Order Number 9009778

Semantics for belief attributions

Savion, Leah, Ph.D.

City University of New York, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

SEMANTICS FOR BELIEF ATTRIBUTIONS

by

LEAH SAVION

A dissertation submited to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1989

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

May 5, 1989
Date

Chair of Examining Committee

May 5, 1989
Date

Executive Officer

Supervisory Committee

The City University of New York

Acknowledgements

I thank my readers, Professors Charles Landesman and Stephen Spielman, and especially my advisor, Professor Arnold Koslow, for their valuable comments on earlier drafts of this dissertation. I am eternally grateful to Anil Gupta, my husband, without whose assistance this work would have been completed years ago.

TABLE OF CONTENTS

Chapter 1

Introduction

SECTION I

A familiar yet problematic feature of belief contexts is that the complement sentence p in 'a believes that p' cannot always be replaced by an intensionally equivalent sentence q without disturbing the truth value of the whole. For example, the sentence "2+2=4" has the same intension as "there is no largest prime number", but the sentence

(1) Jim believes that 2+2=4

may be true, yet

(2) Jim believes that there is no largest prime number

may be false. This familiar feature of belief creates problems both for semantic theory and for a metaphysical account of the notion of belief.[1]

The problem for semantic theory concerns one of its fundamental principles -- **Compositionality**, the principle that the meaning of the whole is explained as a function of the meanings of the parts. Once meanings are assigned to the primitive components of an expression, the semantic

---

1. The metaphysical aspect of the problem will be discussed later.

theory supplies rules using which the meaning of the complex expression can be calculated; and using this functional dependence, properties such as consistency, implication, incompatibility etc., can then be explained. For instance, if the theory takes the meaning of a (declarative) sentence to be its truth value, the meaning of the functor 'and' can be viewed as a function which takes the truth values of two sentences p, q, and yields a truth value (a function that is familiar from truth functional logic). Using that function, one can explain why the sentence 'p and q' implies p.

Note that the principle of compositionality, which is usually attributed to Frege, is indifferent to the choice of how to individuate meanings. The principle puts only the following abstract constraint on any semantic theory: whatever meaning is taken to be, the meaning of the whole is a function of the meanings of the parts (this is often refered to as Frege's principle).[2] The problem that belief contexts create for semantic theory is that standard ways of treating meanings imply, in virtue of Frege's principle, that the above kind of substitution in belief contexts must preserve truth -- a result that goes against the intuitive

---

2. A 1-place function is defined as a set of ordered pairs, the first of which is the argument and the second is the value, such that there is a unique value for every argument. For instance, the following set {<1,2>,<2,3>,<3,4>} is a function, since for each of the objects 1, 2, 3, there is a unique value associated with it. The following set {<1,2>,<1,3>...} is not a function, since the argument 1 does not have a unique value.

behavior of these contexts. This happens regardless of whether meanings are given an extensional or an intensional account.

Extensional semantics identifies the meaning of a singular term with the individual it refers to; the meaning of a predicate with its extension; and the meaning of a sentence with its (absolute) truth value. In light of Frege's principle, extensional semantics implies that all contexts are extensionally transparent: co-referential terms, co-extensional predicates, and materially equivalent sentences are (correspondingly) interchangeable salva veritate. Some contexts, such as those generated by truth functional connectives (e.g., 'and') are extensionally transparent. But, extensional semantics forces one to construe all contexts as extensionally transparent.

Intensional semantics identifies the meanings of expressions with their intensions. The intension of an expression, on one major approach, is identified with a function which assigns to each possible world the extension of that expression in that world. So, on this account, the meaning of a singular term is a function which assigns to each possible world an individual; the meaning of a predicate is a function which assigns to each possible world the extension of that predicate in that world; and the meaning of a sentence is a function from possible worlds to the truth values "The True" and "The False". The meaning of

a sentence can also be regarded as the set of possible worlds in which the sentence is true. (The two conceptions are equivalent.) The intension of a sentence is often referred to as a proposition. Intensional semantics implies that all contexts are intensionally transparent: co-intensional expressions are interchangeable without affecting the truth value of the original sentence (in any world). A modal context is intensionally transparent. A context which is not intensionally transparent is called intensionally opaque. In general, any context which is intensionally opaque is also extensionally opaque, but not vice versa.

By Frege's principle, if p and q are assigned the same meaning, then the meaning of the sentences 'a believes that p' and 'a believes that q' should also be the same. That means that if meaning is construed extensionally, and the complement sentence is taken to be semantically significant, then believing one truth implies believing all truths. Alternatively, if meaning is construed intensionally, believing one truth implies believing all logically equivalent truths. In belief contexts, substitution of extensionally identical expressions obviously does not preserve truth, but neither does the substitution of intensionally equivalent sentences always preserve truth. For instance: all mathematical truths are identified in intensional semantics with the same proposition (since they

are all true in all the worlds), but one may believe one mathematical truth without believing another.

In short, the problem that belief poses for semantic theory is this: traditional semantics seems to imply that all contexts are intensionally transparent, but belief contexts seem to be, intuitively, intensionally opaque. The most natural tendency is to understand this difficulty as a challange to fundamental issues of semantic theory: the issue of the logical vs. the grammatical form, the account of meaning, and the principle of compositionality.

Let us concentrate, for now, on theories that both accord with Frege's principle, and accept the empirical data about believers, (i.e., data that is responsible for the intensional opacity of belief contexts -- hence we shall be concerned with actual, nonideal, believers). If the logical form of a sentence is supposed to reflect major features of its standard grammatical form, then a sentence like

(3) a believes that p

should be regarded as consisting of a two place predicate ('believes'), and two names ('a', 'that p').[3] This form treats the complement sentence 'p' as a genuine component of the whole, and we are confronted by the difficult task of individuating its meaning so as to account for the intensional opacity of the sentence. Denying that

---

3. Hence -- belief has an object, denoted by 'that p'.

'believes' has a definite object entails the assignment of a logical form which is divorced, to one extent or another, from the usual grammatical form, a move which would normally be considered as one that calls for justification.

The standard approach to the problem of intensional opacity of belief sentences can be seen as following one of the two major strategies: (I) deny that p in (3) is a genuine semantic component of the sentence; (II) regard the complement sentence as a semantically significant component, but individuate meanings more finely than in intensional semantics. Either one of these strategies will preserve Frege's principle and overcome the problem of intensional opacity. In the dissertation I will consider critically a number of theories that follow these strategies. Section II provides an overview of the main theories in the literature.

SECTION II

The theories that follow strategy (I), by which the complement sentence of a belief attribution does not have a genuine semantic occurrence, can conveniently be grouped according to the logical form they assign to belief sentences.

(I.1) Quine's theory: A belief attribution 'a believes that p' has a subject-predicate form, a form that it shares

with sentences such as 'Bob is fat', 'Alfred is a man' and the like. This theory views 'believes that p' as a primitive one-place predicate. Thus the expressions 'believes that p' and 'believes that q' are completely distinct for distinct p, q, and have no semantically relevant parts in common. The logical relations that p and q may bear to each other outside belief contexts are irrelevant to their behavior in belief contexts, and so the intensional opacity is accounted for by simply blocking all substitutions.

(I.2) Linguistic theories: belief attributions are relational: they involve a two-place predicate ('believes') and two names. The second name 'that p', however, does not contain p as a genuine constituent. It denotes a sentence type or a sentence token, or an utterance, or an inscription, etc. (These theories are sometimes referred to as following the Direct Approach to the "objects of belief".)

This type of theory clearly overcomes the problem of intensional opacity: the substitution of p by a different sentence q results in names for different sentences, so the substitution need not preserve truth regardless of any relationships between the meanings of p, q. These meanings do not participate in the compositional value of the whole sentence.

