit from discussions of various of the issues addressed below, but for their encouragement as well. My thanks go especially to Bill Demopoulos, Frankie Egan, Bruce Freed, Ausonio Marras, Bob Matthews, Glenn Pearce, and Ed Stabler; and also to Rich Hall, for correspondence. Portions of the dissertation were presented to audiences at the University of Calgary, McGill University, and the University of Alberta, whom I also thank for discussion.

By far the greatest debt I have to acknowledge in this connection is to Bill Harper, who first inspired my interest in this project, and whose continued enthusiasm has been a significant factor in its completion. I consider myself very fortunate to have worked with him.

TABLE OF CONTENTS

INTRODUCTION

In the work to follow I wish to sketch a formal treatment of the belief operator; an operator, that is, in a formal language, so constrained semantically as to make it reflect important features of the English verb 'believes'. The project is thus an exercise in formal philosophy of language. The attempt to make this account responsible to the most pervasive features of common practices of belief attribution necessarily leads to claims about the nature of belief, and evidence of agents' beliefs, and so forth; in other words, it involves forays into philosophy of mind and philosophy of psychology. And in fact, the approach to be developed has striking implications for theories of propositional attitudes. But it is the broadly semantic task which is my principal concern here, and accordingly, its sometimes controversial ramifications for the philosophy of mind are typically not explored. For example, there is a body of literature in which claims are made, and disputed, concerning the sort of construals under which belief statements can play a role in the explanation of behaviour, an issue which I essentially ignore. The formal job has, I believe, an integrity which permits this tack.

The account I offer is a variety of possible worlds semantics, inspired primarily by writings of Richard Montague, and Robert Stalnaker. Montague's exquisite formal work was coupled with an unconcern bordering on disdain for questions of its psychological relevance. Stalnaker, on the other hand, has brought forward a deep and highly original conception of the philosophical underpinnings of possible worlds semantics, and its role in the philosophy of psychology. This conception motivates and underwrites the use of the principal formal device to be employed in the sequel.

The proposal I am going to make, to deal with problems like that of substitution failures in belief contexts, requires that we imagine possible worlds in which certain words of English (names, for example) have different intensions than they do in the actual world. Against this approach it is sometimes objected that words are essentially meaning-bearing items; hence, worlds of the sort I want to consider don't exist-- any linguistic item whose intension differs from that of any given item of English, is a different word. I invite those who share this concern to construe me as speaking, not of the same word having a different meaning, in some other world, but rather of some word, which has the same phonological and syntactic properties as the given item, and a different intension, taking the place of the "original" word, in this other world.

Much of the focus of the discussion to follow will be on belief statements whose complement clauses are singular statements involving proper names. The account itself is, however, general, and in particular is not limited by considerations of grammatical category of items in the complement clauses.

Still, the generality may not always be obvious, in the light of particular examples, and claims which seem straightforward when confined to belief statements of the sort indicated, may appear problematic when applied to the wider class of belief statements. I therefore ask the reader to keep in mind that, in the context of certain examples or arguments, I may sometimes speak simply of "belief statements", where it would be more precise to say "belief statements whose complement clauses are singular statements involving proper names". The intention will be clear, I hope, and the degree of looseness of speech, tolerable.

In a similar vein (but possibly less innocuous) is my use of the "de dicto/ de re" terminology. According to what is perhaps the traditional view, de dicto and de re attitudes are distinguished by their objects. The former have sentences, or sentence-like propositions as objects, and the latter, (typically non-linguistic) things in the world. And de dicto and de re statements are usually distinguished by syntactic tests, based, for example, on the failure (or otherwise) of certain inferential principles, such as existential generalization, and various substitution principles.

This is not the sense in which I use the terms, as will become clear. The objects of belief are not, on my view, intrinsically de dicto or de re. Rather, belief statements admit of de dicto and de re construals, based on pragmatic considerations, and the two sorts of construal may or may not, in a given instance, differ. So when I speak of an agent's de dicto belief that P, I should be understood to refer to that belief which is indicated by a de dicto construal of the statement that the agent believes that P; and likewise when I speak simply of an agent's de dicto beliefs.

Roughly, the distinction I wish to mark is between those construals of belief statements that are, and those that are not, conditioned by certain features of the epistemic circumstances of the subject of the attitude. So, for example, I freely apply the terms to belief statements whose complement clauses are general statements. Despite such idiosyncrasies, I believe that this distinction has enough in common with traditional ones to warrant the common appellation.

In Chapter I, I sketch informally the treatment of belief, and attempt to marshal intuitions in its favour by applying it to a couple of problem cases, most notably Kripke's "Puzzle About Belief". Next, in Chapter II, the theory is characterized in somewhat more general terms, to locate it in a field of possible alternatives, and to facilitate arguments which suggest that the likeliest alternatives, within a broadly model-theoretic approach, are seriously flawed. An example due to Thomason, involving propositional attitude statements whose complement clauses are universal generalizations, is examined as well.

The formal presentation of the theory occupies Chapter III, where, in addition, a number of simple (although abstract) examples are worked through, to show that the model theory has the right properties. There follows a brief discussion of belief statements whose complement clauses contain indexicals, indicating informally how such statements could be accomodated within the framework adopted, and certain puzzles generated by such statements, resolved.

It should be noted at the outset that, in a certain sense, this is not a formal <u>semantic</u> theory at all. The central technical device employed is designed to reflect the role of certain extra-linguistic factors in the assessment of belief statements, factors which are plastic in the extreme. The concluding remarks (Chapter V) address this and related matters directly.

UNPUZZLING *DE DICTO* BELIEF

It is widely held that allowing unrestricted substitution of co-referential singular terms, within the context of attributions of belief, leads to counter-intuitive results. In particular, the attribution of *de dicto* beliefs pose this problem, for plainly an agent may, through the most venial sort of ignorance, simultaneously believe, for example, that the tallest spy is a spy, and that Ortcutt is not a spy, even though, in fact, Ortcutt is the tallest spy. In other words, permitting such substitutions will sometimes lead us to attribute to an otherwise reasonable agent, patently contradictory beliefs. For this reason, it is quite plausibly held that such substitutions are illicit.

The problem is especially acute for those who subscribe to the theory that proper names have no descriptive content, that they simply *name*. In the idiom of possible worlds semantics, names, on this theory, are rigid designators. Consequently, two names which refer to the same object do so necessarily, in the sense that the statement asserting their identity is a necessary truth.

While it seems reasonable enough that the interchange of arbitrary co-referential singular terms be prohibited in *de dicto* belief contexts, it is more difficult to justify the prohibition when applied to co-referential names: statements which differ only in occurrences (in extensional contexts) of co-referential names will be necessarily equivalent, and hence, presumably, they will have the same content, or express the same proposition. What, then, are the grounds for discriminating between belief attributions that differ only in occurrences of necessarily equivalent statements?

The grounds are the same as in the more general case. Where substitution of co-referential names is allowed, it seems, contradictory beliefs will congregate. Surely an agent might, without inconsistency, believe that Cicero was bald, and that Tully was not. As before, it is argued, substitution must be shunned. And the need to do so, even with proper names, suggests that names do after all have some kind of "sense", to which belief contexts are sensitive.

According to Kripke,[1] the above characterization of the problem is too crude. It fails to reflect the role of certain other implicit principles, relating to belief attributions, in producing paradoxical consequences. That is, the "paradoxical" belief statements result not just from the substitution principle, but also from two other principles that pertain to belief attribution. Hence, the substitution principle itself is not necessarily implicated.

The two principles just mentioned are the principle of disquotation (henceforth, Disq) and the principle of translation (henceforth, Trans), and, Kripke argues, they suffice, by themselves, to produce examples quite similar to the familiar problem cases that arise when substitution is permitted.

Disq is the principle that what an agent assents to, he believes (subject to qualifications designed to exclude obviously irregular cases), and Trans is the principle that, roughly, normal practices of tranlation between natural languages, preserve truth even when applied to statements occurring within belief contexts.

It needs to be emphasized that the sort of belief contexts, and belief attributions, to which the principles are intended to apply, are de dicto belief contexts, and attributions of de dicto belief. Disq, for example, is quite implausible when the resulting belief statement is construed de re. Ralph may sincerely assent to the statement 'the tallest spy is a spy', even though there is no individual of whom Ralph believes it to be true, that he is a spy.

Trans, on the other hand, seems uncontroversial when applied to belief attributions construed de re, amounting to little more than the thesis that translation between natural languages is indeed possible. In any case, Kripke is adamant on this point, his concern is with de dicto belief.

The puzzle that Disq and Trans engender is this: Pierre, who is, in effect, an ambi-monolingual speaker of French and English, assents to statements in the two languages, whose standard translations into either language, are contradictory. In the example, he assents to both 'London is not pretty', and 'Londres est jolie'. It is easy to see that Disq and Trans will yield attributions of contradictory beliefs to Pierre. "Does Pierre", Kripke asks, "Or does he not, believe that London is pretty?"[2]

It seems plainly wrong to attribute to him contradictory beliefs-- he doesn't know that 'London' and 'Londres' name a single city. But then the only other alternatives involve rejecting his own testimony, a rather drastic method of restoring coherence to his (putative) beliefs. "As in the case of the logical paradoxes," writes Kripke, "the present puzzle presents us with a problem for customarily accepted principles and a challenge to formulate an acceptable set of principles that does not lead to paradox, is intuitively sound, and supports the inferences we usually make".[2]

I would like now to indicate a theory, or rather, the prototype for a theory, that I believe will, in large measure, meet this challenge.[3] The possibility of the sort of approach I want to take Kripke has noted (on page 244 of Kripke (1979), and especially in footnote 10). He writes:

> ...I would emphasize that there need be no <u>contradiction</u> in maintaining that names are <u>modally</u> rigid, and satisfy a substitutivity principle for modal contexts, while denying the substitutivity principle for belief contexts. The entire apparatus elaborated in "Naming and Necessity" of the distinction between epistemic and metaphysical necessity, and of giving a meaning and fixing a reference, was meant to show, among other things, that a Millian substitutivity doctrine for modal contexts can be maintained even if such a doctrine for epistemic contexts is rejected. "Naming and Necessity" never asserted a substitutivity principle for epistemic contexts.
>
> It is even consistent to suppose that differing

> modes of (rigidly) fixing the reference is responsible for the substitutivity failures, thus adopting a position intermediate between Frege and Mill ...

The articulated form of the approach I will advocate has its source in the writings of Robert Stalnaker, and in particular, his papers "Assertion", and "Propositions".[4]

In the former paper Stalnaker outlines a theory of assertion according to which "what is said", on a given occasion of utterance, may differ in content from the proposition that the utterance would normally be taken to express. Central to the analysis he offers are certain notions characterized in terms of a generalization of the familiar possible worlds semantics used to interpret modal discourse. On the usual account (as elaborated, for example, in Kripke's "Semantical Considerations on Modal Logic"), sentences in a language can be interpreted as expressing propositions, where a proposition is a set of possible worlds, or equivalently, a function from possible worlds to truth values. Then the question of whether the proposition expressed by a sentence is true or false, is just the question of whether or not the actual world is one of the worlds where the proposition takes the value "true".

But possible worlds play another role as well, in determining the truth or falsity of statements-- sometimes the proposition expressed by a given sentence can vary, depending on which possible world is actual. Sentences involving indexicals provide obvious examples.

So, on the generalized semantic account that Stalnaker offers, statements are associated with propositional concepts, which are functions from possible worlds to propositions. The possible worlds in the domain of these functions represent features of possible contexts (of utterance, in this case), that are relevant to determining what proposition is expressed. These notions represent, in a sense, a pragmatic dimension of interpretation.

The semantic apparatus just described is well suited to characterizing formally the distinction that Kripke urges (in "Naming and Necessity", and "Identity and Necessity"), between the metaphysical content of a statement, and the epistemic grounds for making the statement. Kripke's doctrine that names are rigid designators has the consequence that identity statements involving two names are necessarily true, if true at all, and necessarily false, if false. Thus, when one utters such a statement, or its denial, one expresses either a necessary truth or a necessary falsehood. But in many cases, such utterances are quite non-trivial; that is, they are informative.

This is explained, intuitively, by the fact that one might not know whether the statement is true or false, in spite of its metaphysically august character. And formally, this lack of knowledge can be naturally represented by associating with the statement a propositional concept that maps some possible worlds onto the necessarily true proposition, and others onto the necessarily false one.

The possible worlds in the domain of this propositional concept thus reflect an agent's epistemic situation. They are epistemic possibilities according to the agent-- roughly, ways that things could be, for all he knows. What he intends to convey in asserting the identity statement, then, is that the actual world is one of those epistemic possibilities where the statement is necessarily true. In no one possible world is this proposition expressed by the identity statement. It is the proposition which is true, in a world w, just in case whatever proposition is expressed by the identity statement in world w, is true in world w.

Propositions of this sort are determined by the full propositional concept associated with a statement, and Stalnaker refers to them as diagonal propositions. They will for the time being be denoted by *C(P), where P is the statement in question, and C(P) is the propositional concept associated with it. They can be characterized more succinctly thus: *C(P)(w) = true if and only if C(P)(w)(w) = true. This, then, is how "what is said" can differ from the proposition actually expressed by a statement.

In the second of the papers mentioned above, "Propositions", Stalnaker argues that propositions, in the technical sense in which we have been using the term, are properly thought of as the objects of propositional attitudes, such as belief. So to believe a proposition is to believe that the actual world is a member of a certain set of possible worlds.

I do not wish to review those arguments here, but rather, just to look at the suggestion for dealing with one difficulty raised by this view, namely, the difficulty of explaining how an agent can believe one, and disbelieve (or simply fail to believe) another, of two necessarily equivalent statements, such as, for example, two mathematical truths.

The difficulty arises in the general case because two necessarily equivalent statements are true in all the same possible worlds, and so, there corresponds to the statements only one proposition, which an agent may believe or not, but not both. This problem (which differs little from the one just mentioned above), prompts Stalnaker to observe,

> Because items of belief and doubt lack grammatical structure, while the formulations asserted and assented to by an agent in expressing his beliefs and doubts have such a structure, there is an inevitable gap between propositions and their expressions. Wherever the structure of sentences is complicated, there will be non-trivial questions about the relation between sentences and the propositions they express, and so there will be room for reasonable doubt about what proposition is expressed by a given sentence. This will happen on any account of propositions which treats them as anything other than sentences, or close copies of sentences.[5]

In the specific case of mathematical statements, Stalnaker suggests that "... where a person fails to know some mathematical truth, there is a non-actual possible world compatible with his knowledge in which the mathematical statement says something different than it says in this world".[6]

In other words, the suggestion is that possibilities concerning semantic features of natural language, play a role in determining what proposition is in fact the object of belief, when an agent assents to one and only one of two logically equivalent statements, or refrains from assenting to a logical truth. The proposition believed is not necessarily the one that a linguistically omniscient being would interpret the statement(s) as expressing.[7]

This approach is tailor-made for handling problems of substitution in de dicto belief contexts. The problem, recall, grew out of the fact that a Kripkean account of proper names makes statements differing only in occurrences of co-referential names (in extensional contexts), necessarily equivalent. But now note that the fact (if it is one), that a proper name rigidly designates a given object, is in no way mitigated by the possibility that it could have denoted something else. Even if all proper names are always rigid designators, it still makes good sense to say, for example, "'Jocasta' might not have named Jocasta".

Now the intuitive reason for wanting to prohibit the inference from 'BaFb' to 'BaFc' (where 'BaP' represents the claim that a has the de dicto belief that P), even when 'b' and 'c' rigidly designate the same thing, is that the agent, a, might not believe that they do. This possibility will be represented in the propositional concepts for 'Fb' and 'Fc', in the fact that in some worlds 'b' and 'c' do not denote the same thing, even though relative to such worlds they are still rigid designators. In consequence, 'Fb' and 'Fc' will not express the same proposition relative to each possible world in the domain of the propositional concepts associated with them. This, roughly, is how an agent's failure to believe the substitution instances of statements he believes, is explained.

The explanation can be made more precise. On the view being sketched, the object of an agent's de dicto belief that b is F, is *C(Fb), the diagonal proposition determined by the propositonal concept associated with 'Fb'. The objects of de re beliefs, on the other hand, are the propositions actually expressed by the statements used in ascribing the beliefs.

Suppose, then, that an agent, a, has the de dicto belief that b is F; i.e., BaFb. And suppose also that 'b' and 'c' are rigid designators, and that b = c. What can be said about the agent's beliefs regarding 'Fc'? On the view sketched above, one can infer that BaFc only if *C(b = c) is true at all worlds where *C(Fb) is true.

Recalling the characterization of what it is to believe a proposition, and what it is to have the de dicto belief that b is F, it is easy to see that

$$BaFb \; \& \; [{*C(Fb)} \subseteq {*C(b = c)}] \rightarrow Ba(b = c).$$

In other words, the semantic condition which is necessary, for the inference from 'BaFb' to 'BaFc' to go through, is sufficient to warrant the conclusion that Ba(b = c).[8]

In general, the diagonal proposition defined on C(Fb) will not be the same as the proposition expressed by 'Fb' in any given world. That is, *C(Fb) ≠ C(Fb)(w), for arbitrary w, as a general rule. It differs intuitively in that what it corresponds to is, in effect, the proposition that 'Fb' is true.[9] This proposition is true at a world w just in case what 'b' refers to, in w, has

has the property that 'F' expresses, in w. An agent who believes this proposition can be expected to assent to 'Fb' under normal circumstances, since what is believed is that the linguistic items 'F' and 'b' are so related to the world and each other as to make 'Fb' true.

In cases where a sentence P expresses the same proposition relative to each world in the domain of the propositional concept C(P), the diagonal proposition *C(P) will be the same as the proposition which is the value of C(P) at any world w. In such cases (if they exist), the distinction between de dicto and de re belief collapses, and an agent has the one if and only if he has the other. And this, I think, is the result one should want for these cases, where there is no uncertainty about what the reference of the relevant linguistic items might be.

When should belief attributions be understood de dicto, and when de re? In the case of belief attributions involving singular statements, I think they should be, and generally are, understood de re, except when it is distinctly problematic to do so. To my mind, this amounts to no more than taking them at face value wherever possible.

Consider Oedipus. Presumably he believed that it was o.k. to marry Jocasta. Since Jocasta was his mother, if his belief was a de re belief, then he also had the (de re) belief that it was o.k. to marry his mother. But this he doubtless would have disputed. He would have insisted that he believed that it was not o.k. to marry his mother. The problem is avoided if we attribute to him only _de dicto_ beliefs, so that his belief that it was o.k. to marry Jocasta would not imply a belief that it was o.k. to marry his mother. But that just doesn't seem right. Surely his belief was about Jocasta, that it was o.k. to marry _her_; i.e., his belief was _de re_. This commits us to the view that he also had the _de re_ belief that it was o.k. to marry his mother.

But nothing in these attributions now prevents us from attributing to him as well the _de dicto_ belief that it was _not_ o.k. to marry his mother. And nothing in these attributions implies that he held inconsistent beliefs, if _de dicto_ and _de re_ beliefs are understood as outlined above. For these beliefs to be consistent, the propositions which are their objects must have at least one possible world in common. And this can easily be the case, provided Oedipus did not have the _de dicto_ belief that Jocasta was his mother. Which he didn't.

Typically, verbal reports are a reliable indicator of _de dicto_ belief. For _de re_ beliefs, non-verbal behaviour is often a better guide. That Oedipus married his mother is good reason to think that he believed (_de re_) that it was o.k. to do so. But now we are forced to regard his belief that it was not o.k. to marry his mother, as a _de dicto_ belief, if we are not to attribute to him plainly contradictory beliefs.

To make the example clearer, let w_o, w_1, and w_2, be the only relevant possible worlds, and suppose w_o is the actual world. Let there be two objects, x and y, in the universe of each world. Relative to these worlds and objects we can interpret a predicate letter 'F' and individual constants 'b' and 'c' as follows: let 'F' be interpreted as the property which holds of x in w_o and w_1, and holds of nothing in w_2. The interpretation will be indicated diagrammatically like this

	w_o	w_1	w_2
F:	{x}	{x}	Λ .

Let 'b' be interpreted as rigidly designating x. That is,

	w_o	w_1	w_2
b:	x	x	x .

And let 'c' also be interpreted as rigidly designating x. I.e.,

	w_o	w_1	w_2
c:	x	x	x .