The direct approach offers a large spectrum of "sentential" theories (see Chapter 6) centered on fine individuation of the objects of belief via some syntactic correlation it has to the complement sentence. The theories range from those that assign to objects of belief purely syntactic identity conditions (e.g., they identify the object of belief with the sentence token used to describe it), to those that recognize semantic elements as well as syntactic ones in the individuation of the objects of belief.

(I.3) Russell's theory: belief attributions are analyzed as consisting of a name, and a multiple-place predicate, which relates the believer to several entities. E.g., in the sentence 'Bob believes that Alfred is fat' the verb 'believes' is viewed as a three place predicate relating Bob to Alfred and to the property of being fat; in 'Bob believes that Alfred and Mary are fat', 'believes' is a four place predicate, relating Bob to Alfred, Mary and the property of being fat. Note that the complement sentence itself is not considered a genuine semantic component of the belief attribution.

On this theory, Frege's principle is not violated since by dissecting the complement sentence into several semantic components, the theory dispenses with "objects of belief", so the substitution of intensionally equivalent sentences is blocked. A whole host of other substitutions (as well as

existential generalizations) which are responsible for the opacity of belief contexts are not that lucky: Bob's belief that Jekyll is a doctor ends up being identical to the belief that Hyde is a physician, whether or not Bob is aware of the coreferentiality of the names and the synonymity of the predicates, since the relation of belief hold between him and the referents themselves.

Theories that follow strategy (II), i.e., those that consider the complement of a belief attribution to be a genuine semantic constituent, can be classified according to their treatment of the notion of proposition. (These theories are sometimes referred to as taking the Indirect Approach since the object of belief is only indirectly expressed in the complement sentence).

(II.1) Thomason's theory: postulates a proposition as a primitive undefined notion. Distinct complement sentences express, by this account, distinct propositions. Such a move answers to the question of intensional opacity by blocking all substitutions in belief context, as in Quine's theory. (Thomason's theory, though, has some benefits resulting from having objects of belief, which Quine's account does not admit: by (II.1) but not by (I.1) one can infer from 'a believes that p' that a believes something.) By this account propositions are individuated extremely finely -- no two distinct sentences express the same proposition.

(II.2) Carnap's theory: propositions are structured meanings. The proposition expressed by a sentence is built using the meanings of the parts and the semantic structure of the sentence. On this account, propositions are individuated finely, but not so finely as in (II.1) -- distinct sentences may express the same proposition. Sentences express the same proposition, on this theory, iff they are (in Carnap's terminology) intensionally isomorphic. The treatment of objects of belief as complex semantic structures overcomes the problem of opacity associated with the substitutivity of intensionally equivalent sentences, but calls for the help of additional devices in coping with substitutivity of co-intensional singular terms and predicates. A theory of this type will be examined in Chapter 5.

(II.3) Stalnaker's theory: propositions are viewed according to the possible world analysis: they are identified with sets of possible worlds. Thus two sentences express the same proposition iff they are necessarily equivalent. Being sets of worlds, propositions have no syntactic properties and do not mirror the different ways a language may have of conveying the same content. Being so coarsely defined, propositions seem completely unqualified for the job of being the objects of beliefs, a claim Stalnaker attemts to refute. In his defence of (possible worlds) propositional belief, Stalnaker resorts to devices of separate belief states, complicates "expresses" relations

that holds between a sentence and a proposition, and appeals to context dependency to account for opacity, contradictory beliefs, etc. This theory is explained and examined in Chapters 3 and 4.

The second strategy, by which the complement sentence of a belief attribution is viewed as a genuine semantic constituent, seems to me more promising. However, I believe that none of the three types of theories outlined under that strategy is adequate. In the next section I will sketch the kind of approach that may be useful for dealing with the problem of belief.

SECTION III

I wish to argue that the inadequacy of the current theories is due primarily to a common faulty strategy: they mistakenly identify the **objects of belief**, i.e., the denotation of 'that p', with **meanings**. Some theories hold this identity in a strong form: the denotation of 'that p' is identified with the meaning of p. That implies that the object of belief can be directly recovered from the semantics of the sentence p. Others take the identity to hold between the kind of entities that meanings and objects of belief are. On this view, 'that p' denotes a meaning-like entity, though it cannot directly be recovered from the meaning of p alone. Stalnaker, for example, holds that the

object of the belief attributed in '_a_ believes that _p_' is a proposition, but not necessarily the proposition that _p_ normally expresses. The proposition believed cannot then be obtained from _p_ by semantic rules alone.

I suggest that the objects of belief cannot be identified with meanings. The object of belief that is attributed in '_a_ believes that _p_' is not identical to the meaning of _p_, nor is it identical to the meaning of any other sentence. Meanings and objects of belief, I want to argue, are different kinds of entities.

How do we go about demonstrating that two abstract notions are of different kinds? Though this task has been a major preoccupation of philosophers for generations, it has not been made easier. Since the notions of "meaning" and "object of belief" are theoretical notions, one way to distinguish between them is to show that the functions these two notions serve are different to such an extent that no (one kind of) notion can successfully fulfill them all. I cannot offer a conclusive argument of this sort, but I believe that I can motivate the idea that the two notions should be distinguished.

The central role of the concept of **meaning** is to help explain the semantic properties of the language. The major constraint on such an explanation is that it should account for the logical and semantic properties of expressions in a compositional way. Thus it should account for synonymity of

sentences, consistency, implication, incompatibility etc. The major benefit, from a philosophical point of view, of employing the notion of meaning in semantic theory is that it enables one to capture that aspect of sentences that is responsible for these semantic properties.

The role of the concept of **object of belief**, on the other hand, is to help explain basic properties of certain mental states. The main purpose of assigning objects of belief (and of attitudes in general) is to characterize states of the mind in a way that specifies their causal and functional role with respect to each other, to stimuli and to behavior. The major desideratum of employing this notion is to classify belief states according to the cognitive role they play for an agent.

Thus, the role of the objects of belief is clearly distinct from that of meaning. One may argue, however, that it is still possible that the same entity may serve both roles. This possibility is hard to rule out absolutely, but I believe that there are considerations that suggest that entities suitable to be objects of belief are of a different kind than meanings.

A point in favor of the distinction between the two notions could be made by showing that two distinct (in meanings) complement sentences of belief attributions can be used to ascribe the same belief, and/or that the same meaning (expressed by tokens of one eternal sentence) can be

used to attribute distinct beliefs to different agents or on different occasions. Objects of belief are used to characterize a belief state in terms of the informational content of that state. For instance, when we ascribe to Fido, the dog, the belief that there is a bone buried under the tree, we actually attribute to Fido a bit of information or misinformation, which also serves, together with other bits of information and desires that Fido has, to explain and predict the dog's behavior. This bit of information is different from the one we ascribe to the dog's owner, Fred, when we say that Fred believes that there is a bone buried under the tree, although the complements of the two belief attributions are the same.

The belief state ascribed to Fred serves to explain much more complex kinds of behavior than the one ascribed to the dog. For instance, it may function in the explanation of Fred's utterance that there is an organ of a dead body buried under the tree, and the like. The beliefs attributed to Fido and to Fred are, therefore, different. Hence the objects of the two beliefs cannot be the same, regardless of the fact that they are ascribed by sentences with identical components.

Beliefs, like sentences or propositions, are true or false; they can be compatible or incompatible with each other, etc. The fact that objects of belief have some semantic properties has perhaps been responsible for their

identification with the meanings of sentences. However, the characterization of a belief state by meanings alone is similar to the characterization of a computer by its physical dimensions alone: the most important properties are left out. The concept of meaning, which is designed to play a specific role in semantic theory, is not capable, in my opinion, of accounting for our basic intuitions about the concept of belief.