These interpretations of 'F', 'b', and 'c', determine the following interpretations for the sentences 'Fb', 'Fc', '~Fc', and 'b = c':

	w_o	w_1	w_2
Fb:	T	T	F

	w_o	w_1	w_2
Fc:	T	T	F

	w_o	w_1	w_2
b = c:	T	T	T

	w_o	w_1	w_2
$\sim$Fc:	F	F	T

Associated with each sentence above is the proposition expressed by the sentence in the actual world, i.e., in w_o. But now we want to consider other possibilities, in particular, epistemic possibilities according to an agent who is unclear about what the term 'c' refers to. For such an agent, we may imagine that there is a possible world, say w_1, where 'c' rigidly designates y. That is,

	w_o	w_1	w_2
$c(w_1)$:	y	y	y

We can suppose, for simplicity, that this is the only epistemic alternative which differs from the actual world in point of interpretation. In other words, letting $i_{w_k}(\Phi)$ stand for the interpretation of 'Φ' in w_k, we suppose that $i_{w_o}(F) = i_{w_1}(F) = i_{w_2}(F)$. We suppose also that $i_{w_o}(b) = i_{w_1}(b) = i_{w_2}(b)$, and that $i_{w_o}(c) = i_{w_2}(c)$. $i_{w_1}(c)$ is what was just given above.

Relative to these generalized interpretations of 'F', 'b', and 'c', the following propositional concepts are determined for 'Fb', 'Fc', 'b = c', and '$\sim$Fc':

Fb:

w_o	w_1	w_2
T	T	F
T	T	F
T	T	F

Fc:

w_o	w_1	w_2
T	T	F
F	F	F
T	T	F

b = c:

w_o	w_1	w_2
T	T	T
F	F	F
T	T	T

∿Fc:

w_o	w_1	w_2
F	F	T
T	T	T
F	F	T

Now suppose we think of 'F__' as standing for the English predicate 'It is o.k. to marry __', 'b' as standing for 'Jocasta', and 'c' as standing for 'the mother of Oedipus'. It can be seen that in the actual world (w_o) the proposition expressed by 'Fb' is the same as that expressed by 'Fc'. This proposition is the object of Oedipus' de re belief that b is F, and since 'b' and 'c' rigidly denote the same thing, it is the same as the de re belief that c is F-- it is the belief that the actual world is either w_o or w_1.

The de dicto belief that ∿Fc has as its object the diagonal proposition of the propositional concept associated with '∿Fc', i.e., *C(∿Fc). (Diagonal propositions are indicated by the diagonal dotted lines in the diagrams.) And this proposition has world w_1 in common with the proposition which is the object of the de re belief that c is F. So these two beliefs are compatible on the view I've been describing.

But once the agent (Oedipus, in the example) acquires the de dicto belief that b = c, i.e., believes the proposition *C(b = c), something has to give. According to Sophocles, it was the de re belief that b is F which got dropped, but we can easily imagine things turning out otherwise-- Oedipus might have refused to believe (de dicto) that Jocasta was his mother, or he might have decided (de dicto) that in special cases it's o.k. to marry one's mother. I don't think there are any general principles according to which de re beliefs have to yield to conflicting de dicto beliefs, or conversely.

In the simplified model given above, the proposition that is expressed by 'Fb' is the same in each world in the domain of the propositional concept C(Fb) (represented by the vertical axis, in the diagram). Hence, *C(Fb) is the same proposition as i_{w_o}(Fb), and so, the de re belief that b is F and the de dicto belief that b is F, are the same. This feature, I think, properly reflects the fact that Oedipus (whose beliefs these are) is familiar not only with the person Jocasta, but also with the use of the term 'Jocasta' to refer to her. This is not the case with the term, 'the mother of Oedipus', and accordingly, the de re belief that c is F and the de dicto belief that c is F, are not the same.

It is important to note that a world such as w_1 is not an impossible world where, for example, Jocasta is not Oedipus' mother, but a possible world where the terms 'Jocasta' and 'the mother of

Oedipus' name distinct entities, and where they rigidly name these other entities. That such a world might be, for an agent, an epistemic alternative to the actual world, seems to me to be very much in keeping with Kripke's account of names-- it is precisely because a name carries with it no clue as to the identity of its referent, that an agent's de dicto beliefs involving the name may depend on such possibilities.

I have outlined a theory of de dicto and de re belief which permits in the latter, and prohibits in the former, inferences based on "substitution of identicals". It is, moreover, a theory in which names function as rigid designators. And I have argued, largely by example, that it is an intuitively plausible one. It remains to apply the theory to the somewhat different examples Kripke uses to generate his "puzzle". If the puzzle continues to appear intractable, it will surely count against the theory. On the other hand, the theory might lead us to re-examine the intuitions on which the puzzle is built. So let us now try casting Kripke's examples in terms of the approach described above.

Pierre is our agent, and he assents to the statements, 'London is not pretty', and 'Londres est jolie'. On the basis of Disq we say that he believes (de dicto) that London is not pretty, and also that he believes (de dicto) that Londres est jolie.[10]

Next Trans, applied to the latter belief statement, yields the belief that London is pretty. The object of his "English belief" (that is, the belief that is signalled by his assent to the English sentence), is $*C(\text{London is not pretty})$. What about his "French belief", that London is pretty?

We have here two alternatives. We could say that since he believes the proposition $*C(\text{London is not pretty})$, and since $*C(\text{London is not pretty}) \cap *C(\text{London is pretty}) = \Lambda$, that he therefore does not believe that London is pretty, and hence, we reject Trans as a legitimate principle for de dicto belief. Or, we could accept Trans, and argue that, since Pierre doesn't realize that 'London' and 'Londres' name the same city, $*C(\text{London is not pretty}) \cap *C(\text{London is pretty}) \neq \Lambda$.

There is something right about each alternative, I think. Pierre's "English belief" and his "French belief" clearly do not have, as their respective objects, complementary propositions. No agent, on the view I am advocating, believes the actual world to be a member of the empty set of possible worlds.

However, if we choose to accept Trans, we must also either accept as a general principle

$$BaP \; \& \; Ba{\sim}P \rightarrow Ba(P \; \& \; {\sim}P),$$

carefully explaining that when the antecedent is satisfied, $*C(P \; \& \; {\sim}P) \neq \Lambda$; or we must deny it, carefully explaining that in those cases where the antecedent is satisfied, and the consequent is not, $*C(P) \cap *C({\sim}P) \neq *C(P \; \& \; {\sim}P)$, i.e., that we do not have on the part of our agent, a gross failure to draw inferences.

It is easier simply to reject Trans. In any case, the three possibilities are equivalent, in that they each interpret the objects of Pierre's "English" and "French" beliefs as compatible propositions.

Rejecting Trans merely spares us the drudgery of continually explaining away seeming failures of rationality.

And what is more, it does so at no real expense to our intuitions. When we reflect on the details of the example, it becomes increasingly clear, I believe, that Trans is not a reasonable principle to apply. Pierre is in the altogether strange position of being able to speak fluently both French and English, and yet is unable to translate from one language to the other (at the very least, he is unable to translate sentences containing 'London' or 'Londres'). Presumably, then, he would not have assented to 'London is pretty', on the basis of his belief that <u>Londres est jolie</u>, even if he had never acquired his "English belief" that London is not pretty.

This alone should give us cause to pause. If an agent readily assents to one statement, and refuses to assent to its standard translation into another language he knows, it is surely grounds for supposing that the statement and its translation are not, in his view, equivalent. If we further accept the strong principle of disquotation (henceforth, Strong Disq), that is, Disq together with the converse claim that an agent who believes a statement will, <u>caeteris paribus</u>, be disposed to assent to it, then failure to assent to a statement implies lack of belief in it. Hence, Pierre's failure to assent to 'London is pretty' implies that he does <u>not</u> believe that London is pretty, and thus implies the negation of Trans. Briefly,

Strong Disq → ∿Trans.

There is, in other words, a basic incompatibility between Strong Disq and Trans, and even between simple Disq and Trans, if we also accept as general principles

$$BaP \,\&\, BaQ \rightarrow Ba(P \,\&\, Q),$$

and

$$\sim Ba(P \,\&\, \sim P).$$

And this really is the puzzle: something has to go, but what?

I have urged that Trans be rejected, and have considered as well the possibility of denying one or the other of the two general principles just mentioned. I have not questioned Disq. The reason for this lies, naturally, in the characterization of de dicto belief as belief in the diagonal proposition (associated with a statement). For the diagonal proposition associated with a statement P corresponds, in a certain sense, to the proposition that 'P' is true,[11] and so, it is natural to think of de dicto belief, and disposition to assent, as intimately related; related in a way that is well represented by Strong Disq. Moreover, Trans is in essence a kind of substitution principle, and on this account, substitutivity does not generally hold in de dicto belief contexts. In short, it is a basic feature of this model that it yields a definite response to the puzzle: Pierre believes that London is not pretty, and does not believe that London is pretty.

The propositional concepts employed earlier to characterize Oedipus' beliefs will do nicely for Pierre, also. Let 'b' stand for 'London', 'c' for 'Londres', and 'F__' for '__ is not pretty'. (Thus, '∼F__' stands for '__ is pretty', or equivalently, for simplicity, '__ est jolie'.) Then Pierre's de dicto belief that London is not pretty, is the belief whose object is *C(Fb). His de dicto belief that Londres est jolie has as its object *C(∿Fc). These propositions are jointly true in w_1. But w_1 is a world where *C(b = c) is false, and so the inference from BaFb to BaFc, or from Ba∿Fc to Ba∿Fb, is not justified.

The challenge of the paradox, again, was to "formulate an acceptable set of principles that does not lead to paradox, is intuitively sound, and supports the inferences we usually make". The above account eliminates the paradox, and is, I have argued, intuitively sound. Does it support the inferences we usually make? In particular, how is one to respond to Kripke's argument in favour of Trans, to the effect that if Pierre doesn't believe that London is pretty, no monolingual Frenchman ever did?[12]

Trans, it seems to me, owes its initial plausibility to the fact that, applied to de re belief, it is virtually incontestable. If a monolingual Frenchman believes, of Londres, that it is jolie, then obviously he believes, of London, that it is pretty. But why suppose for a moment that an analagous relation should hold between de dicto beliefs? The monolingual Frenchman who has the de dicto belief that Londres est jolie, has, in a certain sense, the belief that a certain statement is true. Such facts do not support any inferences to beliefs about the truth of statements in a language he does not know, nor to beliefs about the truth of statements in a language he does know, if he is ignorant of the translation relation between the languages.

In other words, the argument appears to issue from failing to keep clearly in mind the distinction between de dicto and de re belief. It is true that we are usually prepared to make inferences based on Trans, but this is due to the natural tendency, noted earlier, to construe belief attributions de re whenever possible. Thus, it appears to me, the present approach does countenance the inferences we would make on reflection, and excludes only those that are, on reflection, questionable.

I have focussed on the central example of Kripke's paper, the example concerning Pierre and his beliefs about London. The other examples he discusses are comparable, in spite of their ideosyncrasies, and, I submit, comparable treatment is appropriate and effective. If so, what does this leave us with? It leaves us first of all with the conclusion that de dicto belief contexts are not "Shakespearean", that is, that co-referential proper names cannot freely be substituted in such contexts. This conclusion is at odds with Kripke's final remarks.

However, the model used to characterize de dicto belief gives a precise account of when inferences based on such substitutions are warranted, an account which reflects exactly our intuition that it is when, and only when, an agent believes two names to be co-referential, that the inferences are justified. The model further provides the framework for an analysis of belief statements, in terms of which the various principles, that jointly produce the "paradoxes", can be assessed, and an optimal consistent set of principles chosen. And it does this, all the while conforming to the desideratum that names be treated as rigid designators. That is, it gives a principled basis for denying that <u>de dicto</u> belief contexts are "Shakespearean", without falling back on a description theory of names. It is the right tool for the job.

ENDNOTE

I have claimed that the various alternative ways of dealing with the puzzle are equivalent, in that they each take the objects of Pierre's beliefs to be compatible propositions. In another sense they are plainly inequivalent-- which principles one accepts will determine which belief statements one endorses, and which ones one withholds. This inequivalence is obscured in Kripke's example, since it is Pierre's "French belief" that is intuitively dubious. Those who are inclined to restrict the principle of disquotation, so that it is applicable only in cases where the correct reference of terms is somehow guaranteed, will see Pierre's assent to the French sentence as not meeting the conditions, that is, as not being a candidate for disquotation. And those who reject the principle of translation, will not translate the French belief statement. In either case, it is Pierre's "English belief" alone that gets certified.

But imagine the story changed so that it is assent to the English statement that fails to qualify for disquotation, if either does. Say Pierre is a monolingual Frenchman who takes a tour of England with a bunch of other monolingual francophones, visits London, finds it unlovely, and consequently comes to assent to 'Londres n'est pas jolie'. Back in France, he decides to study English, becomes fluent, and as a result of his indiscriminate reading, comes to assent to 'London is pretty'. As before, he doesn't realize that London and Londres are one city. Now, those who reject (as I will say, meaning, reject or restrict) the principle of disquotation, and accept Translation at face value, will say that Pierre believes that London is not pretty; and those who reject (or restrict) Translation, but accept Disquotation, will say that Pierre believes that London is pretty. And each will withhold the contrary belief attribution. Letting 'P' stand for 'London is pretty', and sticking for the moment with our modified version of Pierre's sayings and doings, we may summarize the relationship between principles and statements as follows:

	principle	belief statement implied
1)	Disq	BaP
2)	Trans	$Ba{\sim}P$
3)	Closure under conjunction	$BaP \& Ba{\sim}P \rightarrow Ba(P \& {\sim}P)$
4)	Non-contradiction	${\sim}Ba(P \& {\sim}P)$.

The above principles form an inconsistent tetrad. To restore consistency, it is not necessary to deny (or restrict) either Disq or Trans. David Lewis, for example, seems to accept both, and thus is obliged to reject 3) or 4) instead.[13] Although there is nothing logically wrong with such an approach, 3) and 4) look a lot like rationality conditions on belief, whereas Disq and Trans do not.

So many will prefer to withhold one or the other of the belief attributions, and thereby avoid the need for telling a long story, like Lewis', to account for the apparently inconsistent attributions.

The difficulty can be avoided by placing strong restrictions on the principle of disquotation, and this is the tack some seem to prefer when confronted with the puzzle.[14] In effect, the application of Disq is confined to cases where the subject of the attitude can be thought of as having a proper understanding of the reference of singular terms in the sentences he assents to. In other words, Disq is applicable only in cases where the resulting belief statement can be construed de re.

Now a universal principle of disquotation is certainly inappropriate if the resulting belief statements are going to be construed de re (as was remarked at the outset of this Chapter). But the puzzle, as Kripke emphasizes, is a puzzle for de dicto belief contexts (indeed, the puzzle can be regarded as a kind of test for whether one takes de dicto construals of belief attributions seriously). Thus, to restrict Disquotation to those cases where the resulting belief statements can plausibly be construed de re is not to address the puzzle at all.

The tension between Disq and Trans has two sources, already noted. The first is their joint inconsistency in the presence of 3) and 4). The second is that Disq is naturally appropriate to de dicto construals of belief statements, and Trans to de re. One can enforce uniformity in this regard by confining Disq to de re contexts, that is, by refusing to apply it in the absence of referential guarantees on the singular terms in the sentences to which assent is given. I have claimed that this does not address the problem, as conceived by Kripke.

Instead, we need a way of eliminating inconsistency, while devising a set of principles uniformly applicable to de dicto belief contexts. This can be done by rejecting Trans outright as a principle appropriate to belief statements construed de dicto, or less radically, by adapting it to de dicto contexts, through the imposition of constraints on the translation of belief statements, constraints which sometimes will yield non-standard translations of the embedded clauses.

An obvious device is to translate the embedded clauses into a semantic meta-language. Thus, 'Pierre croit que Londres n'est pas jolie' might be translated as 'Pierre beleives that the city he calls 'Londres' is not pretty' (ignoring for simplicity the problem of translating items other than singular terms).[15] One can obviously attribute to Pierre as well the belief that London is pretty, without thereby convicting him of inconsistency.

Having indicated a preference for rejecting or restricting Trans, rather than Disq (for the case of belief statements construed de dicto, and where inconsistent belief attributions would otherwise result) I must now point out (if it has escaped notice) that this approach seems to yield grossly counterintuitive results when applied to my own version of Pierre's history. For instead of saying, 'Pierre believes that London is pretty', and, 'Pierre believes that the city he calls 'Londres' is not', shouldn't we be saying much the opposite?

Doesn't Pierre *actually* believe that London is not pretty, and that some city he calls 'London' is?

The answer would be yes, *if* we were construing the belief statements *de re*. But we are not, and since we are not, since we are concerned with a construal of belief statements for which referentiality (of singular terms) may fail, the fact that Pierre's "French belief" is really *about* London, and his "English belief" not, is beside the point. Pierre believes *de dicto* that London is pretty, and we must not naysay him.

FOOTNOTES

1. Kripke (1979).

2. *ibid.*, page 259.

3. In the discussion below I confine my attention to examples like Kripke's, i.e., to belief statements whose complement clauses are singular statements containing (what are presumed to be) rigid designators. Other devices of singular reference introduce complications which are not immediately to the point, and will be ignored in this chapter.

4. Stalnaker (1976), and Stalnaker (1978). I do not mean to suggest that Stalnaker would endorse the treatment of *de dicto* and *de re* belief that follows.

5. Stalnaker (1976), page 87.

6. *ibid.*, page 88. In Stalnaker (1984), he backs away from this suggestion, for mathematical statements.

7. Hartry Field (1978), argues that the suggestion cannot work, since, for an agent who knows the semantic rules for the language of

set theory (for example), the statement asserting that those rules relate the Banach-Tarski conditional (for example) to the necessary truth, itself expresses a necessary truth, and so, on Stalnaker's approach, cannot be doubted. This argument (which appears on pages 14 - 15) is, I think, not so much a refutation of Stalnaker's view as a *reductio* of the claim that an agent might both doubt the truth of a true mathematical statement, and also know the relevant semantic rules, in the strong sense of "know" that Field presupposes.

In the case of ordinary statements involving proper names, it is obvious that if knowing the semantic rules implies knowing the reference relation, then one can fail to know the semantic rules, and still be a normal and competent speaker of the language.

8. Here (and perhaps elsewhere), the term 'semantic' is used somewhat broadly, since the propositional concepts, on which diagonal propositions are defined, depend not just on the actual semantic properties of expressions, but also on how the agent is epistemically situated.

9. What the qualification "in effect" comes to here is spelled out explicitly in Chapter V.

A remark is in order about my use of underlining. I underline phrases lifted bodily from other languages, in the usual manner. Thus, for example, "Pierre believes that *Londres est jolie*". In addition, however, I use locutions of the form '*that* Φ' as a kind of canonical name, in the metalanguage in which I write, of a proposition.

Thus, for example, "Pierre believes the proposition that London is pretty".

10. The dubious linguistic pedigree of this locution has troubled some earlier readers, who seem to feel that a defective sketch of the puzzle is bound to result. However, the translation of 'Pierre croit que ...' as 'Pierre believes that ...' is hors de propos. What is at issue is the propriety of translating what occurs within the belief context.

I have also heard it claimed, as if self-evident, that such locutions, being mere "word salad", are not part of any language at all. This strikes me as a very odd claim. I have heard "word salad" spoken, evidently to the same purposes, and with the same success, as any more familiar natural language. Indeed, English itself, considered historically, is a word salad. Kripke, in discussing word salad (Kripke (1979)), does not explicitly claim that such constructions are not part of any language. He does, however, suggest that they are not governed by linguistic rules (see footnote 10 to "Puzzle"). This too is not correct. Agreement with respect to tense, number, gender, etc., is maintained, transitive verbs take direct objects, and so forth. Linguistic rules do not, of course, determine whether, in a given context, items from one "constituent" language rather than another are used-- but then, linguistic rules do not determine word choice in any language. Other factors do that.

11. Again, this needs to be qualified. See Chapter V.

12. Kripke (1979), page 256.

13. Lewis (1981).

14. Marcus (1981), and Marcus (1983); and John Biro, in a paper read at the 1982 meetings of the Society for Philosophy and Psychology, a version of which appears as Biro (1984).

15. Note that in translating 'Londres' as 'the city he calls 'Londres'', we may in some instances be assuming too much. Perhaps Pierre, obtuse in spite of his logical acuity, has taken too literally a poetic apostrophe to London ("Londres, il se dit que tu n'est pas jolie, et je le croit ..."), and believes that 'Londres' is the name of a woman.

POSSIBLE WORLDS SEMANTICS AND PROPOSITIONAL ATTITUDES

I want now to discuss in a somewhat more general way the semantic device introduced informally in the first chapter, which was there claimed to have natural applications in the interpretation of propositional attitude discourse. First of all, I want to give a rough "map of the terrain", within which to locate the issues I want to address, and to provide a particular perspective on them. I will mention what appear to me to be some significant advantages to the use of this device, and also considerations which suggest that any adequate treatment of propositional attitudes, within the model-theoretic tradition, will have to employ this device, or something essentially equivalent.

The device, to repeat, is a certain generalization of the familiar possible worlds semantics for modal discourse, according to which sentences in a language are thought of as expressing propositions, which are identified with sets of possible worlds (intuitively, those worlds where the sentence in question is true), or with functions from possible worlds to truth values, assigning the value "true" to those worlds where the sentence is true. The question, then, is whether or not propositions in this technical sense can plausibly be regarded as the objects of propositional attitudes, such as belief.

There are various reservations one might have about the use of possible worlds. At one extreme is Quine, who, famously, regards the very notion of "possible worlds" as hopelessly obscure. Others do not, but may, like Storrs McCall, insist on rigourous admission requirements for any candidate worlds. For present purposes, I want to put aside all such scruples about possible worlds. For they do not bear directly on the central issues in the semantics of propositional attitudes; namely, the identity and individuation of the propositional objects of the attitudes, and the closure conditions on the attitudes.[1] What makes the possible worlds approach controversial in this application is that it carries with it a particular determination of these latter issues, a distinctive "taxonomy" of the objects of belief, for example; a taxonomy which is often felt to be inappropriate to the task.

The question of the identity and individuation of propositions can be posed as follows: which pairs of sentences (of a language L) are to be regarded as expressing a single proposition, and which pairs as expressing distinct propositions? Here is a possible (partial) answer to the question: there are exactly two distinct propositions, and every (closed) sentence of L expresses one or the other of them (these propositions might be thought of as the "True" and the "False").

At the other extreme, we could stipulate that no two distinct sentences of L correspond to the same proposition, that each sentence expresses a different proposition from every other sentence. In this case, propositions might simply be surrogates for sentences, preserving all (or nearly all) of the structure of the language L, whereas the two-element algebra of propositions mentioned a moment ago collapses almost all of that structure.[2]

There are, of course, intermediate possibilities. Something equivalent to the Lindenbaum algebra results from mapping all (and only) logically equivalent sentences on to a single proposition-- this algebra preserves the structure of L "up to logical equivalence". Consideration of relevance relations among the sentences of L will yield notions of a proposition that preserve more of the structure of L than does the Lindenbaum algebra-- some pairs of logically equivalent sentences (but not others) will get mapped on to distinct propositions. The point of all this is that there is a field of alternatives available, differing from one another in the amount of structure invested in the algebra of propositions; or equivalently, in how "fine-grained" their propositions are; or again, in how closely (and in what respects) propositions are taken to resemble the sentences that express them.

So the problem is where, in this field of alternatives, to locate our concept of propositions, with a view to treating them as the objects of propositional attitudes. Propositions, in the sense of possible worlds semantics, come with a built-in algebraic structure much like that of the Lindenbaum algebra (the exact nature of this structure depends, as the Lindenbaum algebra does not, on the details of interpretation of the non-logical constants of L). And so their use involves commitments, philosophically contentious ones, concerning the identity and individuation of propositions.

If we conceive the problem in the manner laid out above, then the most familiar arguments against the possible worlds approach can be seen to rest on intuitions exerting a "pull" on the characterization of propositions, in the direction of an algebra that maximally reflects the structure of L. In defence of the possible worlds approach, I will raise some difficulties with these arguments; but ultimately I want to contend that there are fundamental, pre-analytic intuitions concerning belief and meaning, that "pull" in opposite directions, and thus that a major job for any semantic account of belief is to come to terms with these opposing forces. This job requires a special flexibility lacking in most accounts of propositions and belief.

To begin with, let's consider some common objections to the possible worlds approach. Belief in a proposition, on this approach, is characterized intuitively as belief that the actual world is one of the worlds where the proposition takes the value "true". It follows immediately that belief is closed under logical consequence, for logical consequence is represented in the algebra of propositions as set-theoretic inclusion. But set theory is extensional-- hence, it is claimed, sets of possible worlds are going to be "too extensional" to be the objects of belief.

It is also claimed that the deductive closure of belief, in possible worlds semantics, makes agents "logically omniscient". Most objections focus on specific apparent examples of the failure of belief to be closed under various extensionally valid, and even, in fact, intensionally valid, forms of inference. All of these are really versions of the same objection. And the conclusion drawn is that, whatever propositions are, they must reflect more of the structure of natural language than the possible worlds approach allows.

In particular, it must be possible for sentences differing only in occurrences of co-referential proper names to express different propositions (and this must be the case even if names are regarded as rigid designators; that is, even if names have the same intension, and not just the same extension). Likewise in the case of co-intensive predicates, or necessarily equivalent sub-sentences. Call a linguistic context in which substitution even of intensionally equivalent items fails, a hyperintensional context.[3] The claim, then, is that belief contexts are hyperintensional, and that this feature refutes the possible worlds approach.

Well, let's see where this suggestion takes us. Let us suppose, that is, that logically equivalent statements, as well as (non-indexical) statements differing only in occurrences of co-referential names, or co-intensive predicates, may express different propositions. Note, however, that if they do express different propositions, they do so necessarily, in the sense that the propositions they express are fixed, and remain the same no matter in what context the statement occurs.[4] This seems to me to involve a serious distortion of the very phenomena of natural language to which such accounts are meant to be sensitive. The distortion arises not simply from having the propositions expressed by (non-indexical) sentences be invariant with respect to the contexts in which they occur, but from that feature in combination with the supposed hyperintensionality of belief contexts.

Let's consider an example. Suppose 'Jill' and 'Sue' are co-referential names, which are not, however, substitutable in belief contexts. In particular, say 'Jill walks' and 'Sue walks' express distinct propositions, and so an agent might believe that Jill walks without believing that Sue walks, or vice versa.

The trouble is that we apparently have no way of inferring that an agent who believes that Jill walks, and also that Jill is Sue, therefore believes that Sue walks. We might imagine that this agent knows who Jill is as well as any mortal could, and further, that 'Jill' and 'Sue' are, in the agent's ideolect, as synonymous as any pair of natural language items could be. For such an agent, it is tempting to say that, to believe that Jill walks is to believe that Sue walks. But on the present approach we can't say that. If 'Jill walks' and 'Sue walks' express different propositions at all, they do so for everyone. To invert a familiar argument, the semantics of belief should not entail that agents are logically obtuse.

A (deceptively) simple partial remedy to this problem would be to introduce a meaning postulate stating that agents' beliefs are closed under substitution of names believed to be co-referential. Similar provisions would have to be made for expressions of other types, and for other sorts of inference.

In other words, once we have settled on a determinate algebra of propositions and a unique mapping from sentences of L to that algebra, together such as to allow failure of certain inferences in belief contexts, some means will have to be found of accomodating those particular cases in which the inferences plainly ought to hold. Conditional meaning postulates might be proposed to lend the needed additional flexibility to the theory. But the move is not really adequate. In the first place, it seems wrong to attribute to the agent of the previous paragraph <u>two</u> beliefs, rather than saying that either of two belief statements could equally well be used to indicate his belief that Jill (Sue) walks.

A more important difficulty has to do with the underlying motivation for a hyperintensional treatment of belief. It is in deference to our intuition that an agent, who does not know that Jill is Sue, might believe that Jill walks and not believe that Sue walks, that the embedded statements are treated as expressing distinct propositions (and likewise with respect to other pairs of equivalent, logically equivalent, or necessarily equivalent statements).

But for any two statements, however they are logically related, one can always find an (hypothetical) agent who will assent (sincerely, reflectively, etc.) to one and not the other. In addition, there is sure to exist, among the agents who assent to a given statement, considerable disagreement as to the import, presuppositions, consequences, and so forth, of the statement. In short, what I call the testimonial evidence suggests that no two statements express the same proposition, and further, that, in some sense, sentences can mean different things to different folks.[5]

There are serious difficulties with the view that every sentence expresses a unique proposition. For one thing, the problem mentioned above of the closure of belief under various sorts of inference, becomes intractable. That is, suppose the algebra of propositions is isomorphic to the "algebra" of sentences. Then in order to recover even the most trivial sorts of closure of belief (e.g., the sort that would allow us to infer Ba(P & Q) from Ba(Q & P)), the algebra will have to be augmented with explicit closure principles, in the form of "meaning postulates" governing belief statements.[6] But what set of meaning postulates could yield the right sort of closure of belief for all agents? None, plainly; the testimonial evidence is just too plastic. And this robs the account of belief of any generality or naturalness.

Likewise, the view that a sentence may semantically express different propositions, depending on how the agent is epistemically situated, is presumably unacceptable (where non-indexical sentences are concerned).

To deny either of these unattractive views is to deny that the testimonial evidence of agents is sufficient, or necessary, to determine the identity or distinctness of propositions. In other words, a decision to treat the algebra of propositions as any less "fine-grained" than the "algebra" of sentences, is a decision to cease, at some point, respecting the testimonial evidence of agents. At whatever point this occurs, it will become necessary to explain away the failure of belief statements to behave, logically, in accordance with the demands of the testimonial evidence.

Such explanations will be precisely of the kind one gives in support of _de re_ construals of belief statements. One says, for example, that John, who believes that Hesperus is pretty, _therefore_ believes that Phosphorus is pretty, and his protestations to the contrary are due to an (admittedly minor) confusion or ignorance about the name 'Phosphorus'. Now, if there is going to be _some_ identification of the propositions expressed by different sentences, and if the direct testimony of agents is discounted in virtue of "_de re_-like" considerations, what reason is there not to regard any two necessarily equivalent sentences as expressing one proposition?

One might anticipate two sorts of response to this question, one having to do with the logico-linguistic talents that agents might or might not possess, and the other in terms of independently specifiable features of the sentences themselves. One could say, for example, that the assent or dissent of _competent_ speakers (a notion that would have to be made precise) suffices to identify and individuate the propositional objects of belief. Or one might propose that the degree of complexity of sentences, or relevance relations among them, or some such, also helps determine propositional identity, or lack of it. I'm not going to examine these matters here.

For the moment, simply note that the "evidence" appealed to in favour of a general hyperintensional treatment of belief is not compelling, and that there is a serious tension between the desire to accomodate the differing epistemic circumstances of different agents (i.e., by deferring to their testimony), and the desire to keep the content of sentences out in the open, as it were, public, fixed for all.

Let me lead into a sketch of the approach I want to advocate, by giving it a quasi-historical context. Extensional logic may be thought of as taking a point-valued semantics; that is, formulae are interpreted by primitive "0-dimensional" items-- the truth values 0 and 1, for example. Early attempts to give formal semantic substance to already well articulated notions of modality proceeded by expanding the class of values. The use of matrices (truth tables) involving more than two values permits the characterization of non-classical connectives and operators. However, it turns out that no finite expansion of the domain of truth values can characterize properly the "ordinary" modal operators.[7]

What finally did do the trick was classes of truth values indexed by possible worlds. These may conveniently be thought of as a one-dimensional extension of the notion of a truth value-- in place of a single "point" value, 0 or 1, we now have an infinite sequence of such values. In other words, where simply piling on more (point-) values to interpret formulae didn't work, adding a new dimension to the concept of interpretation did. In retrospect, it seems just what intensional contexts require, in that it allows the content of a formula to vary differentially with respect to its truth value.

In his papers "Pragmatics" and "Pragmatics and Intensional Logic", Montague elaborated on the sort of structure the indices would need to have in order to accomodate context sensitive elements of natural language, like tenses and personal pronouns.[8] He expanded the notion of an index to include not just possible worlds, but also moments of time, speakers, objects ostended, and so forth. An apparent shortcoming of his treatment, pointed out by both Stalnaker and Kaplan, is that statements like 'I am here now' come out valid, and that on Montague's semantics, a valid statement is necessary as well.[9]

What is clearly required is a way of letting some aspects of the interpretation of the statement vary independently of others, so that the obvious contingency of the proposition that I am here now can be secured. In particular, possible worlds must be allowed to vary independently of the other elements of indices. At the same time, we don't want to forfeit the intuitive "validity" of the statement-- the intuitive fact, that is, that the statement will be true, at any occasion of tokening, with respect to the index of that tokening (roughly, true whenever uttered). So the non-modal aspects of interpretation should be kept appropriately fixed together (to indices). In short, a second (pragmatic) "dimension" to interpretation is needed. This is just what is provided by systems like Kaplan's Logic of Demonstratives.[10]

Kaplan's system does much to alleviate the problems raised for a one-dimensional intensional logic by natural language indexicals. Similar problems nonetheless persist. These derive from the apparent feature of natural language, that some locutions are "more intensional" than others. Familiar examples of inferences involving modal contexts, that Quine seems to have taken virtually to reduce modal logic to absurdity, have pretty well lost their sting, largely as a result of Kripke's arguments concerning the nature of proper names. Other such examples, formally very similar, continue to trouble many philosophers.

The difficulty with a hyperintensional treatment of belief in a one-dimensional semantics lies in the attempt to accomodate apparently conflicting intuitions— in seeking a middle ground, it sacrifices something of each. The conflicting intuitions are that the content of a statement does not depend on what particular agents know or believe, and that what belief should be attributed to an agent who assents to the statement, does.

Such problem cases as, for example, the apparent non-substitutivity of proper names in belief contexts, seem to call for a way of allowing the propositional objects of belief to vary differentially with respect to the propositions that interpret statements outside of belief contexts. That is, these cases seem to call for a two-dimensional extension of ordinary intensional logic.

Now, how is a second dimension to interpretation to be understood in the case of belief statements?[11] Very much by analogy to the case of indexicals, only the relevant features of context will not be physical locations, in space and time, nor whether the agent is speaker, interlocutor, or something else to the discourse. Rather, it is how the agent is epistemically situated that is pertinent.

What needs to be represented, in order to account for substitutivity failures in belief contexts, is agents' ignorance or doubt concerning the content of the terms in question. This can be done by associating with the terms, not just an intension (i.e., a function from possible worlds to extensions), as is done in ordinary one-dimensional modal logic, but a function from possible worlds to intensions. The worlds in the domain of such a function can be thought of as ways things could be, for all the agent knows. Then an agent's uncertainty about the content of certain expressions can be represented by a function that maps some of these worlds on to different intensions than others. Possible worlds thus figure twice in the interpretation of statements, first as determinants of possible contents, relative to an agent's epistemic situation, and then as determinants of truth.

These two roles will be represented by the axes of a two-dimensional array of extensions, as was done in Chapter I. For example, given a term t of a certain logical type, there will be a class of objects $\{a_1, \ldots, a_i, \ldots\}$ from which the extension of t is drawn. We will again imagine the class of possible worlds to be numerically ordered so that functions from worlds to extensions can be displayed as sequences of objects of the appropriate type, and we will take the actual world to be first in line. Then if the extension of t is a_o, its intension might be $\langle a_o, a_1, a_2, \ldots\rangle$, and a function from worlds to possible intensions for t might look like this:

	w_o	w_1	w_2	...
w_o	a_o	a_1	a_2	...
w_1	b_o	b_1	b_2	...
w_2	c_o	c_1	c_2	...
⋮		⋮		

Such a two-dimensional array associated with a term t will be denoted C(t).[12] The actual intension of t, then, is the top (horizontal) row of C(t), and it will be denoted @C(t). The other horizontal rows, that is, the $C(t)(w_i)$, for $i \neq 0$, are the intensions of t in worlds that the agent does not know to be non-actual. An agent's uncertainty about the content of a term t will be reflected in the fact that $C(t)(w_i) \neq C(t)(w_j)$, for some w_i, $w_j \in$ dom C(t).

Let us now consider an example brought by Thomason to show the inadequacy of possible worlds treatments of propositional attitudes.[13] All groundhogs are woodchucks. Indeed, we might say that this is necessarily so, since the terms 'groundhog' and 'woodchuck' are simply different names for the same critter. But, Thomason argues, the statement 'John is aware that all groundhogs are groundhogs' might be true, while 'John is aware that all groundhogs are woodchucks' is false; hence (supposing propositions to be the objects of propositional attitudes) the embedded statements must express different propositions, and hence, the proposition expressed by a statement is not adequately represented by the set of possible worlds in which it is true.

Now, in a sense, if John knows what groundhogs are, then he is aware that all groundhogs are woodchucks, even though 'woodchuck' may not be part of his vocabulary. Conversely, if John is not aware (in this sense) that all groundhogs are woodchucks, it is because he doesn't know what groundhogs (woodchucks) are. If he can nonetheless be said to be aware that all groundhogs are groundhogs, it must be in some other sense than the above. It is obvious what that sense is: what John is aware of is that the statement 'All groundhogs are groundhogs' (unlike 'All groundhogs are woodchucks') is a logical truth. Whatever property 'groundhog' expresses, the statement 'All groundhogs are groundhogs' will be true. But such is not the case for 'All groundhogs are woodchucks'.

To illustrate the situation, let us take properties to be functions from possible worlds to sets of objects-- those objects that have the property in the world in question. The property of being a groundhog, then, might be $\langle X, Y, Z, \ldots \rangle$, where capital letters from the end of the alphabet stand for sets. If so, then the same function is also the property of being a woodchuck. If x is an object in the domain of discourse, the proposition that x is a groundhog will be the sequence of values $\langle v_1, v_2, \ldots, v_k, \ldots \rangle$, where $v_i = 1$ if x is a member of the ith set in the sequence $\langle X, Y, Z, \ldots \rangle$, and $v_i = 0$ otherwise. The intensions of quantified statements are then determined in the obvious way.

Now suppose an agent (John, say) isn't quite clear as to what groundhogs are. For John, then, there are properties other than the property of (actually) being a groundhog, which are also candidates for the intension of 'groundhog'. That is, there are possible worlds where the intension of 'groundhog' is not the property of being a groundhog, which John is unable to distinguish from the actual world. These are not worlds in which, for example, groundhogs are marsupials, but rather, worlds in which 'groundhog' may express the property of being some marsupial.

Let us say that w_1 and w_2 are two such worlds, and that in w_1 'groundhog' expresses the property <Y, W, X, ...> and in w_2, <X, W, V, ...>. Then relative to John's epistemic situation, the following two-dimensional array is associated with the term 'groundhog':

	X	Y	Z	...
G:	Y	W	X	...
	X	W	V	...

Associated with 'woodchuck' is perhaps,

	X	Y	Z	...
W:	Z	V	V	...
	Y	X	W	...

Now consider the statements 'All groundhogs are woodchucks' and 'All groundhogs are groundhogs', and let 'P' and 'Q', respectively, stand for them. Then @C(P) = <1, 1, 1, ...> = @C(Q). So if 'John is aware that Q' means that the object of John's awareness is @C(Q), then John is also aware *that P*. But the sense in which John is aware *that Q*, and is not aware *that P*, can be captured on this account by noting that *C(Q) = <1, 1, 1, ...> ≠ <1, 0, 0, ...> = = *C(P) (assuming that V, W, X, Y, and Z are distinct sets). That is, when we incorporate into the model John's uncertainty as to what groundhogs (woodchucks) are, there is no problem accounting for the apparently discriminatory nature of his beliefs. Yet the objects of belief are still propositions, understood as sets of possible worlds.

It will be noticed that the above account involves distinguishing two propositions, @C(P) and *C(P), each associated with the statement 'P', and treats them as the objects of John's *de re* and *de dicto* attitudes, respectively. I do not wish to argue here that one and only one of these is *the* correct intension of 'P', nor to claim that statements in propositional attitude contexts are semantically ambiguous. For the moment I would simply like to remark that, if John can truly be said to be aware *that Q*, and unaware *that P* ('P' and 'Q' as above), it must be because the objects of his propositional attitudes, *as we describe them*, are the kind of thing that is sensitive to his epistemic situation, in a way that is well represented in the diagonal propositions *C(P) and *C(Q).

The above treatment can be generalized to apply to items of any grammatical category. Thus, substitution problems do not, as is so frequently claimed, refute the view that the objects of propositional attitudes are propositions, in the sense of possible worlds semantics.