The semantic analysis of belief sentences that I will develop considers, as mentioned earlier, the complement sentence as having a genuine occurrence in the belief attribution, but its meaning is not to be identified with the object of the belief attributed. Let us call objects of belief "Informational Contents",[4] to distinguish them from meanings. The task of constructing a theory on this plan consists of giving an account of the following:

- Informational contents
- How do informational contents characterize the belief states
- The relationship between the complement of the belief attribution and the informational content attributed

The way I believe one should go about searching for the as yet unfound objects of belief, i.e., for informational

4. The way I understand this notion is different from that of other theorists, e.g., Dretske.

contents, is by going from the data about the cognitive role of belief to the theoretical notion that can best be used to classify that data. At this stage in the development of my theory I can offer only a few sketches of what this strategy would yield, sketches that may not yet stand up to close scrutiny.

It is clear in the first place that informational contents are:

(i) abstract entities (e.g., they are not inscriptions, utterances, and the like)

(ii) language independent (they are not, for example, sentence types)

(iii) they can be evaluated as true or false with respect to the state of affairs they are about.

In order to determine further properties of informational contents let us pursue the question: "what is needed to characterize a belief state?"

A belief state is characterized by the functional cognitive role it plays in the organism's psychology, and how we should think of informational content is determined by aspects of this role. One of these aspects concerns the role of belief in our reasoning (e.g., interactions with other beliefs to produce new beliefs, or with other states to produce tendencies to act, etc.), and the other with the

relation of our beliefs to the world (by which the beliefs are judge true or false).

The role our beliefs play in reasoning is an essential part of our folk psychological conception involved in the attribution of beliefs and in the explanation of our acts. This role should be viewed as embodying not only the resulting actions but also the whole process of reasoning that beliefs stimulate in our cognitive systems.[5] Consider the two belief ascriptions:

(4) Jim believes that Robin will win

(5) Jim believes that everyone who does not compete, or loses, will have done something Robin will not have done.[6]

The two ascriptions relate Jim to two distinct belief states: the second one ascribes to Jim a state that requires some logical calculation and is used in different logical inferences than the first. The informational content is thus sensitive to the complexity of the belief attributed because the functional role of a belief state is partially determined by the logical moves in which a believer utilizes his belief.

---

5. Sets of possible worlds could be used to characterize beliefs when outcome behavior alone is considered. Such accounts are inadequate for agents in whose mental life reasoning plays an important role.

6. This example is due to John Bigelow.

The second aspect of belief states mentioned above, i.e., their relation to the world, is best approached by considering the thesis of "methodological solipsism"[7], which analyzes mental states as if the agents were sealed off from the world. For instance, imagine a twin-earth planet, where all facts are identical to those on our planet, except that XYZ is present there wherever water is present here (though the same name is used for the two substances).[8] The complement sentence of a belief attribution 'water is wet' picks out, by the solipsistic approach, the same belief for me and for my doppelganger, since these beliefs function identically in their interactions with other beliefs, desires etc. I wish to argue that these beliefs are not the same, since they are about different things: one of them is about water, and the other is about XYZ (the reference relation of "aboutness" is made within the semantics of English as used on earth). The point can also be illustrated by a simpler example: the beliefs ascribed by the complement 'some windows are open' may cause two believers, who dwell in two different houses, to act (inferentially and behaviorally) in the same way under certain circumstances. (We assume that they share the same beliefs about the damage that a rain storm may cause carpets and the same desire to prevent the damage.) So, according

---

7. This thesis was first put forward by Jerry Fodor.

8. This well known example is due to Putnam.

to methodological solipsism the beliefs are the same, but they clearly differ in what they are about.[9]

So, the functional role of a belief state is determined, in my opinion, both by its broad role in the agent's reasoning and by referential relations between the belief and the world outside the believer. The object of belief should be such that it determines the proper aspects of the functional role of belief states as, for example, the notion of "sharpness" characterizes the utility function of a knife. Characterizing a kitchen knife by its color alone fails to indicate, in a normal context, the functional expectations associated with a knife. Similarly, the characterization of a belief state either by an object of belief that provides only the means for semantic judgment (e.g., a set of possible world or a semantic structure), or by an object which indicates only potential solipsistic interactions of "narrow contents" (as in mental sentence theories) would fail on similar grounds.

I hope that the above gives a rough idea of the notion of informational content and how it can be developed. I turn now to the relation of the complement of a belief attribution to the informational content ascribed. As illustrated by the twin-earth example, two "identical"

---

9. Some writers have attempted to capture the "sameness" of information involved in such cases by utilizing Kaplan's distinction between "character" and "content" of certain expressions, "stretching" the distinction to apply to sentences other than those which involve indexicals.

descriptions of different states of affairs can be used to pick out different informational contents. As illustrated by the examples of Fido and Fred, and by (4) and (5), two descriptions of the same state of affairs (e.g., necessarily equivalent expressions or even two tokens of the same eternal sentence) can be used to pick out beliefs that are employed in distinct inferential processes, hence have different cognitive function, so by this theory they denote different informational contents. This is evidence that the informational content, denoted by 'that p' on a given occasion, is determined not only by the meaning of p, but also by pragmatic considerations.

The central task of the theory of belief sentences (to be proposed) is to articulate what the relevant pragmatic factors are, and how they contribute to the determination of the particular informational content that an occurrence of 'that p' denotes in a given context. Contextual factors put constraints on the interpretation of sentences containing indexicals. In a similar manner, the relevant pragmatic considerations can be viewed as constraints on the denotation of 'that' clauses in belief attributions.

En route to sketching some of these pragmatic factors, let us entertain one of the major questions that this account should answer: under what conditions do distinct complement sentences (which "mean" the same) ascribe the same belief? What are the presuppositions involved in those

cases where a sentence of the form 'p or q' is assumed to attribute the same belief as does a complement sentence of the form 'q or p'? Are these presuppositions similar to those involved in judging the complement sentence 'Bill will come in a fortnight' to ascribe the same belief as 'Bill will come in two weeks'? Or to those involved in identifying the beliefs ascribed by (4) and by (5)? Obviously not. The reluctance we feel about these identifications in different contexts can perhaps be traced to some assumptions about the believers these beliefs are attributed to; assumptions about the relevant background knowledge they are supposed to have, their conceptual, linguistic and logical abilities and performance.

Very little outside the normal need be presupposed about a person's cognitive state in order to identify the belief attributed by 'p or q' with the one attributed by 'q or p'. Some linguistic knowledge is presupposed when synonymity holds between the only different elements of the complement sentences, while the ability to exercise a certain level of reasoning is presupposed for the identification of beliefs in the Robin example.

In my view, a general empirical theory about believers is implicitly presupposed by all participants in every belief attribution context. Such a theory divides the population of believers into large groups, according to some common assumptions we have about their knowledge and about

their linguistic and logical capacities. It is improper, in a normal context, to attribute to a cat a belief by using a complicated construction as in (5). A similar violation of conversational rules occurs in a context in which I claim that my two years old son believes that number of stuffed bears in the room is an even prime number. Another pragmatic element that is often involved is the contextual purpose of the specific belief attribution. The attributor's choice of which complement sentence to use to ascribe a belief is determined both by his picture of the agent's cognitive skills, and by what the belief ascriptions is expected to achieve in the context. We have different goals when we attribute beliefs: understanding the agent's state of mind; explaining or predicting his assent to a sentence; explaining or predicting his overt behavior; grouping him under some political or ethical convictions. These different purposes have a significant effect on the choice of the sentence used to ascribe the belief, and hence on the relationship that the sentence has to the informational content it denotes.

The relation between a complement 'that p' and an information content thus depends upon several parameters: some of those are of semantic nature; some are of cognitive nature; and some are contextual. The parameters together determine the object of belief ascribed in the use of 'that p'. The factors about the agent's cognitive state, which may threaten the objective status of informational content,

are taken as general properties by which the population of believers is divided into large groups, according to how they are classified by the attributor. For instance, the question of whether the informational content associated with the complement 'Jim is an M.D.' is identical with the one associated with 'Jim is a physician' receives different answers, according to the evaluation of the agent's ability to recognize synonymous terms in the language, and according to whether it is important to consider that ability (as would be if the purpose is to predict his assent, but may not if the purpose is to predict, say, his non-verbal actions).