Among the consequences that distinguish this approach from more familiar ("one-dimensional") ones, are the following. First, since the domain of a propositional (or other) concept is determined by the epistemic situation of an agent, the different circumstances in which different agents find themselves will often determine different concepts, with respect to a given statement (or other linguistic item). As a result, the diagonal propositions may differ, and so, two agents who have the "same" de dicto belief, in the sense that a single statement may commonly be used to attribute belief to each of them, can still have different propositional objects for their belief(s).[14] It is not plausible, I think, to suppose that all agents who "share" a (de dicto) belief, do so in virtue of believing exactly the same proposition. If they did, the only candidate would be the proposition that the statement in question is true. But this is a notoriously problematic suggestion, at once too weak and too strong. De dicto belief has to have a certain pragmatic plasticity to it.[15]

Second, under sufficiently favourable epistemic conditions, an agent's de dicto belief that P may coincide with the de re belief that P. This seems intuitively right, if we take the de dicto belief that P to be that belief which prompts assent to 'P'. We are thus able to maintain a strong form of monism regarding the objects of de dicto and de re belief, without the vexation of reducing one to the other. There is no reducing to be done.

Third, closure of belief under logical consequence has the right properties, without requiring special meaning postulates to guarantee them; that is, the monism of the theory does not compromise the inferential idiosyncrasies of de dicto and de re construals of belief attributions. This is important if, as I suggested earlier, no complete list of meaning postulates is possible which will provide the right closure conditions on (de dicto) belief for all hypothetical agents. On the present account, one believes the consequences of one's belief, where consequence is given the familiar set-theoretic characterization: a proposition B is a consequence of A iff $A \subseteq B$; or equivalently, if $B = \langle b_1, b_2, \ldots \rangle$ and $A = \langle a_1, a_2, \ldots \rangle$, and the a_i and b_i are all in $\{0, 1\}$, then for all i, $b_i \geq a_i$.

Of course, if one's de dicto belief that P differs from the de re belief that P, then the consequences of that belief will differ as well, and may or may not include particular propositions that follow from the proposition that P. A sufficient condition for an agent who believes de dicto that P, to believe also that Q, is for $*C(P) \leq *C(Q)$. This will be the case iff $*C(P \rightarrow Q) = \langle 1, 1, 1, \ldots \rangle$; that is iff the agent believes de dicto that Q is a consequence of P.

To sum up: I've argued that the usual objections to possible worlds semantics, in its applications to propositional attitudes, are founded on considerations which urge a de dicto-like construal of attributions of such attitudes as belief. These considerations turn out to make life difficult for the theory that tries to accomodate them within a "one-dimensional" framework. They can only be accomodated at the expense of equally fundamental, and contrary, intuitions. No happy medium is available. The attempt to find one results in a conception of the objects of belief which is appropriate neither to the de dicto nor the de re construals of belief statements. By forswearing the closure of belief under consequence that automatically accompanies the possible worlds conception of propositions, one contracts the task of explicitly stating the closure conditions for the full range of possible inference-- a problem which probably has no adequate general treatment.

None of these difficulties attach to the two-dimensional possible worlds approach. I conclude that it is to be preferred.

FOOTNOTES

1. This is not to suggest that ontological concerns are wholly irrelevant to the matter. Rather, the point is that even if possible worlds were abjured, certain problems in the semantics of propositional attitude discourse would persist, and it is these which are distinctively problems of the interpretation of propositional attitude discourse.

For present purposes, possible worlds are primitive model-theoretic items that play a certain role in formal theories-- roughly, they are "carriers" of non-extensionality. Does our use of possible worlds in doing formal semantics commit us to some sort of realism about them? Only in the sense that the ordinary semantics for classical (extensional) first-order logic commits us to Platonism about mathematical objects-- namely, the sets, functions, ordered n-tuples, and so forth, used to interpret quantified predicate logic. For that matter, the usual semantics for classical propositional logic commits us to realism about truth functions in this same sense. And just as appeal to a domain of individuals involves no particular commitment as to what individuals really <u>are</u> (i.e., as to whether they are physical objects, constructions out of sense data, ideas in the mind of God, or whatever), so the formal semantic use of possible worlds is metaphysically neutral. I can see no basis for singling out possible worlds, from among the various model-theoretic devices we use, for special worry.

2. This way of characterizing the problem, in terms of the algebraic structure to be imposed on the "domain" of propositions, is exploited by Richmond Thomason, probably under the influence of relevance logic. See Thomason (1980).

3. The term is due, I believe, to Cresswell. See Cresswell (1975). Cresswell's approach involves investing the objects of belief (etc.) with more structure than is the case in ordinary possible worlds semantics; structure that "mirrors" to some extent the syntactic structure of sentences. Thomason, by contrast, treats propositions as primitive items in the model theory, lacking even the internal structure of sets, but whose algebraic relations to one another are imposed, externally, as it were, by meaning postulates. The present discussion applies equally well to both, since the immediate concern is not with how hyperintensionality is achieved. This is not to suggest that there are no philosophically important differences between the two.

4. This is not, of course, a consequence of the view that belief contexts are hyperintensional, but a natural concomitant of it. If it were allowed that non-indexical sentences might express different propositions, depending on the context in which they occur, there would be no need to regard belief contexts as hyperintensional in order to account for substitution failures, and the like. So I

take it that hyperintensional approaches are motivated by a presumption of the sort indicated.

The alternative to a hyperintensional treatment of belief contexts that I am going to suggest, involves introducing a certain plasticity into mappings from sentences to propositions, when sentences occur in the complement clause of a belief statement. Possible worlds semantics then suffice to account for intentional phenomena, and have the virtue (as I see it) of being minimally non-extensional.

5. Testimonial evidence I take to include, in addition to straightforward assent to and dissent from sentences, reports of general judgements about beliefs (e.g., "Believing that P is not the same as believing that P and, Q or not-Q"), and perhaps some non-verbal responses to linguistic stimuli, such as requests or commands. In brief, testimonial evidence is behaviour, linguistic or otherwise, that is relevant to determining agents' beliefs about meanings.

6. In Thomason's model theory (<u>op. cit.</u>) meaning postulates play two roles. First they determine the algebraic structure of the domain of propositions, and second, they determine the truth conditions of propositional attitude statements (indirectly, by constraining the relations of propositions to truth values, relations which, on his theory, are not intrinsic, as they are in possible worlds semantics). It is the latter role we are, in effect,

considering here.

7. That is, the modal operators of the Lewis systems, for example. See Dugundji (1940). Many-valued logics were of course objects of independent interest, and were themselves sometimes motivated by modal considerations, such as the presumed failure of the law of excluded middle to hold where "future contingents" are concerned.

8. Montague (1968) and Montague (1970a).

9. Stalnaker (1972); and Kaplan (1977). In fact, the resources of Montague's "Pragmatics" suffice to avoid the difficulty. See the discussion of "product models" in Montague (1974), page 107.

10. ibid. Also published in French, Uehling, and Wettstein (1979).

11. The sketch to follow is, again, inspired by work of Stalnaker, especially Stalnaker (1972) and Stalnaker (1976). See also Lewis (1980) for a discussion of "double indexing".

12. The 'C' stands for 'concept', coming from Stalnaker's term 'propositional concept' for those arrays whose elements are all truth values, and which are associated with sentences.

13. Thomason (1980).

14. The scare-quotes around 'same', and the parenthetical 's' in 'belief(s)' are used in order to remain neutral between two ways of speaking about the individuation of belief. What matters, of course, is how we individuate the propositional objects of belief, and not whether we say of two people, who assent to the same sentence, that they share a belief.

15. See Chapter V, § 1, for a fuller discussion of this point.

THE MODEL THEORY

1. Sketch of a Model Theory for a Non-quantificational Predicate Language with a Belief Operator

Let D be a domain of individuals, W a class of possible worlds, and I a set of interpretation functions i_w indexed to possible worlds (intuitively, i_w gives the interpretation of non-logical constants relative to w).

We define $|w| = \{w_k : i_{w_k} = i_w\}$, the equivalence class of worlds that all index $i_w \in I$. A set $x \subseteq W$ is <u>index-closed</u> if $w \in x \Rightarrow |w| \subseteq x$.

Also let $\mathcal{P}_w(W) = \{x \in \mathcal{P}(W) : w \in x\}$, the principal filter on $\mathcal{P}(W)$ generated by w.

A <u>model</u> Γ is a quintuple $\langle D, W, \mathcal{P}_w(W), I, i_w \rangle$ for some $w \in W$ and $i_w \in I$; intuitively, i_w is the standard interpretation according to Γ.

Let $\mathcal{L}$ be a language containing the truth-functional connectives, and the following non-logical constants:

i) a class $\mathcal{C}$ of individual constants {a, b, c, ...}

and for each $n \geq 1$,

ii) a class $\mathcal{P}^n$ of n-place predicate constants {F, G, ...}.

The sentences of $\mathcal{E}$ are formed in the usual manner.

The following conditions on interpretation functions hold: for each $i_w \in I$

$$\text{i)} \quad i_w : \mathbb{C} \rightarrow {}^W D;$$

and

$$\text{ii)} \quad i_w : \mathbb{P}^n \rightarrow {}^W \underbrace{(D \times D \times \ldots \times D)}_{n \text{ times}},$$

(where ${}^X Y$ denotes the set of functions from X into Y).

These clauses represent minimal constraints on interpretations. More restrictive constraints are possible, and perhaps (for some purposes) desirable. For example, it could be stipulated that each $i_w \in I$ be such that $rng(i_w | \mathbb{C})$ contain only constant functions; that is, that the members of $\mathbb{C}$ are rigid designators.

Let Γ now be a particular model $\langle D, W, \mathcal{P}_{w_o}(W), I, i_{w_o} \rangle$.

Then $[\ Fa\ ]^{\Gamma} =_{df}$ $\iota f \in {}^W 2$ such that

$$(\forall w \in W)(\ f(w) = 1 \leftrightarrow i_{w_o}(a)(w) \in i_{w_o}(F)(w)\ ).$$

$[\ Fa\ ]^{\Gamma}$ is the <u>intension</u> of 'Fa' according to Γ. The <u>extension</u> of 'Fa' in w_k (according to Γ) is then $[\ Fa\ ]^{\Gamma w_k} = [\ Fa\ ]^{\Gamma}(w_k) = f(w_k)$, where 'f' is as above. The intensions (and extensions) of truth-functionally complex sentences are determined in the usual way from the atomic cases.

Let Γ be as above. Then $w \in W$ is a <u>standard</u> world with respect to Γ if and only if $i_w = i_{w_o}$.

Suppose $A \in \mathcal{P}_{w_o}(W)$. Then $I_A =_{df} \{i_w \in I\colon w \in A\}$. We define

$$i_A = \iota f \text{ such that } (\forall x \in \bigcup\{\mathbb{C}, \bar{\mathbb{P}}^1, \ldots, \mathbb{P}^k, \ldots\})(\ f(x)(w) = $$
$$= i_w(x)(w), \text{ for } i_w \in I_A;$$
$$= i_{w_o}(x)(w) \text{ otherwise}).$$

i_A is the diagonal interpretation function, relative to A, and in terms of it we can define something roughly equivalent to Montague's "product model" construction. A might be regarded intuitively as a possible epistemic situation of an agent; that is, as a set of worlds that an (hypothetical) agent cannot distinguish from the actual world in point of interpretation. To facilitate this understanding of A, we henceforth assume that A takes only index-closed sets as values.

The <u>diagonal</u> <u>intension</u> of an expression Φ of $\mathcal{E}$, relative to Γ and A is $[\ \Phi\ ]^{\Gamma(i_A/i_{w_o})}$, where $\Gamma(i_A/i_{w_o})$ is the model like Γ with i_A in place of i_{w_o}. $[\ \Phi\ ]^{\Gamma(i_A/i_{w_o})}$ will be abbreviated $[\ \Phi\ ]^{\Gamma A}$. This corresponds to what was earlier denoted $*C(\Phi)$.

Let $\mathcal{E}_B = \dot{\mathcal{E}} \cup \{B\}$, where 'B' is a binary operator taking an individual constant and a sentence, to yield a sentence. We stipulate that iii) $i_w(B) \in {}^W(\mathcal{P}(D \times {}^W2))$.

Belief in a proposition was characterized informally as belief that the actual world is one of the worlds where the proposition takes the value 1. Accordingly, in the set-theoretic algebra of propositions, the beliefs of an agent form a non-trivial filter. This condition needs to be imposed as an additional constraint on the operator 'B', if 'B' is to reflect our intended conception of belief.

In stating this condition, we employ set-theoretic notation to characterize properties of, and relations among, members of ${}^{W}2$, in a straightforward, although literally incorrect way. For example, where $\Phi, \Psi \in {}^{W}2$, we will write $\Phi \subseteq \Psi$, to mean $(\forall w \in W)(\Phi(w) \leq \Psi(w))$. In effect, we are treating the members of ${}^{W}2$ as the sets of which they are in fact the characteristic functions.

iii)' $(\forall w, w' \in W)(\forall d \in D)(\langle d, \Lambda\rangle \notin i_w(B)(w'))$,

(where Λ denotes the empty set);

iii)'' $(\forall w, w' \in W)(\forall d \in D)(\langle d, \Phi\rangle \in i_w(B)(w')\ \&$
$\&\ \langle d, \Psi\rangle \in i_w(B)(w') \rightarrow \langle d, \Phi \cap \Psi\rangle \in i_w(B)(w'))$;

iii)''' $(\forall w, w' \in W)(\forall d \in D)(\langle d, \Phi\rangle \in i_w(B)(w')\ \&$
$\&\ \Phi \subseteq \Psi \rightarrow \langle d, \Psi\rangle \in i_w(B)(w'))$.

Then, for a given value of A, $[\ Ba\Phi\ ]^{\Gamma} =_{df}$ ιf $\in {}^{W}2$ such that $(\forall w \in W)(\ f(w) = 1 \leftrightarrow \langle i_{w_o}(a)(w),\ [\ \Phi\ ]^{\Gamma A}\rangle \in i_{w_o}(B)(w))$.

This yields a <u>de dicto</u>-like reading, or a <u>de re</u>-like reading, of the belief statement, according as i_A deviates, or not, from i_{w_o} on the non-logical constants in Φ.

In chapter I it was remarked that <u>de re</u> construals of belief statements are, so to speak, the designated construals, in the absence of indications to the contrary. Corresponding to this view is an understanding of $|w_o|$ as a "default value" for A. Thus, $[\ Ba\Phi\ ]^{\Gamma}$ is not semantically ambiguous. This point is elaborated somewhat in Chapter V.

It should be noted that a model does not tell us what the epistemic situation of any agent is, nor how to determine it. It only tells us what the intension (and the extension-in-a-world) of a sentence like 'BaΦ' is, relative to a range of possible epistemic situations. In this respect it resembles a model for a language containing indexicals, say, personal pronouns. Such models do not tell us who is speaking, and who is addressed, but relative to a determination of speaker and addressee, they give the intensions of sentences like "You owe me $5".

In earlier discussion, we saw in an informal way how the varying epistemic circumstances of agents could result in substitution failures in belief contexts. We want now to show that this feature is captured by the above apparatus, by means of somewhat more abstract examples. As in the earlier informal discussion, we will suppose W to be numerically ordered, so that functions with domain W can be identified with sequences of objects from the range of the function; in particular, propositions can then be displayed as sequences of 0's and 1's. This is an expository convenience only, and should not be taken as a real feature of the model theory.

Example 1.

Let Γ be as above, and let D = N , the set of natural numbers.

Set $\quad i_{w_o}(a)(w) = 0$, for all $w \in W$,

$i_{w_o}(b)(w) = 1$, for all $w \in W$,

$$i_{w_o}(c) = i_{w_o}(b),$$

$$i_{w_o}(F)(w) = \{1\}, \text{ for all } w \in W,$$

$$i_{w_1}(c)(w) = 2, \text{ for all } w \in W,$$

and set $$A = |w_o| \cup |w_1|;$$

and suppose that $(\forall x \in \bigcup\{\mathbb{C}, \mathbb{P}^1, \ldots, \mathbb{P}^k, \ldots\} - \{c\})(i_{w_o}(x) = i_{w_1}(x))$. Set $i_{w_o}(B)(w) = \{\langle 0, \langle 1, 1, 1, \ldots \rangle\rangle\}$ for all $w \in W$. Then $[\text{ BaFb }]^{\Gamma}(w) = 1$, for all $w \in W$, since $[\text{ Fb }]^{\Gamma}A = [\text{ Fb }]^{\Gamma} = \langle 1, 1, 1, \ldots \rangle$. On the other hand, $[\text{ BaFc }]^{\Gamma}(w_1) = 0$, since $[\text{ Fc }]^{\Gamma}A(w) = 1$ iff $i_A(c)(w_1) \in i_{w_o}(F)(w_1)$ iff $i_{w_1}(c)(w_1) \in i_{w_o}(F)(w_1)$; but $i_{w_1}(c)(w_1) = 2$, and $i_{w_o}(F)(w_1) = \{1\}$.

In other words, according to the model Γ and relative to A, 'BaFb' and 'BaFc' differ in truth value, even though the (standard) interpretation function i_{w_o} of Γ has both 'b' and 'c' rigidly designating the same object. That is, substitution of co-referential names may sometimes fail on Γ for A. The crucial feature of the treatment (in Chapter I) of the examples involving Oedipus and Pierre, namely, the presence of non-standard interpretations of proper names, indexed to possible worlds, is precisely what the parameter A contributes in the example above, by including $|w_1|$. i_{w_1} has the constant 'c' rigidly designating something other than the "actual" referent of 'c'.

2. Model Theory for a Quantificational Predicate Language with a Belief Operator

Let $\mathcal{E}_{BQ}$ be $\mathcal{E}_B$ supplemented with the symbols '$\forall$' and '$\exists$', and a class V of individual variables {x, x_1, x_2, ...}. A <u>variable assignment</u> is a function g: V → D. The class of such assignments is denoted G.

Let Φ be a formula having at most the variable 'x' occurring free. Φ(a/x) is the formula like Φ, but with a ε $\mathbb{C}$ everywhere substituted for free 'x'. We will define $[\ \Phi\ ]^{\Gamma g} = [\ \Phi\ ]^{\Gamma}$ if Φ has no free variables, and $[\ \Phi\ ]^{\Gamma g} = [\ \Phi(e/x)\ ]^{\Gamma'}$ if Φ has free occurrences of 'x', where 'e' is an individual constant not occurring in Φ, and Γ' is the model like Γ, with the (possible) difference that $i'_{w_o}(e)(w) = g(x)$, for all w ε W (i'_{w_o} being the standard interpretation function of Γ'). The generalization to cases of formulae in more than one free variable is straightforward. Note that, although i'_{w_o} is characterized in terms of i_{w_o}, in general w_o will <u>not</u> be the index of i'_{w_o}. Only if $i_{w_o}(e)(w) = g(x)$, for all w ε W, is $i'_{w_o} = i_{w_o}$.[1]

We define $[\ (\forall x)\Phi\ ]^{\Gamma} = (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow$
$\leftrightarrow (\forall g \in G)([\ \Phi\ ]^{\Gamma g}(w) = 1))$,
and $[\ (\exists x)\Phi\ ]^{\Gamma} = (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow$
$\leftrightarrow (\exists g \in G)([\ \Phi\ ]^{\Gamma g}(w) = 1))$.

Some further definitions are required, to cover the cases of statements involving quantifiers and the belief operator together. For given $e \in \mathfrak{C}$, $g \in G$, and $x \in V$,

$$I_{A'} =_{df} \{i'_{w_k} : w_k \in A\};$$

$$A' =_{df} \{w : i_w \in I_{A'}\};$$

$$\Gamma'_A =_{df} \Gamma'(i_{A'}/i_{w_o}).$$

These definitions have the consequence that " ' " and "A", considered as operators on models, commute. That is, $\Gamma'_A = (\Gamma_A)'$, or in other words, $\Gamma g_A = \Gamma_{Ag}$. In addition, if $\Gamma_A = \Gamma$, then $\Gamma'_A = \Gamma'$. What this means is that epistemic situations of agents, however bizarre, cannot disrupt the interpretation of quantification. The underlying intuition is that agents whose semantic ignorance or confusion extends to logical devices like the quantifiers, are beyond the pale, as far as an adequate theory of de dicto construals of belief attribution is concerned.

Example 2.

We want to show that it is not in general the case that $[\ Ba(\exists x)Fx\ ]^{\Gamma} = [\ (\exists x)BaFx\ ]^{\Gamma}$. Let $\Gamma = \langle D, W, \mathcal{P}_{w_o}(W), I, i_{w_o}\rangle$, and D = N, as in the previous example, and set

$$i_{w_o}(a)(w) = 0, \text{ for all } w \in W,$$

$$i_{w_o}(F)(w_o) = \Lambda (\text{the empty set}),$$

$$i_{w_o}(F)(w_k) = \{k\}, \text{ for } k > 0,$$

and $\quad i_{w_o}(B)(w_k) = \{<0, <0, 1, 1, 1, \ldots>>\}$, for some fixed w_k.[2] Let $A = |w_o|$ (hence, $i_A = i_{w_o}$).

i) Now, $[\ Ba(\exists x)Fx\ ]^{\Gamma} = (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow$ $\leftrightarrow <i_{w_o}(a)(w), [\ (\exists x)Fx\ ]^{\Gamma A}> \in i_{w_o}(B)(w))$. Because $A = |w_o|$, the right hand side of the equivalence is the same as $<i_{w_o}(a)(w), [\ (\exists x)Fx\ ]^{\Gamma}> \in i_{w_o}(B)(w)$. $[\ (\exists x)Fx\ ]^{\Gamma} =$

$$= (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow (\exists g \in G)([\ Fx\ ]^{\Gamma g}(w) = 1)).$$

Given the characterization of i_{w_o} above, and the definition of $[\ \phi\ ]^{\Gamma g}$, it is easy to see that $\sim(\exists g \in G)([\ Fx\ ]^{\Gamma g}(w_o) = 1)$,

$$(\exists g \in G)([\ Fx\ ]^{\Gamma g}(w_1) = 1),$$

$$(\exists g \in G)([\ Fx\ ]^{\Gamma g}(w_2) = 1),$$

$$\vdots$$

etcetera.

That is, $[\ (\exists x)Fx\ ]^{\Gamma} = <0, 1, 1, 1, \ldots>$.

Hence, $[\ Ba(\exists x)Fx\ ]^{\Gamma}(w_k) = 1$.

ii) Next consider $[\ (\exists x)BaFx\ ]^{\Gamma} =$

$$= (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow (\exists g \in G)([\ BaFx\ ]^{\Gamma g}(w) = 1).$$

For any particular choice of $g \in G$, $[\ BaFx\ ]^{\Gamma g} = [\ BaFe\ ]^{\Gamma'} =$

$= (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow \langle i'_{w_o}(a)(w),\ [\ Fe\ ]^{\Gamma' A}\rangle \in i'_{w_o}(B)(w))$.

Again, $\Gamma'_A = \Gamma'$. Now suppose that $g(x) = 1$. Then $[\ Fe\ ]^{\Gamma'} =$

$= (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow i'_{w_o}(e)(w) \in i'_{w_o}(F)(w))$.

Given the definition of Γ', this will be $\langle 0, 1, 0, 0, 0, \ldots\rangle$.

Similarly, if $g(x) = 2$, $[\ Fe\ ]^{\Gamma'} = \langle 0, 0, 1, 0, 0, \ldots\rangle$, and so forth. Since i'_{w_o} differs from i_{w_o} <u>at most</u> on $\{e\}$, it can be seen that $\langle i'_{w_o}(a)(w_k),\ [\ Fe\ ]^{\Gamma'}\rangle \notin i'_{w_o}(B)(w_k)$, for any choice of $g \in G$. Hence, $[\ (\exists x)BaFx\ ]^{\Gamma}(w_k) = 0$. This concludes example two.

In the above example $A = |w_o|$. This shows that the failure of '$\exists$' to commute with 'B' is not due to the ideosyncrasies of agents' epistemic situations-- $|w_o|$ is, in effect, the situation of semantic omniscience, according to Γ.

The converse implication to the above, however, holds. That is, $[\ Ba(\exists x)Fx\ ]^{\Gamma}(w) = 1$, whenever $[\ (\exists x)BaFx\ ]^{\Gamma}(w) = 1$. For suppose that $[\ (\exists x)BaFx\ ]^{\Gamma}(w_k) = 1$, for some fixed w_k. Then $(\exists g \in G)([\ BaFx\ ]^{\Gamma g}(w_k) = 1$; i.e., $[\ BaFe\ ]^{\Gamma'}(w_k) = 1$, where $i'_{w_o}(e)(w) = g(x)$, for all $w \in W$. So $\langle i'_{w_o}(a)(w_k),\ [\ Fe\ ]^{\Gamma' A}\rangle \in$

$\in i'_{w_o}(B)(w_k)$. Now i'_{w_o} differs from i_{w_o} at most on $\{e\}$, so it follows that $\langle i_{w_o}(a)(w_k),\ [\ Fe\ ]^{\Gamma' A}\rangle \in i_{w_o}(B)(w_k)$;

i.e., $\langle i_{w_o}(a)(w_k),\ [\ Fx\ ]^{\Gamma g A}\rangle \in i_{w_o}(B)(w_k)$;

so $\langle i_{w_o}(a)(w_k),\ [\ (\exists x)Fx\ ]^{\Gamma A}\rangle \in i_{w_o}(B)(w_k)$;

hence $[\ Ba(\exists x)Fx\ ]^{\Gamma}(w_k) = 1$.

Notice that the above considerations turn on no assumptions concerning the value of the parameter A, and that the penultimate step depends on condition iii)''', the "deductive closure" of belief. Parallel results for the universal quantifier, naturally, obtain as well.[3]

A further example will serve to give a bit more of the flavour of the treatment of belief, and also to anticipate later developments. Let us consider whether, and under what conditions, 'BaFb' implies '$(\exists x)$BaFx'.

Example 3.

Suppose that $[\ BaFb\ ]^{\Gamma}(w_k) = 1$, for some fixed w_k. Then $\langle i_{w_o}(a)(w_k),\ [\ Fb\ ]^{\Gamma A}\rangle \in i_{w_o}(B)(w_k)$. Now we want if possible to show that $[\ (\exists x)BaFx\ ]^{\Gamma}(w_k) = 1$. This would follow if

$[\ BaFx\ ]^{\Gamma g}(w_k) = 1$, for some $g \in G$; that is, if

$<i_{w_o}(a)(w_k), [\ Fx\ ]^{\Gamma g}A> \in i_{w_o}(B)(w_k)$, with respect to the same variable assignment g. And we could get this if it were possible to substitute $[\ Fx\ ]^{\Gamma g}A$ for $[\ Fb\ ]^{\Gamma}A$. Now, $[\ Fb\ ]^{\Gamma}A =$ $= [\ Fx\ ]^{\Gamma g}A$ provided that

i) $A = |w_o|$.

ii) $i_{w_o}(b)$ is a constant function,

and iii) $g(x) = i_{w_o}(b)(w)$.

Variable assignments act like rigid designators. i) and ii) suffice to guarantee that $i_A(b)$ is a constant function, and iii) yields co-referentiality. Thus, if conditions i) and ii) are satisfied, 'BaFb' implies '$(\exists x)$BaFx'.

Where $i_A(b)$ is not a constant function, it is easy to see that the implication fails; i.e., that there is a model Γ and a world w such that $[\ BaFb\ ]^{\Gamma}(w) = 1$ and $[\ (\exists x)BaFx\ ]^{\Gamma}(w) = 0$. For suppose that there exist w, w' $\in$ W such that $i_A(b)(w) \neq i_A(b)(w')$. In particular, set $i_w(a)(w') = 0$, for all w, w' $\in$ W;

$i_{w_k}(b)(w) = k+1$, for all w $\in$ W, and all k $\in$ N;

$i_w(F)(w_k) = \{k+1\}$, for all w $\in$ W, and all k $\in$ N.

Let $A = \bigcup_k |w_k|$,

and $i_w(B)(w') = \{<0, <1, 1, 1, \ldots>>\}$, for all w, w' $\in$ W.

According to the above, $[\ BaFb\ ]^{\Gamma}(w) = 1$, for all w $\in$ W. But

$[\ Fx\ ]^{\Gamma g_A} = \langle \underbrace{0, \ldots, 0}_{k \text{ times}}, 1, 0, \ldots\rangle$, for $g(x) = k$. So,

$\langle i_w(a)(w'), [\ Fx\ ]^{\Gamma g_A}\rangle \notin i_w(B)(w')$, for any $w, w' \in W$. In other words, $[\ (\exists x)BaFx\ ]^{\Gamma}(w) = 0$, for all $w \in W$.

$i_A(b)$ is constant only if $i_w(b) = i_{w'}(b)$, for all $w, w' \in A$ (assuming that for $w \in A$, $i_w(b)$ is constant). Thus, the constancy of i_A reflects the agent's correct understanding of the reference of 'b'.

3. Identity and Definite Descriptions

Let $\mathcal{E}^*$ be $\mathcal{E}_{BQ} \cup \{=, \iota\}$. Then

$$[\ \alpha = \beta\ ]^{\Gamma} =_{df} (\iota f \in {}^{W}2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow$$
$$\leftrightarrow [\ \alpha\ ]^{\Gamma}(w) = [\ \beta\ ]^{\Gamma}(w)), \text{ where } [\ \alpha\ ]^{\Gamma} = i_{w_0}(\alpha), \text{ for } \alpha \in \mathbb{C}.$$

'ι' is the description operator, generating terms of the form $\iota x\Phi$ from formulae containing free occurrences of 'x'.

$$[\ \iota x\Phi\ ]^{\Gamma} =_{df} (\iota f \in {}^{W}D)(\forall w \in W)(\ f(w) = d \leftrightarrow (\exists g \in G)$$
$$([\ \Phi\ ]^{\Gamma g}(w) = 1 .\&.\ g(x) = d .\&.\ (\forall g' \in G)(\ g'(x) \neq d \rightarrow$$
$$\rightarrow [\ \Phi\ ]^{\Gamma g'}(w) = 0))).$$

When the right hand side of the equivalence is not satisfied, $[\ \iota x\Phi\ ]^{\Gamma}$ is undefined. One is not thereby obliged to admit truth-value gaps; and in fact, we will stipulate that formulae containing undefined terms in "primary" occurrence are false. Semantically, then, the present treatment of descriptions is Russellian, without being committed to Russell's views on "logical form".

Three examples follow, to illustrate the properties of sentences of $\mathcal{E}^*$ on the semantics given.

Example 4.

We want to show that $b = \iota x BaFx \not\models Ba(b = \iota x Fx)$.

Let $i_{w_o}(a)(w) = 0$, for all $w \in W$,

$i_{w_o}(b)(w) = 1$, for all $w \in W$,

$i_{w_1}(b)(w) = 2$, for all $w \in W$,

$i_w(F)(w') = \{1\}$, for all $w, w' \in W$.

Let $A = |w_o| \cup |w_1|$, and $g(x) = 1$.

i) Then $[\ Fx\ ]^{\Gamma g}(w) = 1$, for all $w \in W$; i.e., $[\ Fx\ ]^{\Gamma g} =$ $= \langle 1, 1, 1, \ldots \rangle$. Now, $[\ Fx\ ]^{\Gamma g} = [\ Fx\ ]^{\Gamma g}A$ (since, for all $w, w' \in A$, $i_w(F) = i_{w'}(F)$), and this is the same as $[\ Fe\ ]^{\Gamma'}A$.

So $\langle i_{w_o}(a)(w), [\ Fe\ ]^{\Gamma'}A \rangle \in i_{w_o}(B)(w)$, for any $w \in W$.

That is, $[\ BaFe\ ]^{\Gamma'}A(w) = 1$, where $g(x) = 1$. But for any $g' \in G$ such that $g'(x) \neq 1$, $[\ Fx\ ]^{\Gamma g'} = \langle 0, 0, 0, \ldots \rangle$; i.e., $[\ Fx\ ]^{\Gamma g} \not\subseteq [\ Fx\ ]^{\Gamma g'}$. Hence, $[\ BaFe\ ]^{\Gamma'}A(w) = 0$, whenever $g(x) \neq 1$, for any w.

So, $\quad i_{w_o}(b)(w) = [\ \iota xBaFx\ ]^{\Gamma}(w)$,

and so, $\quad [\ b = \iota xBaFx\ ]^{\Gamma}(w) = 1$.

ii) $\quad [\ Ba(b = \iota xFx\ ]^{\Gamma}(w) = 1$ iff

$\langle i_{w_o}(a)(w),\ [\ b = \iota xFx\ ]^{\Gamma}A\rangle \in i_{w_o}(B)(w)$.

Now, $\quad [\ b = \iota xFx\ ]^{\Gamma}A(w) = 1$ iff $[\ b\ ]^{\Gamma}A(w) = [\ \iota xFx\ ]^{\Gamma}A(w)$.

But $\quad [\ b\ ]^{\Gamma}A(w_1) = 2$, and $[\ \iota xFx\ ]^{\Gamma}A(w_1) = 1$.

So $\quad [\ Ba(b = \iota xFx)\ ]^{\Gamma}(w_1) = 0$.

Example 5 (converse to example 4).

To show that $Ba(b = \iota xFx) \not\models b = \iota xBaFx$,

set $\quad i_w(a)(w') = 0$, for all $w, w' \in W$,

$i_{w_k}(b)(w) = k + 1$, for all $w \in W$, and all $k \in N$.

Let $\quad i_w(F)(w_o) = \{1\}$,

$i_w(F)(w_1) = \{2\}$,

$i_w(F)(w_k) = \{3\}$, $k > 1$, for all w,

$i_w(B)(w') = \{\langle 0, \langle 1, 1, 1, 0, 0, \ldots\rangle\rangle\}$, for all w, w',

and let $\quad A = \bigcup_k |w_k|$.

i) $\quad [\ b = \iota xFx\ ]^{\Gamma}A(w) = 1$ iff $i_A(b)(w) = [\ \iota xFx\ ]^{\Gamma}A(w)$.

So $[\ b = \iota x Fx\ ]^{\Gamma A} = \langle 1, 1, 1, 0, 0, 0, \ldots \rangle$;

Hence $[\ Ba(b = \iota x Fx)\ ]^{\Gamma}(w) = 1$, for all w.

ii) $[\ \iota x BaFx\ ]^{\Gamma}(w) = (\iota d \in D)([\ BaFx\ ]^{\Gamma g}(w) = 1 \leftrightarrow g(x) = d)$.

Now, $[\ BaFx\ ]^{\Gamma g}(w) = 1$ iff $\langle i_{w_o}(a)(w), [\ Fx\ ]^{\Gamma g A} \rangle \in i_{w_o}(B)(w)$.

So if $g(x) = 1$, $[\ Fx\ ]^{\Gamma g A} = \langle 1, 0, 0, 0, \ldots \rangle$;

if $g(x) = 2$, $[\ Fx\ ]^{\Gamma g A} = \langle 0, 1, 0, 0, \ldots \rangle$;

if $g(x) = 3$, $[\ Fx\ ]^{\Gamma g A} = \langle 0, 0, 1, 0, 0, \ldots \rangle$;

and if $g(x) > 3$, $[\ Fx\ ]^{\Gamma g A} = \langle 0, 0, 0, 0, \ldots \rangle$.

That is, for <u>no</u> $g \in G$ does $[\ Fx\ ]^{\Gamma g A} = \langle 1, 1, 1, 0, 0, \ldots \rangle$.

Hence, $[\ \iota x BaFx\ ]^{\Gamma}$ is undefined, and so $[\ b = \iota x BaFx\ ]^{\Gamma}(w) \neq 1$, for any w.

Example 6.

'$BaG\iota xFx$' can be true even though 'ιxFx' is improper.

Let $i_w(a)(w') = 0$, for all $w, w' \in W$;

$i_{w_o}(F)(w) = \{1, 2\}$, for all w (alternatively, $i_{w_o}(F)(w)$ could be any set whose cardinality is not 1, for all w);

$i_{w_k}(F)(w) = \{1\}$, $k > 0$, for all w;

$i_w(G)(w') = \{1\}$, for all $w, w' \in W$;

$i_w(B)(w') = \{<0, <0, 1, 1, 1, \ldots>>\}$, for all w, w';

and set $\quad A = \bigcup_k |w_k|$.

Now $\quad [\text{ BaG}\iota x\text{Fx }]^{\Gamma}(w) = 1$ iff $<i_{w_o}(a)(w), [\text{ G}\iota x\text{Fx }]^{\Gamma A}> \in i_{w_o}(B)(w)$,

and $\quad [\text{ G}\iota x\text{Fx }]^{\Gamma}A(w) = 1$ iff $[\ \iota x\text{Fx }]^{\Gamma}A(w) \in i_A(G)(w)$; i.e.,

iff $(\exists g \in G)([\text{ Fx }]^{\Gamma}Ag(w) = 1$, and $g(x) \in i_A(G)(w)$, and for any

$g' \in G$ such that $g'(x) \neq g(x)$, $[\text{ Fx }]^{\Gamma}Ag'(w) = 0)$.

This latter condition fails for w_o, but holds for w_k, $k > 0$. In

other words, $\quad [\text{ G}\iota x\text{Fx }]^{\Gamma}A = <0, 1, 1, 1, \ldots>$.

Hence, $\quad [\text{ BaG}\iota x\text{Fx }]^{\Gamma}(w) = 1$, for all $w \in W$.

The forgoing examples are meant to show that the present approach accomodates a certain range of intuitions concerning the inferential properties of belief statements. Except for the first and the last, all the examples involve problems of commutation of the belief operator with other, variable binding, operators. These are the sorts of cases which are easily handled syntactically, by means of scope distinctions. And in many of the examples, the fundamental device of the present models, the "epistemic situation" A, played no role (where A is set equal to $|w_o|$, it does no work). Moreover, there are well know theories in which proper names are treated as logically complex, allowing similar treatment of belief statements whose embedded clauses contain proper names. In Montague's PTQ, for example, names are thus complex, being of the same logical type as other "denoting phrases" (to use Russell's term). One might then wonder whether the present approach really offers any advantage over more familiar accounts.

The advantage of the present approach lies first of all in its generality. Not only the devices of singular reference, but items of any logical type, can be given the same sort of treatment as are, for example, proper names. (Indeed, example 6 turns crucially on the fact that A contains worlds where the predicate 'F' has a different intension than on the standard interpretation.)

Second, it offers an account of the failure of substitution of co-intensive items, such as co-referential names (if names are regarded as rigid designators). This feature is not shared by PTQ, which does not incorporate the "product model" construction from Montague's earlier "Pragmatics". In PTQ, the difference between de dicto and de re construals of belief statements whose complement clauses contain proper names, shows up as a scope difference in the intensional logic into which the fragment of English is translated. However, it is a consequence of MP.1, according to which names are rigid designators, that the two construals are logically equivalent.

Finally, the present approach explicitly represents relevant features of agents' epistemic situations, which are thus directly implicated in such inferential failures as occur. A "logic of belief" which does not do so will be insufficiently sensitive to the very considerations which motivate such logics in the first place (as was argued in Chapter II).

In short, the forgoing model-theoretic account of belief appropriately reflects the virtues, earlier claimed and informally illustrated, of a "two-dimensional" treatment of belief. Its relatively greater power than one-dimensional, "syntactic" approaches, will perhaps become more evident in the sequel.

4. Iterated Belief

I want to conclude this chapter with some remarks on iterated belief, that is, on the semantics of statements of the form BaBbϕ, and related forms having nested occurrences of the operator 'B'. The semantic theory of this chapter is easily extended to give an account of such statements, but the extension is not quite so straightforward as one might expect.

Consider the statement, 'Saul believes that Pierre believes that London is pretty.' Putting aside the possibility that Saul is unclear about the reference of 'Pierre', or the meaning of 'believes', in other words, confining our attention to the complement clause of the embedded belief statement, a moment's reflection reveals in the statement a kind of "ambiguity", a bit like that involved in cases where pronominal reference is underdetermined (for example, 'Saul believes that Pierre believes that he is late'). Is it Saul's understanding of 'London is pretty', or Pierre's, that is operative? Whereas in the case of the statement 'Pierre believes that London is pretty' we had two main construals, here we have four, corresponding to the various possible combinations of epistemic circumstances of Pierre and Saul, that might play a role in interpreting the statement.

In terms of our formal device for representing epistemic situations, these possibilities are

i) $A_s = |w_o| = A_p$;

ii) $A_s = |w_o| \neq A_p$;

iii) $A_s \neq |w_o| = A_p$;

and iv) $A_s \neq |w_o| \neq A_p$ (where A_s is Saul's epistemic situation, and A_p, Pierre's). Case iv) in fact comprises two subcases, namely $A_s = A_p$, and $A_s \neq A_p$.