Suppose that a theory of informational contents along the lines sketched above can be constructed, some difficulties still remain. Of these, perhaps the most urgent one concerns the issue of deductive closure of our beliefs. The problem is that belief attribution makes sense only if we presuppose some minimal rationality in the believer. It would count as a deviant practice to attribute to Jones the beliefs, for example, that if he has four quarters then he has a dollar, and that he has four quarters, yet refuse to attribute to him the belief that he has a dollar. It seems to be a feature of the concept of belief that we must allow at least a little bit of logical closure. The problem is that if we allow some logical closure then we obtain total logical closure, since even most complicated logical inferences can be reduced to a

series of applications of a simple rule such as Modus Ponens (as in some axiomatizations of quantificational logic). Hence, if we say that beliefs are closed under one application of Modus Ponens, we then must say that they are completely closed under logical implication.

An easy way out of this problem is to deny any closure condition on beliefs. But, this move ignores a major aspect of our concept of belief: some minimal rationality is presupposed for the attribution of a belief to an organism. Even those who might object to such a generalization for all believers are bound to agree that it conventionally holds at least for human (normal) agents. Another way out of the problem is to bite the bullet and admit total closure of beliefs. Such a move, as we shall see in Chapter 3, is unworkable. The practice of belief attribution seems to presuppose partial deductive closure, but not a total one. How can we have the one without the other?

I suggest that this problem can be solved by treating "belief" as a vague relation: whether this relation holds between an agent and an informational content may be a matter of degree. This is analogous to the treatment of "red" as a predicate which may apply to an object only to a certain degree. The notion of belief seems to be tolerant to marginal changes in its objects; though a large change in the same aspect may make a difference in the correctness of the belief attribution. For instance, given the belief in

the axioms governing first order logic, the ascription of the belief that $(\forall x)(Fx \vee -Fx)$ is intuitively more justified, in a normal context, than the attribution of the belief that $(\exists x)(\forall y)(Fx \rightarrow Gy)$ implies $[(\forall x)(Fx) \rightarrow (\forall x)(Gx)]$.

The "partial" deductive closure is achieved by assigning progressively lower degrees of believing to progressively remote or complicated deductive consequences of recognized beliefs, where the rate of "deterioration" depends on some contextual factors, but total deductive closure is avoided.[10]

The semantic analysis of belief sentences along these lines may thus be able to account for the empirical observation that agents believe some of the consequences of their beliefs, and fail to believe others. Another difficulty, that the above strategy of treating 'believes' as a vague predicate may solve, concerns implicit beliefs such as that Frege's earlobe is smaller than the Big Ben. We may be able to distinguish between various kinds of beliefs by assigning them different degrees. For example, the members of the following groups may be so distinguished: informational contents tacitly incorporated in the agents

---

10. A similar technique has been used to block the paradoxical conclusion that every man is bald, which follows by repeated applications of simple logical rules to the reasonable assumptions that a man with no hair is bald, and that the addition of one hair to a bald head does not change its baldness.

cognitive system, explicit but "stored away" beliefs, beliefs currently entertained, etc.

In summary, I wish to argue that the ideas put forward above, namely that (i) the objects of belief are distinct from meanings, and (ii) that 'believes' is a vague predicate, may enable us to explain the ordinary behavior of the concept of belief. For instance: the fact that the same belief sentence can be used to attribute different beliefs; the fact that belief attributions are generally sensitive to some underlying classification of believers that we presuppose in our communications; the fact that belief has a certain degree, short of complete, of deductive closure; and the observation that people sometimes hold contradictory beliefs, though the contradiction is not usually flatly visible as in 'p and not p'.

Chapter 2

The Logical Structure of Belief Sentences

The logical analysis of a sentence of the form:

(1) Daniel believes that Bob is smart

consists of two parts. (i) It specifies the logical structure of the sentence. This part involves, amongst other things, a specification of the logically significant components of the sentence and their logical categories. (ii) It specifies rules that show how the semantic values of the complex constituents are built out of the values of the simpler constituents by means of semantic rules. These rules are standardly articulated in the form of a recursive definition that traces the syntactic construction of the complex expression. The major issue in the first of these parts has been whether belief should be considered an operator or a predicate, and if the latter, how many place predicate it is. The major issue in the second part of the analysis has centered on the identity conditions of the "objects" of belief.

I shall present in this dissertation an argument that begins by undermining the importance of these controversies. I shall argue for the following two claims:

A. The logical form assigned to a belief sentence does not determine its semantics. The choice of a correct or desirable logical form bears mostly on the power of expressibility the resulting system may have. That is, the choice of logical form for belief sentences, within the context of first order language, affects the claims that can be expressed in the language.

B. The semantics one chooses to associate with a belief sentence does not determine the content of the belief ascribed. The choice of semantics bears primarily on what inferences are deemed semantically valid. The linguistic meaning assigned to the complement of a belief attribution fails, in my opinion, to pin down the belief attributed.

In this chapter I argue for the first claim. The chapter is divided into three sections. The first one displays five major alternative theories for the logical form of (1); four of these theories treat belief as a predicate and the fifth as an operator. The question whether the treatment of belief as a predicate is available to us is discussed in the second section, culminating with arguments for the superiority of the predicate approach over the operator approach. In the third section I argue for the inadequacy of two of the four "predicate" approaches.

SECTION I

A theory of logical form for a segment of a language assigns logical form to every interpretation of every sentence of that segment. It views each sentence as constructed from nonlogical[1] elements (e.g., names, predicates, function symbols) by means of logical operations such as negation, quantification and the like. Different analyses carve up the sentence into expressions of different grammatical categories. Let us begin with a survey of the grammatical categories before turning to the possible syntactic analyses of (1).

In the following short detour through logical grammar, I rely on the categorial framework originated by Lesniewski and others, and further developed by Nuel Belnap.[2] In this framework there are three fundamental grammatical categories in logical grammar:

A. **Sentence** (containing also open sentences).

---

1. I follow here the standard division of logical and non-logical elements.

2. Belnap 1975.

B. **Term** (containing proper nouns, noun phrases, abstract nouns, abstract noun phrases, that clauses, pronouns etc.). A term purports to denote a single object, and I take it to exclude common nouns, mass nouns, indefinite descriptions and the like.

C. **Functor** -- a grammatical function, whose argument is a list of items from the grammatical categories, and whose value is an item of one of these categories.

The first two of these categories, that of sentence and term, are relatively clear. The category of functor deserves some explanation. Functor is a broad category, whose elements correspond to connectives, predicates, function symbols etc. A functor takes items from the category of sentence (denoted by 0), or items from the category of term (denoted by 1), or other functors (whose category is determined by their input and output) and yields an item which is either a sentence or a term or a functor. A functor whose input belong to categories $\sigma_1,\ldots,\sigma_n$ and whose output belongs to category τ is an item of the category $(\tau/\sigma_1,\ldots,\sigma_n)$. For instance, 'and' belongs to the category (0/0,0) -- it is a functor that takes as input two sentences and yields a sentence; 'runs' belongs to category (0/1) -- it combines with a name (e.g., 'Bill') to give a sentence ('Bill runs'); 'the father of', is of category (1/1), since it takes a term to produce a term. Note that a functor may take as arguments other functors, e.g.,

'slowly', which takes a functor (e.g., 'runs') to form a functor ('runs slowly'). 'Slowly', thus, belongs to the category ((0/1)/(0/1)).[3] A functor that does not take other functors as values (i.e., whose arguments belong to the categories of sentence and term only) Belnap calls a **first-level functor.** Note also that a functor may take as arguments items from several different categories. Belnap calls a functor all of whose arguments belong to the same category a **pure functor.**

Functors can be classified in this way according to their inputs and outputs. There are infinitely many different kinds of pure first-level functors, four of these are of special interest. These four are described in the table below.

Name	Inputs	Output	Examples
operator	terms	term	--- + --- ; ---'s father
predicate	terms	sentence	--- < --- ; --- is nice
connective	sentences	sentence	--- and --- ; John surmises that ---
subnector	sentences	term	'---' ; that ---

The above framework can be used to clarify differences between five major treatments of belief attribution sentences. Note that the disagreement amongst the following

3. This kind of categorical grammar is further discussed in Chapter 5.

alternatives is reflected in the category that they assign to the functor 'believes' or 'believes that'.

1. The belief attribution 'Daniel believes that Bob is smart' contains two syntactic components: 'Daniel', which is of category 1, and 'believes that Bob is smart', which is viewed as an unanalyzed pure first-level predicate functor. It takes a name to form a sentence, and thus belongs to the category (0/1).[4] This analysis attributes to belief sentences a structure similar to the one standardly given to 'John is tall'.

2. 'Daniel believes that Bob is smart' is analyzed as consisting of a three place functor '--- believes that --- is ---', that combines with two names ('Daniel', 'Bob') and a one-place predicate ('is smart', of category (0/1)) to yield the sentence. As such, '--- believes that --- is ---' in the above sentence is classified as of the category (0/1,1,(0/1)). On this analysis,[5] the logical category of belief varies depending on the complexity of the complement sentence. (For instance, in the sentence 'Daniel believes that Tom loves Sarah' the category of belief is (0/1,1,(0/1,1),1), since the input consist of a two place predicate 'loves' and three names.) In this analysis, neither the complement sentence 'Bob is smart' nor the term

---

4. This theory, defended by Quine, is discussed in section III of this chapter.

5. Endorsed by Russell, and further explained in section III.

'that Bob is smart' are considered genuine logical constituents of the belief attribution. Belief here is neither a first-level nor a pure functor.

3. The role of "belief" in 'Daniel believes that Bob is smart' is that of a two place predicate functor '--- believes ---' which combines with two names ('Daniel', 'that Bob is smart') to form a sentence. Thus, 'believes' is a pure first-level functor of category (0/1,1). The second name 'that Bob is smart' is further analyzed as containing a functor 'that', which combines with a name ('Bob') and a predicate ('is smart') to yield a term. This analysis,[6] like the previous one, denies the complement sentence 'Bob is smart' the status of a genuine constituent of the belief attribution. The difference between the two is in the treatment of the term 'that Bob is smart', which is viewed as a genuine logical constituent on this analysis, but not on the previous one. Note that belief is multiply ambiguous on alternative (2), thought not on alternative (3). On the latter alternative, however, 'that' is ambiguous: the input arguments of 'that' vary according to the number and categories of the recognized components of the complement. In both analyses (2) and (3) the logical form of the belief attribution is sensitive to the complexity of the complement sentence.

---

6. Put forward by Cresswell, and discussed in chapter 5.

4. 'Daniel believes that Bob is smart' contains a two place pure first-level predicate functor '--- believes ---', which combines with two names 'Daniel', 'that Bob is fat' to form a sentence. The difference between this analysis and the previous one lies in the treatment of the complement 'Bob is smart': here it, as well as the term 'that Bob is smart', is considered a genuine component of the sentence.[7] The logical category of "belief" does not depend, according to this view, on the complexity of the complement sentence.

5. 'Daniel believes that Bob is smart' contains a functor '--- believes that ---' of category (0/1,0). I.e., its input are a name 'Daniel' and a sentence 'Bob is smart', which is analyzed in the usual way.[8]

The first four analyses of the belief-attribution sentence take belief to be a **predicate** functor, but disagree about the arguments the predicate takes. The fifth treatment views belief as a sentence **operator.**[9] Let us concentrate for now on the question whether 'believes' should be analyzed as a predicate or as an operator. Which alternative is the adequate one? From the grammatical point of view, it looks as if English is indifferent to whether

---

7. Theories along this syntactic line, e.g., Stalnaker's theory and Sentential theories, are discussed later in Chapters 3, 4 and 6.

8. The 'operator' approach, favored by Hintikka and Montague, is dealt with later on in this chapter.

9. Henceforth I will neglect the contribution of the subject term 'Daniel', since all approaches agree on it.

the functor in the belief attribution is a predicate or an operator. However, there is an argument, due to Montague, that purports to show that the only viable theory of the logical grammar of the belief attributions is the one given by the operator approach. The next section is devoted to the analysis of Montague's result and its lesson concerning the syntactic analysis of belief.

SECTION II

The **Montague argument** was given in the context of a parallel dispute that arises in the literature concerning modal notions. It has been disputed whether 'that p is necessary' should be viewed as consisting of an operator ('that ... is necessary') and a sentence (p), or as consisting of a predicate ('is necessary') and a term ('that p'). Quine suggested that there is no need to choose between these two options. In "Three grades of modal involvement" (1953), he states that the use of necessity as a sentence operator is easily converted into its use as a predicate.[10]

Montague came up with a troublesome response to Quine's suggestion. He showed that if necessity is treated as a

---

10. See also "Intensions revisited" (1977), where Quine asserts, from the position of a jewish chef preparing ham for a gentile clientele, that necessity as a predicate and necessity as an operator are interdefinable.

predicate (of sentences) then standard laws of modal logic imply a contradiction. In light of Montague's result it would seem that we do not have the liberty to choose between the predicate and the operator treatments. The "predicate" analysis is excluded on logical grounds.

Montague's argument relies on a formalization of the "Paradox of the Knower".[11] A simple version of this paradox is as follows: a judge decrees on Sunday that a prisoner shall be hanged on noon of either the following Monday, or Tuesday or Wednesday; that he shall not be hanged more than once; and that he shall not know until the morning of the hanging the day on which it will occur. The prisoner reasons that he can not be hanged on Wednesday, since it is the last possible day for the hanging, and he would know about the hanging on Tuesday afternoon, contrary to the stipulation that he would not know about it before the morning on which it occurs. Since Wednesday has been eliminated, Tuesday is the last possible day on which he could be hanged. But the same reasoning that resulted in eliminating Wednesday as a possible hanging day applies to Tuesday. Thus, Tuesday is eliminated as well as a day of possible hanging. By the same reasoning, Monday is eliminated as well. The prisoner, who logically concludes that the decree could not be fulfilled, is unpleasantly surprised to find himself at the end of a rope at noon of,

---

11. Also known as "The Hangman Paradox", and in another version as "The Unexpected Examination Paradox".

say, Tuesday. It appears that by simple logical moves and some basic intuitive assumptions it follows both that the decree cannot be fulfilled and that it can.

Montague does not attempt to solve this paradox. In a paper written with Kaplan,[12] he attempts to give the paradox a formalization that captures the basic epistemic assumptions involved. Montague and Kaplan showed that when the judge's decree is construed in a self-referential manner, it can be made to yield a contradiction in the presence of simple epistemic principles. I will not reiterate their formulation, which is quite complex, but I will illustrate the principle involved with a simple example. Consider the sentence

(2) The negation of this sentence is known to you.[13]

Suppose that (2) is true. Since only truths can be known, and by the assumption that you know the negation of (2), it follows that the negation of (2) must be true. This yields a contradiction. So the negation of (2) is true. But, as you and I have established that the negation of (2) is true, the negation of (2) is known to you. Since this is exactly what (2) says, (2) ends up being true after all, and this is a contradiction.

---

12. 1960.

13. The referent of the pronoun 'you' is you, the reader. Please play along.

This argument can be formalized within elementary syntax (as formalized in, say, Robinson's arithmetic, henceforth RA), extended by the predicate K$\underline{x}$, which can be read as: the truth of the sentence with Godel number $\underline{x}$ is known. The principles needed are:

(i) $K\ulcorner A\urcorner \rightarrow A$

(ii) If $\vdash A$ then $\vdash K\ulcorner A\urcorner$

(iii) $\vdash D \equiv K\ulcorner \neg D\urcorner$

The first assumption expresses the characteristic veridical property of knowledge: if the truth of A is known, then A is true. The second assumption states that every provable sentence is known.[14] The third assumption asserts the existence of a self referential sentence, that behaves in the manner of (2): D says that the negation of D is known.