We would like each of the above cases to correspond to a construal of the statement which is (possibly) distinct from the others. We would like, for example, to be able to distinguish the reading of the statement according to which Saul believes that Pierre believes the proposition that he, Saul, takes 'London is pretty' to express, from the reading according to which Saul believes that Pierre beleives the proposition that he, Saul, takes Pierre to understand 'London is pretty' to express.

Suppose our (formal) belief statement is BaBcΦ, and that a's epistemic situation is A, and c's is C. Then $[\text{ BaBc}\Phi\ ]^{\Gamma}$ is the propoqsition which is true in a world w just in case

$$<i_{w_o}(a)(w), [\text{ Bc}\Phi\ ]^{\Gamma A}> \in i_{w_o}(B)(w).$$

And $[\text{ Bc}\Phi\ ]^{\Gamma A}$ is the proposition which is true in a world w just in case $<i_{w_o}(c)(w), [\ \Phi\ ]^{\Gamma AC}> \in i_{w_o}(B)(w).$

What is Γ_{A_C}? Γ_A was earlier defined as $\Gamma(i_A/i_{w_o})$. If we take this definition to be entirely general, then $\Gamma_{A_C} = \Gamma_A(i_C/i_A)$. This would have the consequence that, for any A, C $\in \mathcal{P}_{w_o}(W)$, $\Gamma_{A_C} = \Gamma_C$. So it is only c's epistemic situation which plays a role in the construal of BaBcΦ (where, again, it is only the interpretation of 'Φ' which is in question), and the four cases above collapse to two.

What is needed is a definition of Γ_{A_C} which reflects appropriately both A and C. So, let

$$\Gamma_{A_C} =_{df} \Gamma(i_{A \cup C}/i_{w_o}).$$

Then Γ_{A_C} has the following properties.

i) $A \subseteq C \rightarrow \Gamma_{A_C} = \Gamma_C$;

and ii) $C \subseteq A \rightarrow \Gamma_{A_C} = \Gamma_A$.

Together these imply that $A = C \rightarrow \Gamma_{A_C} = \Gamma_A = \Gamma_C$, and that $A = |w_o| \rightarrow \Gamma_{A_C} = \Gamma_C$, and $C = |w_o| \rightarrow \Gamma_{A_C} = \Gamma_A$, since $|w_o|$ is a subset of any (index-closed) A, C $\in \mathcal{P}w_o(W)$. In addition,

iii) $A \nsubseteq C \nsubseteq A \rightarrow \Gamma_{A_C} \neq \Gamma_A$;

$\neq \Gamma_C$;

$\neq \Gamma$.

In view of these properties of Γ_{A_C} it is obvious that the four-fold distinction among construals of BaBcΦ can be maintained. (The characterization of Γ_{A_C} may seem inappropriate in that, it implies that $\Gamma_{A_C} = \Gamma_{C_A}$. However, there is no compulsion, from the point of view of the theory, to construe C as c's epistemic situation as known to God, rather than as conceived by a. The model theory, again, does not tell us what the epistemic situation of any agent is, but only how such things go into determining the intensions of belief statements, under certain construals. So if we interpret a statement BaBcΦ with respect to particular index-closed members A and C of $\mathcal{P}_{w_o}(W)$, it does not follow that BcBaΦ must be interpreted with respect to the same C and A.)

A special sort of iterated belief statement is of some interest, namely, ones in which the different occurrences of 'B' are followed by the same individual constant. On the present account, BaBaΦ does not entail BaΦ, and Ba$\sim$BaΦ does not entail $\sim$BaΦ. This, I think, is as it should be— a semantic theory should not entail Cartesian principles of priviledged access.

To see that the latter implication does not hold, recall that

$$[\ \text{Ba}{\sim}\text{Ba}\Phi\ ]^{\Gamma}(w) = 1 \text{ iff } \langle i_{w_o}(a)(w),\ [\ {\sim}\text{Ba}\Phi\ ]^{\Gamma A}\rangle \in$$

$$\in i_{w_o}(B)(w);$$

and

$$[\ {\sim}\text{Ba}\Phi\ ]^{\Gamma}(w) = 1 \text{ iff } [\ \text{Ba}\Phi\ ]^{\Gamma}(w) = 0 \text{ iff}$$

$$\langle i_{w_o}(a)(w),\ [\ \Phi\ ]^{\Gamma A}\rangle \notin i_{w_o}(B)(w).$$

Let $i_{w_o}(a)(w) = 2$, for all $w \in W$, and let Φ be such that

$[\ \Phi\ ]^{\Gamma} = \langle 1, 0, 1, 0, \ldots\rangle$. Set $A = |w_o|$ (hence, $\Gamma_{A_A} = \Gamma_A = \Gamma$),

and set

$$i_{w_o}(B)(w_o) = \{\langle 2, \langle 1, 0, 1, 0, \ldots\rangle\rangle, \langle 2, \langle 0, 1, 1, 1, \ldots\rangle\rangle\}$$

$$i_{w_o}(B)(w_k) = \{\langle 2, \langle 1, 1, 1, \ldots\rangle\rangle\}, \text{ for all } k > 0.$$

Then $\langle i_{w_o}(a)(w_o), [\ \Phi\ ]^{\Gamma_A}\rangle \in i_{w_o}(B)(w_o)$, and so $[\ {\sim}Ba\Phi\ ]^{\Gamma}(w_o) = 0$.

But $[\ {\sim}Ba\Phi\ ]^{\Gamma_A} = [\ {\sim}Ba\Phi\ ]^{\Gamma} = \langle 0, 1, 1, 1, \ldots\rangle$, and so

$\langle i_{w_o}(a)(w_o), [\ {\sim}Ba\Phi\ ]^{\Gamma_A}\rangle \in i_{w_o}(B)(w_o)$; i.e., $[\ Ba{\sim}Ba\Phi\ ]^{\Gamma}(w_o) = 1$.

A version of the Moore paradox is a consequence of the failure of $Ba{\sim}Ba\Phi$ to imply ${\sim}Ba\Phi$ (together with the fact that 'Ba' distributes over '&'). That is, on the present theory $[\ Ba(\Phi\ \&\ {\sim}Ba\Phi)\ ]^{\Gamma}$ can take the value 1 at some worlds, but only ones where $[\ \Phi\ \&\ {\sim}Ba\Phi\ ]^{\Gamma_A}$ is false (w_o, in the above example, is such a world). In other words, $\models_{\mathscr{K}} Ba(\Phi\ \&\ {\sim}Ba\Phi) \rightarrow {\sim}(\Phi\ \&\ {\sim}Ba\Phi)$. Such beliefs are well described as "self-refuting".

FOOTNOTES

1. The definition of $[.\Phi]^{\Gamma g}$ (the intension of a formula Φ on a model Γ, with respect to a variable assignment g), is given in terms of a substitution instance of Φ, $\Phi(e/x)$. The treatment of quantification is not, however, substitutional, but referential, since the truth (at a world) of $\Phi(e/x)$ on Γ' is itself defined in terms of g(x). That is, although the definition of truth for quantified sentences does not proceed in the familiar way via a definition of satisfaction, the concept of satisfaction is implicit in the truth definition.

2. Strictly speaking, it is the closure of this set in accordance with the clauses iii)' - iii)''', which is the extension of 'B' in w_k; and likewise in the following examples.

3. The results for the universal quantifier are not, however, natural. I suppose that, intuitively, one should want entailment between 'Ba($\forall$x)Fx' and '($\forall$x)BaFx' to fail in both directions, on the grounds that agents might be mistaken about what things there are. Owing to the fact that our models contain fixed domains of individuals, which (domains) are not relative to worlds, the former sentence entails the latter. The latter, however, fails to entail the former, despite the fixed domains, due to a peculiarity of the closure conditions imposed on the extensions of 'B'. In particular,

clause iii)'' leaves open the possibility that the set of propositions believed by an agent, in a world, is, so to speak, ω-incomplete. That is, an agent could believe [Fx]$^{\ulcorner g}$, for each g ε G, without believing [(∀x)Fx]$^{\ulcorner}$. If clause iii)'' were replaced by a condition of closure under (possibly) <u>infinite</u> intersections, this would no longer be the case. Such a replacement would amount to stipulating that agents' belief sets contain a unique strongest (contingent) proposition. '(∀x)BaFx' would then entail 'Ba(∀x)Fx'.

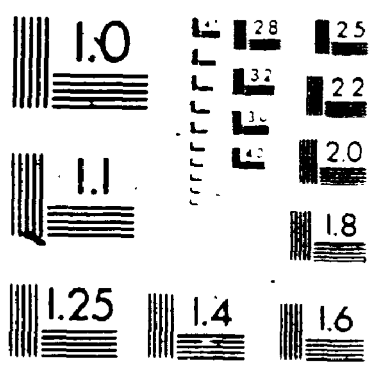
NAMES, INDEXICALS, AND DEFINITE DESCRIPTIONS IN BELIEF CONTEXTS--
WHY SCOPE ISN'T ENOUGH

In $\mathcal{E}^*$ we have two sorts of singular terms, definite descriptions, which are logically complex, and individual constants, which are not. Descriptions designate "non-rigidly"; that is, a description '$\iota x\phi$' will in general have as its intension $\ulcorner [\ \iota x\phi\ ] \urcorner$, a non-constant function. So far, nothing has been said in this regard about the members of $\mathcal{C}$ generally. Now, suppose we want $\mathcal{E}^*$ to contain analogues of both proper names, and of indexical expressions, such as the personal pronouns, and demonstratives. How are we to characterize semantically the difference between these two sorts of singular terms?

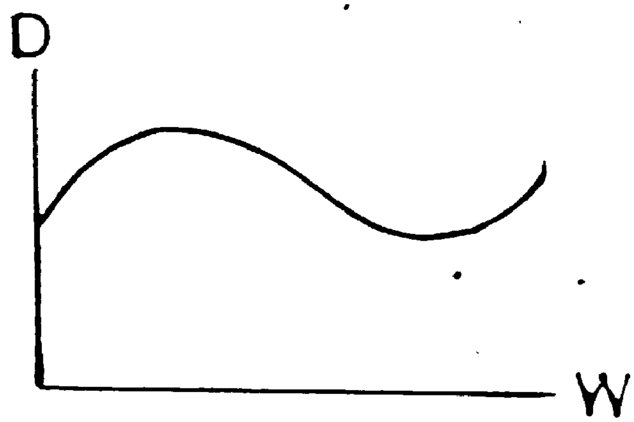
Names, plainly, refer rigidly in a sense in which indexicals do not. But indexicals are not descriptions. Relative to a context of tokening, indexicals refer rigidly, whereas descriptions typically do not. And with respect to different contexts, indexicals will vary in reference, whereas the reference of descriptions will typically not vary with variations in the context of utterance (descriptions, that is, that do not themselves contain indexicals-- examples like 'the candidate I favour' are, so to speak, doubly non-rigid).

In short, names, indexicals, and definite descriptions all differ from one another with respect to rigidity of designation. Pictorially, functions from possible worlds to objects in a domain

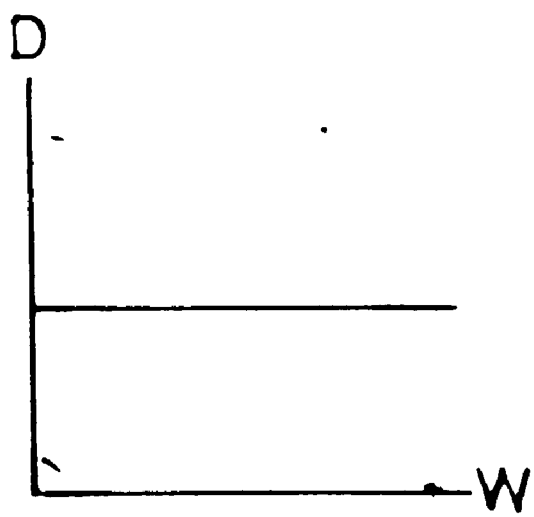
2 of/de 2



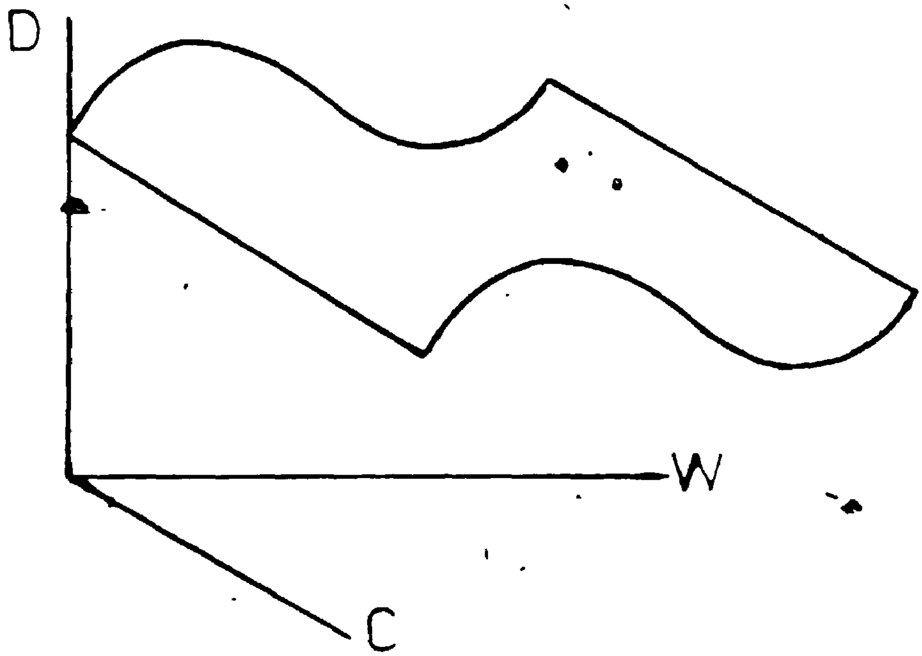
can be represented thus:[1]



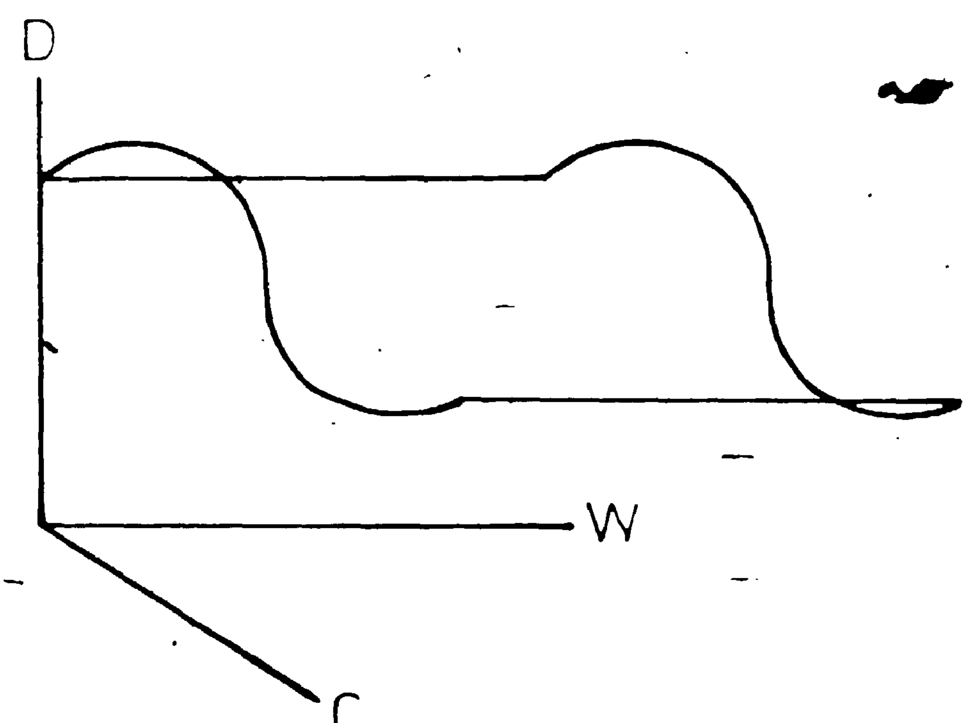
The above diagram represents a non-constant function in ${}^{W}D$, the sort of thing that is appropriate to be the intension of a definite description. Names, on the other hand, will have intensions that look like this:



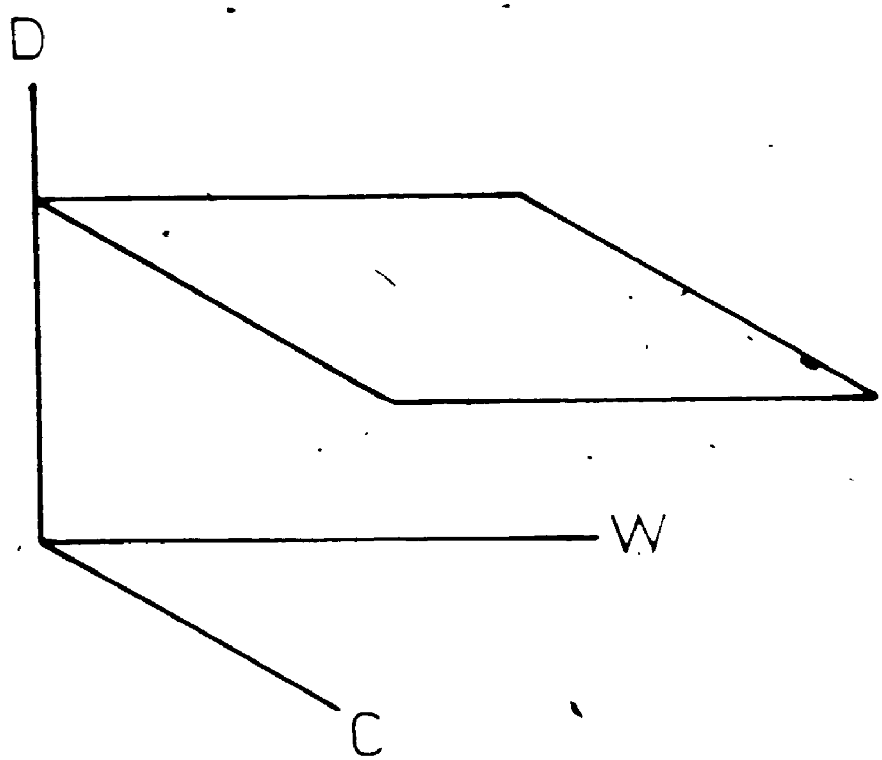
To distinguish indexicals, both from proper names and definite descriptions, along the lines just indicated, it is necessary to introduce another dimension into the diagrams, corresponding to context of tokening. Letting 'C' stand for the class of such contexts, we can give the following picture of the different sorts of singular terms:



descriptions



indexicals



names

Thinking of the members of W and C just as two sorts of indices that go into determining the extensions of singular terms, we can see that a two-dimensional model theory is required just to represent adequately singular terms, independently of any consideration of agents' epistemic situations and the interpretation of belief statements. Moreover, diagonal propositions are required in this two-dimensional model theory in order to distinguish truth at all occasions of tokening from necessary truth, and thus to account for the contingency of examples like "I am here now", also independently of any consideration of agents' epistemic circumstances. So when one turns to the task of representing the role of agents' epistemic circumstances in the semantics of a belief operator, it seems that what is really needed is a three-dimensional account.

Strictly speaking, that is correct. We can, however, ignore this complication for most purposes, for the following reason. When an indexical occurs within a belief context, its reference is determined by the context of tokening of the full belief statement, and not by anything to do with the epistemic situation of the agent to whom the belief attribution is made. For example, if someone says, "Mary believes that I am in Toronto," the occurrence of the indexical 'I' refers (rigidly) to the speaker. Ambiguity is of course possible, as in statements like "Fred believes that George believes that he is in Toronto"; but this is a garden variety problem of anaphora, which turns wholly on the occurrence of two candidates for the referent of 'he' prior to the occurrence of the pronoun itself-- the verb 'believes' is in no way implicated. In short, the reference of indexicals is fixed "externally" to belief context, when the two interact, and so indexicals act as rigid designators within such contexts, as indeed they do within intensional contexts generally.