The informal argument above can be formalized as follows:

1.	$\vdash D \equiv K\ulcorner \neg D\urcorner$	(from iii)
2.	$\vdash (K\ulcorner \neg D\urcorner \rightarrow \neg D)$	(from i)
3.	$\vdash (D \rightarrow \neg D)$	(from 1,2, by PC)
4.	$\vdash \neg D$	(from 3 by PC)
5.	$\vdash K\ulcorner \neg D\urcorner$	(from 4,ii)
6.	$\vdash D$	(from 1,5, by PC)

14. This assumption is very strong as it stands, and will be discussed below.

If 'K' is read as **"necessity"**, then the first two assumptions are axioms of weak systems of modal logic. The argument, therefore, seems to show that if "necessity" is treated as a predicate, basic laws of modal logic (e.g., that a necessary sentence is true) have to be given up. However, it can be shown that if necessity is added to RA as an operator, the inconsistency disappears. Thus the argument favors the operator treatment of modality over the predicate treatment.

If 'K' is read as **"knows"**, assumption (ii) is too strong: obviously not every sentence provable in a given system is necessarily known. The Montague argument actually contains a weaker version of (ii), according to which if a sentence is provable within first order logic alone then it is known. One could construct even weaker versions of this principle, by limiting the number or the complexity of the logical rules the application of which results in a sentence that is known. Note that the weaker the provability relation is, the stronger is the argument against the predicate treatment of knowledge. The principles necessary for obtaining a contradiction are very minimal, so all objections which question assumption (ii) can be met. The contradiction is obtained even when all that is required of the agent is that he employs three or four PC rules he already knows and goes through a short simple proof on just one occasion.

Obviously, the problematic epistemic assumptions do not apply to the notion of belief, but unfortunately the issue cannot be dismissed so easily, since a problem arises also when 'K' is read as **"believes"**. Sentence (iii) says, on this reading: D iff the negation of D is believed. By assuming a minimal level of awareness and rationality, a contradiction can be derived. If you believe D, you recognize that belief and, at least in this one case refrain from believing a simple sentence and its contradiction simultaneously -- it follows that you do not believe the negation of D. This implies, by (iii), that D is false. Once you have reached this conclusion, you believe the negation of D, but that implies that D is true. Thus, you are led to believe both D and the negation of D. The epistemic (or doxastic) assumptions used to generate the "believer" contradiction can not seriously be objected to as a strategy for avoiding the difficulty, since these assumptions need not be generalized -- we only need to imagine that the reader lives up to them in a particular case. The problem is that by extension on Montague's result we obtain that it is not even possible for an agent to behave according to the above assumptions.

The literature on the subject presents a large variety of strategies for avoiding the paradox, while maintaining that 'knows' is a predicate: some impose limits on constructing self-reference, some redefine 'knows' as an

indexical, contextual or metalinguistic predicate. Some reject one or more of the epistemic assumptions as improper extension of principles that hold for alethic contexts to epistemic context. Others deny that every sentence known (especially self-referential ones) expresses a proposition. Most theories about belief sentences ignore the Knower paradox all together, but perhaps the question whether it bears significantly on the choice of the appropriate account of psychological attitudes should be answered before such a dismissal is warranted.[15]

---

15. Most of the attempted solutions for the Knower paradox in the literature can be classified under one of the following three strategies:
I. Question the acceptability of one of the premisses in the argument. Binkley (1968) rejects assumption (ii) as pertaining to ideal knower only; Cresswell (1985) extends the same rejection to the first two assumptions by observing that Propositional Attitudes are not alethic -- they do not satisfy modal conditions. Thomason (1980) rejects the third assumption as containing a sentence that does not express a proposition (since it is self-referential); Prior (1961) rejects the third assumption as one which does not have a truth value.
II. Put restrictions on correct inference. Quine (1952) views the argument as containing a fallacy in reasoning, not a real antinomy: the conclusion the agent is prepared to accept as a certainty, namely that the decree cannot be fulfilled, should have been considered as a possibility from the beginning (in which case no contradiction arises). Halpern and Moses (1986) claim that knowing the conclusion must be compatible with the truth of the premises. Since the premisses of the argument contain inconsistent information, the conclusion is not knowable.
III. Redefine the predicate 'knows'. Skyrms (1978) suggests that we construct a sequence of metalanguages, each containing a distinct predicate of knowledge. Anderson (1983) extends the hierarchy to inference predicates, and Burge's solution (1984) involves indexicality of the Propositional Attitude verb.

The inconsistency demonstrated rests on the ability to form self reference in a language that also contains 'is true', 'necessary', 'knows', 'believes', 'is provable' and the like as predicates of sentences. The conclusion that Montague draws is that the treatment of these notions as sentence operators is the only syntactic option available to us.

I will not attempt to solve the Knower paradox. I will argue rather that it cannot be used to justify the operator treatment of "propositional attitudes" as the only viable one. One rather inconclusive piece of evidence for my claim follows from an analogy between the Knower and the Liar: The liar paradox shows that when truth is treated as a predicate of sentences, the natural principles concerning truth (namely, the Tarski biconditionals) imply a contradiction. If 'truth' is treated as an operator (within a first-order language), no inconsistency can be deduced. But this result is not taken to show that the operator treatment is the only proper syntactic treatment of truth.

My recommendation is to adopt the same attitude towards the lesson from the Knower paradox: The paradox does not eliminate the option of taking 'knows' as a predicate on logical grounds. My recommendation rests on the following observation. Notice that the difference between adding "knows" as a predicate versus adding it as an operator to a first-order language is that the former gives us greater

expressive power. For instance, on the predicate approach we can infer from (1) that there is something Daniel believes, represented as:

(3) (∃x)(Daniel believes x),

and we can formulate a sentence like 'Everything Mary believes John believes', represented as:

(4) (∀x)(Mary believes x → John believes x)

These simple sentences cannot be constructed on the operator approach, since within first-order logic variables are allowed only in name position, not in sentence position. In the sentence

(5) (∃p)(Daniel believes that p)

"believes" as an operator applies to sentences, not to names; hence 'p' occupies here a sentence position, a formalization which is outside the scope of first-order logic.

It is precisely this expressive impotence of the operator that saves this syntactic option from inconsistencies. The situation is particularly clear for the concept of truth: addition of a truth operator to the language does not contribute to the language's expressive power -- the sentence 'It is true that p' says no more then does p. As a result, the addition of the truth operator is not problematic, but that of the truth predicate results in

paradoxes. Observe that if we enrich the language's expressive resources by allowing quantification over sentence positions then self-referential sentences like:

(6) (∀p)(John knows that p → -p)

can be constructed, resulting in paradoxes similar to that of the Liar or the Knower.

The price the language which accommodates the operator 'knows' pays for preventing the formulation of self reference seems to me to disqualify it from being a proper tool for the analysis of belief (and other attitude) sentences. We cannot express in that language sentences like: there is something I believe; John and Mary believe the same thing; some of my beliefs are false; not everything people assert they believe; some of my beliefs influence my behavior, etc. Without resources to formulate the above sorts of sentences, we cannot even express adequate theories that explains how people form their beliefs, how their beliefs relate to their assertions, the logic of belief - desire explanation etc. On the other hand, any language that is sufficiently rich in expressive resources to be used in the analysis of belief sentences would generate paradoxes like the Knower. Thus, Montague did not prove that the only syntactic category that is available for 'believes' is that of a sentence operator.