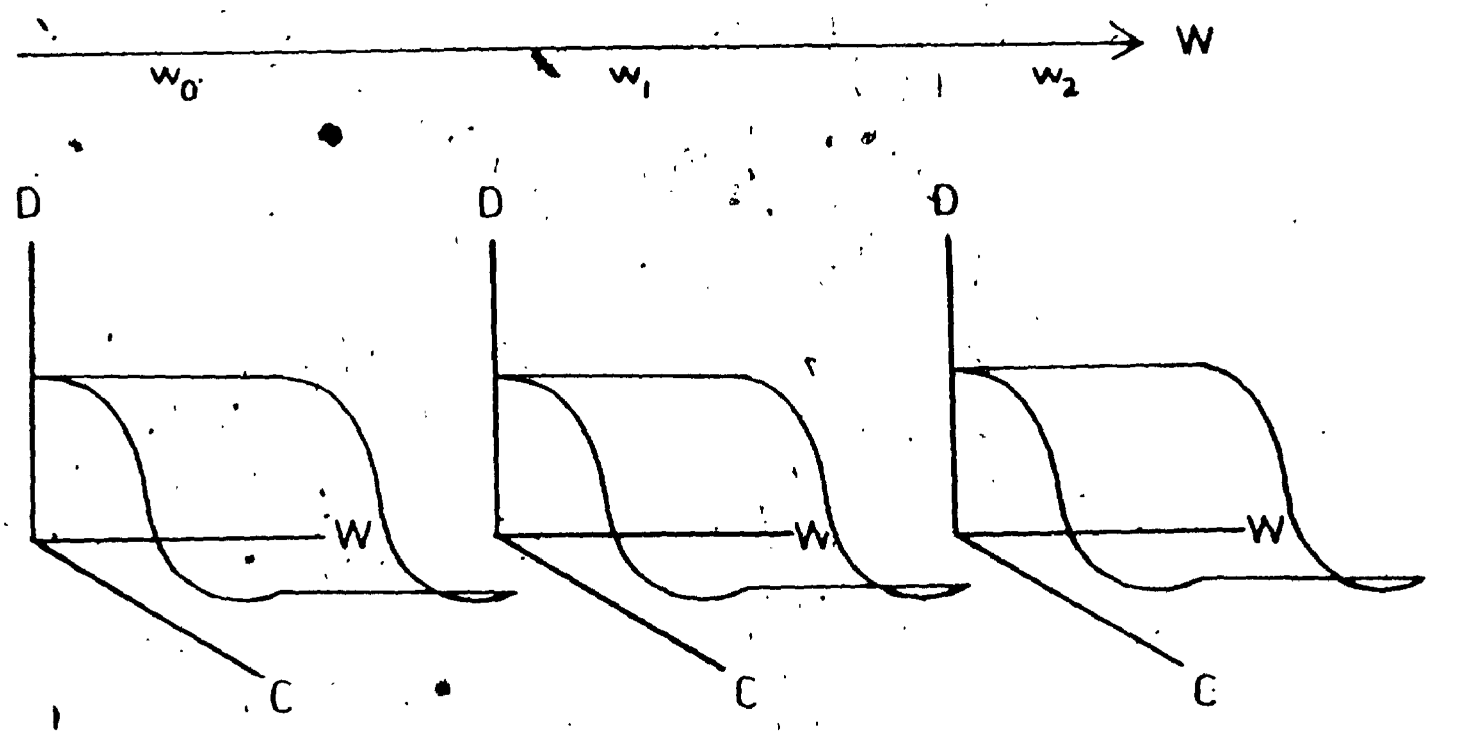
Relative to a particular context of tokening, then, indexicals have fixed, rigid, intensions, which are insensitive to belief contexts. Names, on the other hand, although semantically the most rigid of singular terms, are sensitive to belief contexts, and can display in such contexts a (pragmatic) variability rather like that shown by indexicals with respect to differing contexts of tokening.

Now the variability of indexicals with respect to context of tokening, which was treated by Montague as part of his theory of "pragmatics", is more properly seen as a semantic feature of those terms.[2] The context sensitivity of such terms is highly constrained, and the speaker who does not appreciate the mode of context sensitivity of, say, personal pronouns, is not a competent speaker of the language.

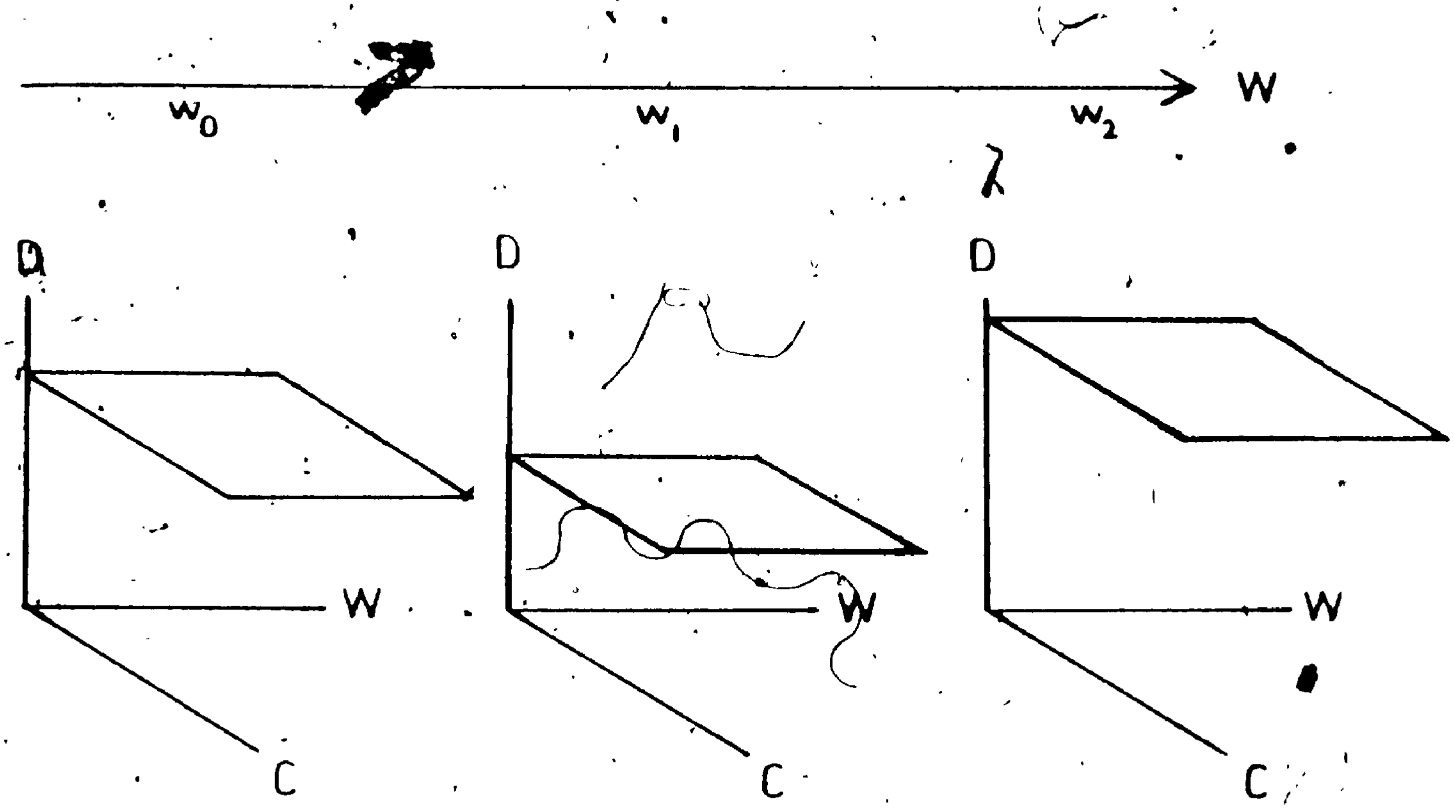
By contrast, the sensitivity of names to differing epistemic situations of the sort we've been considering is relatively unconstrained, and is not a semantic feature of names, which is why ignorance or error about the reference of names does not impugn the linguistic competence of an agent. (It is not entirely unconstrained, of course— the agent who thinks that a certain proper name is say, a verb, can be convicted of linguistic confusion.)

So it seems that names and indexicals undergo a kind of role reversal in belief contexts (where it is ignorance or error concerning the reference of proper names only which is in question, and not ignorance or error concerning the meaning or correct use of indexicals).

To illustrate this point, let us extend the two-dimensional diagrams for names and indexicals to a third dimension, also indexed by possible worlds, representing epistemic circumstances. (Note that the terms 'two-dimensional', 'three-dimensional', etc., refer to the number of dimensions of the various indices, and do not include the dimension D corresponding to possible extensions.) We can indicate this third dimension by a (horizontal) sequence of two-dimensional diagrams, as follows.



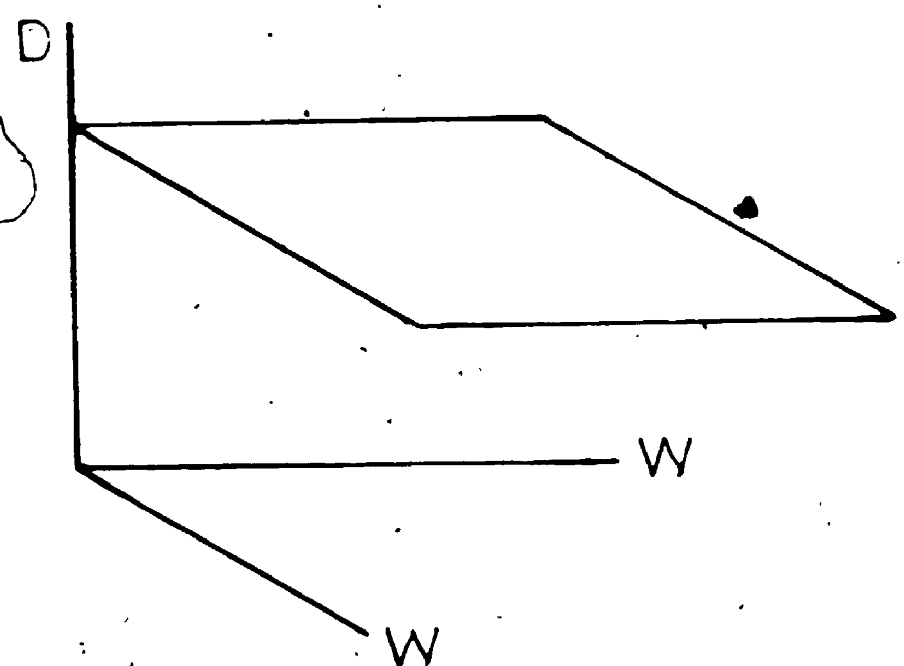
Indexicals



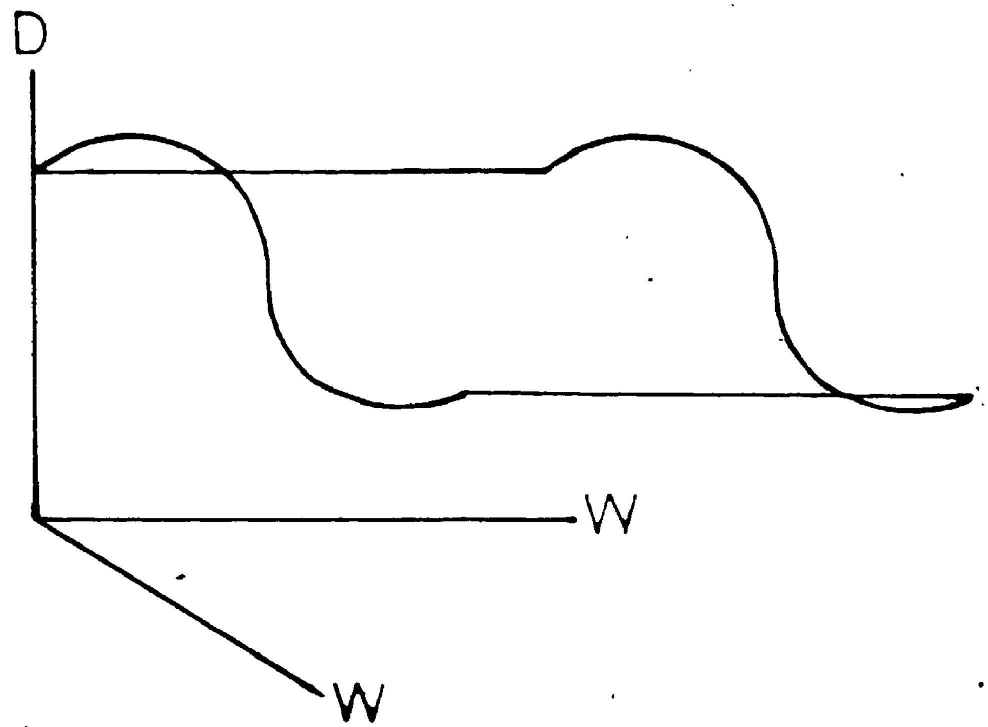
Names

The diagram for the indexical is the same, i.e., is constant, along the just introduced horizontal axis corresponding to the agent's epistemic situation, in keeping with our assumption that the agent is not ignorant or confused about the meaning of the indexical. The diagram for the name changes, however, indicating uncertainty concerning the referent of the name. Even so, in each world the name is a rigid designator whose referent is invariant with respect to different contexts of utterance (the members of C), indicating that the agent appreciates that the name _is_ a name, and not some other kind of singular term.

Now, for a fixed choice of context in C we will again have a two-dimensional diagram in $W \times W$ (x D, of course). And for the indexical, it will look like this



since the indexical varies in reference only along the dimension C. For the name it will look something like this



since the name varies in reference along one of the W-axes (the one corresponding to the agent's epistemic situation), and not the other. This illustrates the sense in which names are "like" indexicals, while maintaining the obvious semantic distinctions.

There are various straightforward ways in which the model theory developed thus far could be modified to reflect the forgoing considerations. The notion of a model could, for example, be expanded to include a class C of contexts (of tokening), members of which would then serve as additional arguments in the determination of the intensions of non-logical constants. Or, since a context of tokening of sentences is part of certain possible worlds, contexts could be identified with classes of worlds (those that are "congruent" with respect to the relevant features of the context), or with representatives of such classes.

The latter approach has the advantage that contexts are automatically cross-indexed with possible worlds, something that must in any case be accomplished one way or another, in order for diagonal intensions to be definable. And the definability of diagonal intensions is an adequacy requirement on any treatment of indexicals--they are what allow one to characterize the "contingent validity" that statements like 'I am here now' intuitively possess.

For the reasons already given, we will here omit actually making such modifications. We may note informally a pair of constraints that might plausibly be imposed on admissible values for the parameter A, amounting to a decision to confine attention to agents whose linguistic competence is not in question.

i) if $c \in \mathbb{C}$ is an indexical, then $(\forall w \in A)$ $(i_w(c) = i_{w_0}(c))$, where, again, i_{w_0} is the designated standard interpretation function of Γ.[3]

ii) if $c \in \mathbb{C}$ is a proper name, then $(\forall w \in A)$ $(i_w(c)(w_j) = i_w(c)(w_k))$, for all j, k.

These constraints reflect our intuition that it is a condition of linguistic competence that an agent's understanding of indexicals not be distorted, and that agents recognize proper names as rigid designators (although they may be mistaken about what is thus designated). This is not to suggest that agents may not also be mistaken or confused about such things. But the hard cases for a semantic theory are those which cannot readily be dismissed as deviant. In making the above assumptions, we address ourselves to the hard cases.

A certain difficulty arises which has received much attention in the literature on belief, and whose proper treatment requires appeal to the full account of singular terms sketched above. It is sometimes claimed that belief statements whose complement clauses contain indexical expressions have a special status among belief attributions. In the particular case where the indexical in question is a pronoun referring to the object of the attribution (i.e., the subject of the belief) we have so-called "first person" belief. It is argued (variously, by various authors; e.g., John Perry, and David Lewis) that some indexical beliefs are irreducibly so-- the indexical element is ineliminable.[4]

Typical examples purporting to show this are cases where an agent apparently has a certain indexical belief, but lacks the corresponding _de dicto_ or _de re_ belief; or vice versa. That is, the statement 'a believes that he is F' may be true, while 'a believes that b is F', where 'b' refers to a, is apparently false on both the _de dicto_ and _de re_ construals.

Against this view it has been argued (variously, by, e.g., Stalnaker, and Boer and Lycan) that first person belief is not thus irreducible.[5] I do not wish to survey the arguments pro and con, but rather to make a couple observations on the subject.

First, such examples are a special type of substitution problem, which in many respects resembles the more familiar (non-indexical) ones. In the schematic case above, for example, it is a's failure to realize that 'b' refers to himself, a, that accounts, intuitively, for a's beliefs. An amnesiac might be mistaken, or uncertain, about the reference of his own name; and anyone might fail to recognize a definite description as applying to himself. Such cases do not seem to me to differ importantly from the more familiar ones; on the contrary, they seem to admit of quite similar treatment. That is, where 'b' is a name, or a definite description not itself containing indexicals, the model theory already articulated will suffice to accomodate the example, in the manner already employed to deal with Oedipus and Jocasta, or Pierre and London.[6]

When, on the other hand, 'b' is itself an indexical expression referring to a, the matter is more complicated.[7] Suppose, for example, that 'b' is a demonstrative expression, like 'that man', which the agent uses, or would use, to refer to what happens to be himself, seen reflected in a mirror (without realizing that he is looking in a mirror, or at least, not realizing that it is his own reflection he sees). So the agent believes, perhaps, that <u>he</u> is F, and that <u>that man</u> is not.

The difficulty, intuitively, is this: the belief that "that man" is not F appears to qualify both for <u>de dicto</u> and <u>de re</u> status, since the agent will assent to the sentence 'That man is not F', and since there is direct perceptual acquaintance with the referent of 'that man' (putting aside the fatuous view that what one <u>is</u> acquainted with in such circumstances is not a person, but an image). Thus the indexical element of a's belief that he is F appears, in this case, to be genuinely ineliminable; one cannot appeal, in the manner of previous examples, to ignorance about the singular term 'that man', in order to explain the agent's seemingly conflicting beliefs.

Formally, the problem is that, on the account of indexicals given above, both 'he' and 'that man' will (relative to a context of tokening) rigidly designate the same object in all worlds in A, the agent's epistemic situation, as illustrated above (in the penultimate diagram). Hence the diagonal intensions of both terms will coincide with each other, and with their (common) actual intension. There seems to be no room to maneuver, given our assumption, embodied in the constraint i) above, that the agent understands the proper use of personal pronouns and demonstratives.

The solution to this difficulty rests in noticing that in a case of the sort described, the agent, although not ignorant of the mode of reference of the demonstrative 'that man', is ignorant or mistaken about the context of tokening in which the demonstration is made. There is relevant information about the context which the agent lacks, and in the absence of which, the reference of the term cannot be determined. There is, in other words, a range of possible contexts of tokening that the agent is unable to distinguish from the actual one, which would determine different referents for the expression 'that man'.

In a fully articulated model theory, the parameter A would contain not only possible worlds, but also contexts of tokening (however these are represented), to give a fuller characterization of agents' epistemic situations. Because context of tokening is the dimension along which indexicals are variable, the parameter A would then sometimes determine diagonal intensions for indexicals which differ from their actual intensions with respect to any particular context of tokening. The substitution failures can then be accounted for in a fashion exactly parallel to that employed in non-indexical cases.

It would be possible, and perhaps desirable, to impose a further constraint on the parameter A, stipulating that one is never mistaken or confused about the relevant features of the context of one's own tokenings of the first person pronouns. If this is done, the so-called first person belief attributions will, in a certain sense, have a special status among belief statements. But this special status does not imply that there is anything irreducibly first person, or <u>de se</u>, about the propositional attitude states themselves, or their objects. On the contrary, the specialness of such attributions consists simply in the fact that they can always appropriately be construed <u>de re</u>. So, if an agent believes that he is F, and if 'b' is a demonstrative expression like 'that man', referring to the agent, then the agent does not believe (<u>de re</u>) that that man is not F. The lesson here is that perceptual acquaintance does not suffice to underwrite <u>de re</u> construals of belief attributions, as should, perhaps, have already been clear from earlier discussion.

FOOTNOTES

1. These graphs should not be taken too literally. No suppositions are made regarding the cardinality of D or W, ordering relations on D or W, or the character of functions represented, except with respect to constancy and non-constancy.

2. This observation has been made by Bernard Linsky, in conversation.

3. That is, if $c \in \mathfrak{C}$ is a term whose intended interpretation is as an indexical. I do not mean to suggest that the members of $\mathfrak{C}$ are intrinsically indexicals or proper names.

4. Lewis (1979); Perry (1979).

5. Stalnaker (1981); Boër and Lycan (1975).

6. See Chapter V, §1.

7. The problem of indexical belief which I am about to discuss is not precisely that raised by any of the authors just cited, but rather, is a problem presented, as it seems to me, by certain of their examples, for the sort of approach I want to take.

CONCLUDING REMARKS

Two themes are prominent in the forgoing discussion of the semantics of belief. First of all, *de dicto* construals of belief statements have been treated, at an intuitive level, as involving a kind of semantic ascent, with respect to (items in) the complement clauses of such statements. Second, belief statements have been likened to statements containing indexicals (especially personal pronouns). Both of these points require comment.

1. *De Dicto* Belief and Semantic Ascent

It was remarked in Chapter I that the agent who believes, *de dicto*, that b is F, in effect believes the proposition *that 'Fb' is true*. While it seems to me that there is some truth in this, it needs qualification. In particular, I want to resist the conclusion that *de dicto* belief is simply meta-linguistic belief. The story is a bit more complicated than that.

Because diagonal propositions are a function of particular agents' epistemic circumstances, there is no unique diagonal proposition associated with any given sentence, belief in which can be identified with *de dicto* belief. Different values of A determine different propositions as $[\ Fb\ ]^{\Gamma}A$. A, recall, is to be thought of as containing those worlds that an agent is unable to distinguish from the actual world with respect to the interpretation of certain non-logical items of the language. In short, the belief one attributes to an agent, under de dicto construal of the belief statement, is relative to the agent.

On the other hand, the proposition that 'Fb' is true does not, in any obvious way, depend on the circumstances of particular agents, any more than does the proposition that Fb itself. Semantic facts, like other natural facts, may vary from world to world, but not, typically, in a way that has anything to do with our epistemic access to them. Consequently, the proposition $[\ Fb\ ]^{\ulcorner A}$ (what was in Chapters I and II denoted *C(Fb)) will not in general be the same as the proposition that 'Fb' is true.

Intuitively, the difference is this. The proposition that 'Fb' is true is one which is true at any world in which the referent of 'b' is a member of the extension of 'F', and false in all others. Thus, any interpretation of 'F' and 'b' which meets the minimal constraints on interpretation functions introduced at the outset of Chapter III, is within the domain (loosely speaking) of the proposition.

For example, the proposition that 'London is pretty' is true is true in a world in which 'London' refers to the axiom of choice, and 'is pretty' means equivalent to the well-ordering principle. But such worlds are not relevant to Pierre's beliefs, on the story Kripke tells. Formally, such worlds are not to be members of A, when A is thought of as representing Pierre's epistemic circumstances.

And the definition of $[\ Fb\ ]^{\Gamma A}$ is such as to make it agree with $[\ Fb\ ]^{\Gamma}$ on any worlds not in A.

This shows the sense in which <u>de dicto</u> beliefs have non-linguistic content on the present theory. For the exclusion of certain worlds from A, such as those in which 'London' refers to the axiom of choice, reflects the fact that, however unclear he may be about the referent of 'London', Pierre knows perfectly well that it's a city, perhaps that it's a major financial centre, the capital of England, and so forth. And so Pierre's <u>de dicto</u> belief that London is pretty differs from the mere belief that 'London is pretty' is true, and also perhaps from the <u>de dicto</u> belief that London is pretty, of an agent who doesn't know that London is the capital of England, or that it's a large city.

This feature of the present approach is among its principal virtues, in my view; and it is a virtue which (as was argued in Chapter II) cannot be matched by any theory which maps sentences onto determinate propositional objects, without consideration of the particular circumstances of the subject of the belief.

Despite this relativity of diagonal propositions to particular agents, and the concomitant non-linguistic content of <u>de dicto</u> beliefs, there is an important sense in which believing <u>de dicto</u> that b is F, <u>is</u> believing that 'Fb' is true. <u>Within</u> the realm of the agent's ignorance or confusion about the reference of 'b' (or the meaning of 'F') the two are the same.

To bring these points into sharper focus, let us augment the language £* with a predicate 'T' and a term-forming operator '⌜ ⌝', and call the expanded language $\mathcal{E}^*_T$. The set of sentences of $\mathcal{E}^*_T$ is the smallest set meeting the following conditions.

i) every sentence of $\mathcal{E}^*$ is a sentence of $\mathcal{E}^*_T$;

ii) if Φ is a (closed) sentence of $\mathcal{E}^*_T$, then so is $T^\ulcorner\Phi^\urcorner$;

iii) the set of sentences of $\mathcal{E}^*_T$ is closed under truth-functional compounding and (non-vacuous) quantification, in the normal way.

So, for example, $(\exists x)(Fx \rightarrow T^\ulcorner Gb^\urcorner)$ is a sentence of $\mathcal{E}^*_T$, but $(\exists x)T^\ulcorner Fx^\urcorner$ is not.

$$[\ T^\ulcorner\Phi^\urcorner\ ]^\Gamma =_{df} (\iota f \in {}^W 2)(\forall w \in W)(\ f(w) = 1 \leftrightarrow$$

$$\leftrightarrow [\ \Phi\ ]^{\Gamma(i_w/i_{w_o})}(w) = 1).$$

'T' is thus a truth predicate for $\mathcal{E}^*_T$.[1]

Notice, first, that $[\ T^\ulcorner\Phi^\urcorner\ ]^\Gamma$ will not in general be the same as $[\ \Phi\ ]^\Gamma$; indeed, it will not in general be the case that either is a subset (i.e., a sub-proposition) of the other. So an agent might believe either proposition without believing the other. However, $[\ \Phi\ ]^{\Gamma(i_w/i_{w_o})}(w) = [\ \Phi\ ]^{\Gamma A}(w)$ for any $w \in A$, and so, in particular, $[\ T^\ulcorner\Phi^\urcorner\ ]^\Gamma = [\ \Phi\ ]^\Gamma A$, for $A = W \in \mathcal{P}_{w_o}(W)$. Since the possible values of A are all subsets of W, $[\ \Phi\ ]^{\Gamma A}$ always coincides with $[\ T^\ulcorner\Phi^\urcorner\ ]^\Gamma$ on A. This shows the sense in which agents' de dicto beliefs are "meta-linguistic". The sense in which they are not is reflected in the fact that it is not generally true that $[\ \Phi\ ]^{\Gamma}A = [\ T^{\ulcorner\Phi\urcorner}\ ]^{\Gamma}$.

It has been observed (e.g., by Kaplan, and Kripke)[2] that, normally, agents use singular terms (for example) with the intention of referring to the same things as do others of their linguisitc community; i.e., with the intention of achieving "standard" reference, despite possibly ideosyncratic notions about the referents of the terms. Consequently, the differences between <u>de dicto</u> and <u>de re</u> belief are, as it were, invisible to the subject of belief (or, echoing Descartes, there are no sure signs by which one may know that one's belief is <u>de re</u>).

This feature also is accounted for on the forgoing characterization of <u>de dicto</u> belief, according to which the propositional objects of <u>de dicto</u> and <u>de re</u> belief coincide on W - A (the complement of A), and diverge only within A. But A just is the set of worlds the agent in question cannot distinguish from the actual world, with respect to the interpretation of certain non-logical items in the language. So, naturally, divergence from the standard interpretation of those items, within A, is something the agent will not recognize.

The definition of $[\ T^{\ulcorner\Phi\urcorner}\ ]^{\Gamma}$ was, in a certain sense, indirect, being based on the values of 'Φ' in certain models, rather than proceeding directly by defining the interpretation of 'T', and of '⌜Φ⌝'. As a result, non-standard interpretations of 'T' are ruled out. This seems appropriate, for much the same reason as was given in the cases of the quantifiers, and indexicals. As in those cases, a correct understanding of the predicate 'true' seems to lie close.

to the heart of linguistic competence.

Because of the exclusion of non-standard interpretations of 'T', $[\ T\ulcorner\Phi\urcorner\ ]^{\Gamma}$ will always be the same as $[\ T\ulcorner\Phi\urcorner\ ]^{\Gamma A}$, for any (admissible) A. It was remarked above that $[\ T\ulcorner\Phi\urcorner\ ]^{\Gamma} = [\ \Phi\ ]^{\Gamma A}$, when A = W. It follows that $[\ T\ulcorner\Phi\urcorner\ ]^{\Gamma A} = [\ \Phi\ ]^{\Gamma A}$, or in other words, that $[\ T\ulcorner\Phi\urcorner \leftrightarrow \Phi\ ]^{\Gamma A}$ is the necessary proposition, for A = W, even though $[\ T\ulcorner\Phi\urcorner \leftrightarrow \Phi\ ]^{\Gamma}$ is not. Hence, $[\ T\ulcorner T\ulcorner\Phi\urcorner \leftrightarrow \Phi\urcorner\ ]^{\Gamma}$ is the necessary proposition. That is, the "Tarski sentences" are not, in general, valid on the present semantics. Still, the proposition that a given Tarski sentence is true, is necessary, and so cannot be doubted-- one cannot doubt that the actual world is one where '$T\ulcorner\Phi\urcorner \leftrightarrow \Phi$' is true.

A parallel to this feature of the treatment of the truth predicate is of some interest in the examination of indexical belief. Let us expand the language $\mathcal{L}^{*}_{T}$ (without, this time, renaming it) by adding an operator 'R' and a clause concerning it to the definition of "sentence of $\mathcal{L}^{*}_{T}$".

ii)' if α is an individual constant of $\mathcal{L}^{*}_{T}$, then $R\ulcorner\alpha\urcorner$ is a singular term of $\mathcal{L}^{*}_{T}$.

$$[\ R\ulcorner\alpha\urcorner\ ]^{\Gamma} =_{df} (\iota f \in {}^{W}D)(\forall w \in W)(\ f(w) = d \leftrightarrow$$
$$\leftrightarrow i_{w}(\alpha)(w) = d).$$

'$R\ulcorner a\urcorner$' can be read, "the referent of 'a'", and sentences like '$R\ulcorner a\urcorner = b$' can be read, "'a' refers to b".

The parallel with truth is this. $[\, R\ulcorner\alpha\urcorner = \alpha \,]^{\Gamma}$ is not the necessary proposition, but $[\, T\ulcorner R\ulcorner\alpha\urcorner = \alpha\urcorner \,]^{\Gamma}$ is. Furthermore, $[\, R\ulcorner\alpha\urcorner = \alpha \,]^{\Gamma}A$ coincides with $[\, T\ulcorner R\ulcorner\alpha\urcorner = \alpha\urcorner \,]^{\Gamma}$ on A. So, for example, an agent might know that "'Fred' refers to Fred" is true, and believe de dicto that 'Fred' refers to Fred, without knowing (de re) who 'Fred' refers to. The agent might not be able to distinguish from the actual world, a world in which 'Fred' refers to George. But in such a world, the sentence "'Fred' refers to Fred" will express the proposition that 'Fred' refers to George, which, in that world, is true.

Now consider Lingens, the famous amnesiac reading his own biography in the Stanford Library (and believing what he reads). According to Lewis, following Perry, Lingens could know every true proposition about Lingens, and still not know that he is Lingens. Thus, indexical knowledge (or belief) is not reducible to "propositional" knowledge (belief).

But (I claim), the case is misdescribed. What Lingens knows is that a certain body of statements involving 'Lingens' is true, among them, no doubt, the statement "'Lingens' refers to Lingens". What he does not know is that 'Lingens' refers to Lingens. Since he is Lingens, knowing that 'Lingens' refers to Lingens would be knowing that it refers to him. So his epistemic defect is, after all, a failure to know the truth of certain true (non-indexical) propositions. In short, once one admits knowledge (or lack thereof) of such propositions as that 'a' refers to a, and recognizes that this proposition is not the same as the proposition that "'a' refers *to a" is true*, and need not be the propositional object of the *de dicto* belief that 'a' refers to a, the arguments for irreducibly indexical belief begin to fall apart.[3]

2. Pragmatic Ambiguity and the Primacy of *De Re* Construals

The language $\mathcal{E}^*_I$ is a pragmatic language in the sense of Montague, and the model theory described above makes it non-trivially so. There are items in the language, in particular the operator 'B', whose semantic role is governed, in part, by indices in the models which may vary independently of other features of the models.

But there is a further sense in which the present account of belief might be termed "pragmatic". It has already been remarked that Montague's pragmatics might more properly be seen as a central part of an adequate theory of the semantics of certain expressions, such as tenses and other indexicals. This is reflected in the fact that the context sensitivity of such expressions is "built into" the interpretation functions that are an element of every model. By contrast, the index to which the belief operator is sensitive, on the present account (the parameter A), is, so to speak, external to the interpretation function of a model— it is in fact defined over classes of such functions. The pragmatics of belief, in other words, is an extra-semantic feature of our models, in a sense which contrasts with Montague's pragmatics (of indexicals), which is infra-semantic.

In actual discourse it may happen that features of the context of discourse, which are represented in the model theory by indices of various sorts, are not known, and perhaps, cannot be determined. In such cases the precise content of an utterance may be in doubt (as when, for example, one answers the phone and hears an unfamiliar voice say, "Hi, it's me"). Although circumstances may preclude determining what proposition has been expressed by the utterance, the sentence uttered is not thereby rendered semantically ambiguous. Still, one may wish to say that the utterance is in some other sense ambiguous.

It is sometimes claimed that belief statements are ambiguous between *de dicto* and *de re* construals, and this is treated as a semantic ambiguity in the interpretation of "believes". On the present account, the interpretation of the operator 'B' is univocal—it has a determinate extension in each world, which may contain or exclude any agent/proposition pair, but not both. It is semantically unambiguous.

It is moreover determinate in a way in which indexicals are not, for the extension of the operator in a world, on a model, depends on no other indices, no other features of context. On the other hand, the parameter A, representing agents' epistemic circumstances, can reflect different readings of a belief statement, for values of A that diverge in specified ways from the designated standard interpretation function of a model. The felt "ambiguity" that sometimes results is thus not semantic, but pragmatic.

It may be thought that a further ambiguity arises in the case where $A = \{w_0\}$. Is this to be taken as indicating a *de re* construal of the belief statement in question, or a <u>de dicto</u> construal, relative to an agent who happens not to be confused or ignorant about the intensions of the relevant terms? Or to put the question somewhat differently, when do such cases constitute attribution to an agent of semantic knowledge?

In fact, it seems to me that this distinction isn't <u>semantically</u> significant. The truth conditions of belief statements, and their inferential properties, ought to be (and are, on the present theory) indifferent to such matters.

It was suggested early on that the <u>de re</u> construal of belief statements enjoys a priviledged status, in that it is the natural construal in the absence of contrary indications. In fact, features of the context in which a belief attribution is made may be such as to incline one toward one construal or another, with varying degrees of force. (The features of context may include such things as the particular choice of words used in making the attribution-- more generally, any discernable aspect of the mode of attribution may be relevant to determining the content of the belief attributed.)

The point that needs to be made is that whatever these contextual features suggest as the appropriate construal, and however strongly or weakly they suggest it, the <u>de re</u> construal has a special status-- it is after all the construal under straightforward semantic interpretation of the embedded sentence in question. The readings of a belief statement that diverge from the <u>de re</u>, although frequently the intended ones, are semantically anomalous (under a rather strict notion of semantic norms).

The anomaly is rather like that which accompanies violations, or apparent violations, of Gricean maxims governing discourse. In both cases, extra-semantic aspects of context invest statements with a kind of import beyond their strictly semantic import; and this further import will not uncommonly constitute the real point of the statement, in its context.

Grice's analysis of implicature allows him to accomodate certain pervasive features of language use, while retaining a "classical" formal conception of the logical devices of natural language. The forgoing analysis of belief statements is meant to achieve a precisely parallel result.

FOOTNOTES

1. $\mathcal{L}^*_T$ has rather limited expressive powers, for a language containing a truth predicate. Clause ii) of the definition of "sentence of $\mathcal{L}^*_T$" excludes such things as 'Tb' (for $b \in \mathbb{C}$), and '$(\exists x)Tx$'. This may seem undesirable since, whatever semantic anomaly attaches to natural language expressions like "London is true", the latter does not seem ill-formed. Moreover, if $(\exists x)Tx$ is not a sentence, EG does not apply unrestrictedly, and this too is somewhat counterintuitive-- the inference from $T(\phi)$ to $(\exists x)Tx$ is, on the face of it, as reasonable as other instances of EG. This somewhat narrow definition is adopted to avoid semantic paradoxes, which are not a present concern, and because $\mathcal{L}^*_T$ suffices to illustrate the relation between <u>de dicto</u> belief, and "believing-true" a sentence, which is.

2. Kaplan (1977); and Kripke (1979).

3. The example of Lingens is just one of several that Lewis employs. The others are not precisely equivalent, but nonetheless, succumb to similar treatment. The formal treatment just sketched is essentially equivalent to that proposed, informally, by Stalnaker, in "Indexical Belief".

4. Grice (1975).

BIBLIOGRAPHY

Biro, John (1984). "What's in a Belief?", Logique et Analyse 107: 267 - 282.

Boër Steven, and William Lycan (1975). "Knowing Who", Philosophical Studies 28: 299 - 344.

--------- (1980). "Who, Me?", Philosophical Review 89: 427 - 466.

Cresswell, M. J. (1975). "Hyperintensional Logic", Studia Logica 34: 25 - 38.

--------- (1985). Structured Meanings (Cambridge: MIT Press).

Davidson, Donald, and Gilbert Harman, eds. (1972). Semantics of Natural Language (Dordrecht: D. Reidel Publishing Company).

Dugundji, J. (1940). "Note on a Property of Matrices for Lewis and Langford's Calculi of Propositions", Journal of Symbolic Logic 5: 150 - 151.

Field, Hartry (1978). "Mental Representation", Erkenntnis 13: 9 - 61.

French, Peter, Theodore Uehling, and Howard Wettstein, eds., (1979). Contemporary Perspectives in the Philosophy of Language, (Minneapolis: University of Minnesota Press).

Grice, H. P. (1975). "Logic and Conversation", Syntax and Semantics 3: 41 - 58.

Kaplan, David (1977). Demonstratives, manuscript.

--------- (1978). "On the Logic of Demonstratives", in French et al. (1979).

Kripke, Saul (1963). "Semantical Considerations on Modal Logic", Acta Philosophica Fennica 16: 83 - 94; reprinted in Linsky (1971).

--------- (1979). "A Puzzle About Belief", in A. Margalit, ed., Meaning and Use (Dordrecht: D. Reidel Publishing Company).

Lewis, David (1979). "Attitudes De Dicto and De Se", Philosophical Review 88: 513 - 543; reprinted in Lewis (1983).

--------- (1980). "Index, Context, and Content", in S. Kanger and S. Öhman, eds., Philosophy and Grammar (Dordrecht: D. Reidel Publishing Company).

--------- (1981). "What Puzzling Pierre Does Not Believe", Australasian Journal of Philosophy 59: 283 - 289.

--------- (1983). Philosophical Papers (New York: Oxford University Press).

Linsky, Leonard, ed. (1971). Reference and Modality (London: Oxford University Press).

Marcus, R. (1981). "A Proposed Solution to a Puzzle About Belief", Midwest Studies in Philosophy VI: The Foundations of Analytic Philosophy (Minneapolis: University of Minnesota Press).

--------- (1983). "Rationality and Believing the Impossible", Journal of Philosophy 80: 321 - 338.

Montague, Richard (1968). "Pragmatics", in Montague (1974).

--------- (1970a). "Pragmatics and Intensional Logic", in Montague (1974).

---------- (1970b). "The Proper Treatment of Quantification in Ordinary English", in Montague (1974).

---------- (1974). Formal Philosophy, Richmond Thomason, ed. (New Haven: Yale University Press).

Perry, John (1977). "Frege on Demonstratives", Philosophical Review 86: 474 - 497.

---------- (1979). "The Problem of the Essential Indexical", Noûs 13: 3 - 21.

Segerberg, K. (1973). "Two-Dimensional Modal Logic", Journal of Philosophical Logic 2: 77 - 96.

Stalnaker, R. (1972). "Pragmatics", in Davidson and Harman (1972).

---------- (1976). "Propositions", in MacKay and Merrill, eds., Issues in the Philosophy of Language, (New Haven: Yale University Press).

---------- (1978). "Assertion", Syntax and Semantics 9: 315 - 332.

---------- (1981). "Indexical Belief", Synthese 49: 129 - 151.

---------- (1984). Inquiry, (Cambridge: MIT Press).

Thomason, R. (1980). "A Model Theory for Propositional Attitudes", Linguistics and Philosophy 4: 47 - 70.