Furthermore -- the logical form assigned to a belief attribution does not settle the semantic difficulties of opacity, compositionality etc. Following the operator strategy, for instance, still leaves the question of the semantic value of the complement sentence open. Let us survey briefly the most prominent account of belief along the operator lines, due to Hintikka.[16] The sentence

(7) X believes that p

contains, on this approach, an operator (X believes that), and a sentence (p). As such, p has a genuine semantic occurrence in the sentence: it denotes the object of the attributed belief, which Hintikka takes to be a proposition defined as a set of possible worlds. Attribution of a belief to a person involves a division of all possible worlds into those compatible with the belief in question and those incompatible with it. This division is characterized semantically by a function which associates, for each individual and each possible world, a set of possible worlds that are doxastic alternatives to the given one.

The coarse individuation of the objects of belief does not pose a problem for Hintikka, since he views the logic of belief sentences as part of a theory about ideal believers. The presuppositions that make the agent an ideal believer are that he is deductively omniscient; he is fully aware of

---

16. 1962.

his beliefs (if x believes that p then x believes that he believes that p); he has a complete knowledge of references of names, pronouns, indexicals and of all synonymy relations, etc. The ideal believer also operates reliably upon his knowledge: he has no performance limitations, and has unlimited storage capacity.

The way Hintikka's theory solves the problem of compositionality obviously does not rest on adopting the operator treatment of belief, but on its appeal to ideal believers. The intensional opacity of belief sentences is simply not recognized by the theory: a belief that p is identical to the belief that q when p and q are logically equivalent. When the semantic values of the complement sentences are taken as sets of possible worlds, as in Hintikka's system, the theory collapses into one of the modal systems (e.g., S4), and gains its elegance and familiarity. The problem is that what Hintikka offers is not a theory of belief sentences, but at best a theory of a notion that is misleadingly spelt like the word "believes" in English. One important consequence of Hintikka's theory is that the set of sentences occurrently believed by an agent is closed under logical consequence and contains not only infinitely many beliefs, but also beliefs that are too complex to be entertained in a life time. This consequence disqualifies the theory from being part of a semantic explanation of ordinary language.

In short, if belief is treated as an operator, and we wish to account semantically for non-ideal beliefs, the problem of individuating the objects of belief remains. Hence, as far as the semantic difficulties go, there is no advantage to the operator treatment. There are, however, several advantages to the predicate treatment: we can work within a classical framework, and the customary laws of predicate logic with identity may be employed. We can infer (3) from (1), since the variables can occupy term positions. On the operator option, in contrast, this inference is not possible within a first-order language, since we can not quantify on a sentence position. The inference can be preserved by enriching our language with an apparatus for quantifying on sentence position, but this is known to have formal difficulties on top of being very complicated.

SECTION III

Taking belief to be a predicate allows for at least four major incompatible syntactic strategies; these differ on the number of syntactic elements that the predicate applies to. In this section we will examine a theory that treats belief as a multi-place predicate (option 2 above), and another that takes belief as a one-place predicate (option 1

above).[17] My aim is to show that these syntactic treatments of belief sentences do not solve the essential semantic difficulties that belief contexts generate, but rather create additional difficulties. The following chapters concentrate on various accounts based on the alternative strategy according to which belief is a two-place predicate.

The **multi-grade predicate approach** seems to offer the advantage of avoiding the problem of the individuation of the objects of belief, since it denies that there is a single object to a belief. An example of such a theory is found in Russell's works.[18] The problem Russell tackles is not that of opacity, but that of accounting for false beliefs, which seem to correspond to "negative facts". This notion is completely incomprehensible to Russell, who denies its existence vehemently. Russell concludes that 'that p' is not a denoting term, for if p is false then it cannot denote anything. 'Believes', according to Russell, relates the agent to a various number of terms with which the belief is concerned. For instance:

In 'x believes that A is F', 'believes' is a three place relation, applying to x, A, being F.

In 'x believes that A loves B', 'believes' is a four place relation, applying to x, A, B, loving.

---

17. Cresswell's theory, which advocates option 3, is discussed in chapter 5. Theories advocating option 4 are discussed in chapters 3, 4 and 6.

18. Russell 1910; 1919.

In '$\underline{x}$ believes that A loves B in May', 'believes' is a five place relation. In '$\underline{x}$ believes that A is jealous of B's love for C', 'believes' is a six place relation, and so on.

On this account, belief is rendered ambiguous: belief sentences have infinitely many logical forms, depending on the infinitely many forms that their complement sentences can have. 'Believes' is not "logically one" in all different cases, but must be distinguished according to the "nature" of the proposition believed. Thus, the logical form of (1) denies both 'that $\underline{p}$' and '$\underline{p}$' the status of genuine components.

Russell coined the term "propositional verbs" (he actually opposes the replacement of 'verbs' by 'attitudes', because of the psychological flavor the latter carries) for such words as 'knows', 'believes', 'desires' and the like, since these verbs have the form of relating an object to a proposition. But, according to Russell, that is not what they really do. The awkwardness of saying that we believe propositions is that "obviously propositions are nothing",[19] they do not have independent existence and are not part of the inventory of the world. In a complete description of the world, in Russell's opinion, false propositions will not be mentioned, but false beliefs will; so propositions are not the objects of belief. Russell concludes that "...the

---

19. In "The philosophy of logical atomism", 1918, p.222.

belief does not contain a proposition as a constituent, but only contains the constituents of the proposition as constituents".[20]

What are these constituents? In 1910 Russell identified the constituents of a proposition with the **constituents of the fact** which is the objective of that proposition (i.e., the fact that makes the proposition true or false by some intuitive correspondence relation).[21] Thus, the proposition that Desdemona loves Cassio has three constituents: Desdemona, Cassio, the loving relation by which the other two are "knit together". Two propositions have the same logical form iff they differ only in regard to their constituents, e.g., 'A loves B' has the same form as 'C hates D', but differs from 'A loves B on Sundays'. The importance of the role of the position of the constituents, which distinguished 'Socrates loves Plato' from 'Plato loves Socrates' seems to bear not on the the logical form of the proposition, but on the "form" of the fact (which takes the positions of its constituents into account) to which the proposition corresponds.

A few years later (1919) Russell changed his mind about the constituents of belief: no longer are they constituents of the objective facts corresponding to the proposition, but

---

20. Ibid p.224.

21. "On the nature of truth and falsehood", in Philosophical Essays.

images (which have meanings) and sometimes **sensations** (which don't). The content of a belief is defined as a proposition, which, in turn, is defined as consisting of interrelating images. These images refers to facts by virtue of their contents. The advantage Russell believes to have gained by distinguishing the logical classification of beliefs (via propositions) from the psychological classification (by sensations and images) is the possibility to admit propositions as actual complex occurrences, thereby do away with the difficulty of false beliefs.

The substitution of intensionally equivalent complement sentences of the belief attribution is blocked by Russell's account. The constituents of 'Tom loves Sarah' are different from those of 'Sarah is loved by Tom', since the relations of "loves" and "is loved by" are distinct. Hence believing the first cannot be identified, on Russell's theory, with believing the second. Unfortunately, on the first theory the substitution of co-referential or co-intensional atomic constituents of a belief is not blocked, because on this theory the referents of names themselves are semantically significant for the semantic value of the whole sentence. Hence on this theory the belief that Tully is bald is identified with the belief that Cicero is bald, since their constituents are identical. On the revised theory the constituents of belief are "images", so this specific problem does not arise. But, the employment of the notion of "image" generates other problems due to the

subjective and elusive nature of images. The truth or falsity of a belief, as a relation between a believer's images and the world, becomes totally obscure; explanation of behavior and communication by means of belief attribution are questionable, and worse there is this problem: how can Jo's belief, for instance, that round squares do not exist, be understood as a relation Jo has to an image of a round square, when it is very likely that Jo does not (and cannot) have any such image?

Another difficulty with Russell's account concerns the large gap between the logical form assigned to belief sentences by the theory and their intuitive grammatical form. This divergence seemed to Russell justified on the basis of his conception of propositions: 'that p' cannot be considered a denoting term, since p may be a false proposition, in which case 'that p' denotes, on Russell's theory, a "false fact". Obviously, propositions need not be (and usually are not) thought of as so intimately tied to concrete facts.

A treatment of belief as a **one-place predicate** is proposed by Quine in Word and object. According to him, everything that follows the first singular term X in the construction 'X believes that p' should be treated as an unstructured predicate. Each occurrence of the belief verb together with the sentence that follows it is rendered an isolated independent predicate. This proposal does not recognize any

common element between two belief predicates 'believes that p' and 'believes that q', despite the employment of the same verb 'believes' and regardless of any semantic relations that p and q may have. Quine's theory is sometimes called the "fusion theory", since it "fuses" the belief verb with its object. Some of Quine's motivations for the above proposal stem from his view about intensional objects -- like propositions, properties etc., -- that should be dispensed with, since the question of their individuation is a "mistaken ideal".[22]

The problematic logical inferences, like the inference of

(8) a believes that q

from

(9) a believes that p

and

(10) p is logically equivalent to q

are blocked for all distinct p and q, because an inference from (9) to (8), according to Quine's view, is similar in form to the inference of 'a is F' from 'a is G'. Unfortunately, all other logical inferences are blocked too; e.g., the inference that a believes that Bill loves Mary,

---

22. 1960, p.206.

given that a believes that Mary is loved by Bill. Quine sees a great advantage in this radical result. On his view, there seems to be no intuitively acceptable inferences involving the structure of the complement sentence, so he sees no point in assigning any logical structure to belief sentences. According to Quine, even inferences that do not involve the logical structure of the complement sentence are not allowed. For instance, given that a believes that p and that b believes that p we cannot deduce, on Quine's theory, the conclusion that a and b believe the same thing. From (9) we cannot even infer that there is something that a believes, but Quine is not disturbed by that: he regards such quantifications as trivial, dismissable, expendable.

Even if we grant Quine his disregard for these intuitively acceptable inferences, his theory is paying too high a price in the attempt to avoid the undesirable ones. According to the fusion theory, too many phenomena that call for some connection end up being just accidentally alike. Thus a's belief that p has no relation to a's desire that p or to fearing that p. The occurrence of an identical complement sentence in those constructions is treated as accidental, since the theory does not take that complement to be a semantically significant component in the analysis of these sentences. So, even though the proposal is extendable to all Propositional Attitude sentences, the benefit gained is meager if the theory cannot account for the interactions between the mental states described by

these sentences. This theory cannot provide any explanation of how people modify their beliefs on the basis of other accepted beliefs, of how their fears and hopes depend on the way they perceive the world, or how their actions stem from or reflect the relevant mental attitudes. There is also no relations between the sentence 'pigs fly' and the complement of the sentence '_a_ believes that pigs fly'. Furthermore -- it is a sheer accident that what _a_ believes, when he believes that _p_, is true iff that _p_ is true.

Another objection that is often raised against Quine's theory is that if every 'believes-that-_p_' forms a new predicate, then primitive linguistic resources end up being infinite: there are infinitely many semantically distinct sentences of the form 'believes-that-_p_'. Davidson observed that the problem is not in having infinitely many primitives, but rather that there is no rule for determining the meanings of these primitives in a finite way, and therefore cannot account for learnability of language.

Some of the criticisms made above of proposals for the logical analysis of belief sentences rest implicitly on the following constraints on a theory of logical form:[23]

1. There should be syntactic evidence for the logical form assigned to a sentence.

---

23. Some of these are suggested by Harman, 1975.

2. The logical form should minimize novel rules of logic. I.e., the logical rules applicable to the categories of the constituents should be kept as close as possible to the rules of ordinary Predicate logic with identity.

3. The valid logical inferences should mirror the intuitive inferences made with these sentences, and most of the intuitively valid inferences should be demonstrable in the theory.

The operator strategy fails to comply with the second principle. Russell's and Quine's approaches do not conform to the first one. All theories criticized above violate the third, perhaps the most important, constraint.

The two syntactic alternative remaining from the list given at the beginning of this chapter, namely (3) and (4), agree on the treatment of belief as a two-place predicate and on recognizing 'that p' as a genuine component of the belief attribution. Alternative (3), which does not take p as a genuine component of the sentence, is discussed in Chapter 5. The syntactic alternative I favor is (4): it views 'that p' as obtained by combining p with a subnector 'that'. The basic problem for a theory that follows this syntactic option is to figure out the denotation of 'that p', i.e., it has to give an account of the object of belief. Some prominent theories along this line are discussed in

Chapters 3, 4 and 6, while my proposal is outlined in the last chapter.

Chapter 3

The Possible-Worlds Theory of Propositions

A natural grammatical analysis of sentences like 'Alfred believes that pigs can fly' treats "believes" as a two-place relation that holds between the believer Alfred and the object denoted by 'that pigs can fly'. Accordingly, this analysis views the sentence as consisting of three parts:

1. **'believes'**-- a two place predicate, that takes two names and yields a sentence

2. the **subject term** 'Alfred' which denotes the agent (the one doing the believing)

3. the **nominal** 'that pigs can fly' which denotes the proposition expressed by the sentential complement 'pigs can fly' (the object of believing).

The notion that most needs an explanation in this analysis is that of "proposition". I will discuss in this chapter the account of propositions that is provided by the possible-worlds semantics.

Possible-world semantics treats propositions as functions from possible worlds into truth-values. Thus, relative to a domain of possible worlds, a proposition is fully determined by that subset of the domain in which the proposition takes the value "true". So an equivalent

definition of proposition, in possible-worlds theory, is to say that it is a set of possible worlds, i.e., the worlds in which it is true.

This account of proposition has proven fruitful in several areas of semantics, but it has one apparent drawback: propositions which are true in exactly the same worlds are represented by the same function. Consequently, in this theory necessarily equivalent propositions turn out to be identical. Belief contexts, as mentioned, seem to require a more refined individuation of the objects of belief in order to account for the actual data about this mental state. John Bigelow brings the following pair of sentences to illustrate the point:

1. Robin will win.
2. Everyone who does not compete, or loses, will have done something Robin will not have done.

These two sentences express necessarily equivalent propositions. By the Possible-world analysis they mean the same. So the analysis implies that the following two sentences are logically equivalent:

3. Marian believes that Robin will win.
4. Marian believes that everyone who does not compete...

But it is not difficult to imagine the possibility that Marian, not realizing the identity of the propositions involved, believes the first sentence but not the second.

This failure of substitutivity of necessarily equivalent propositions for one another in belief context seems to disqualify the possible world account from properly representing the objects of belief.

Stalnaker, in his book Inquiry, offers an interesting defense of the Possible-worlds account. This defense has two major components. The first component, to be discussed in section I, brings several independent arguments in favor of treating propositions as sets of possible worlds. Stalnaker argues that the proper account of belief and desire, something he calls "the causal-pragmatic theory", implies that the objects of belief and desire should be individuated in exactly the way that the Possible-worlds theory individuates them. Stalnaker argues in favor of the "causal-pragmatic theory" that it solves the problem of intentionality (explained below) and that it gives a successful overall account of belief attributions. The second component of his defense is his response to the logical problems that propositions so individuated face in belief contexts. We will examine his methods for dealing with these difficulties in section II of this chapter and in the next chapter.

SECTION I

The pragmatic theory views rational creatures as essentially agents, performers of actions. Representational mental states, like belief and desire, are treated as correlative functional dispositions to behave in certain ways. One major key for understanding these mental states is by tracing the role they play in the characterization and explanation of the behavior they tend to produce. This intuitive picture is similar in its broad outlines to the functionalistic account endorsed by Putnam, Fodor and others, which individuates mental states of an organism by their functional role in the workings of that organism. The causal aspect of the causal-pragmatic theory provides an account of how the interactions between the world and the agent produce these mental states.

In the following pages I will discuss (1) the pragmatic aspect of the theory, (2) the need for the addition of the causal aspect, (3) how the two integrate into a complete picture which leads, according to Stalnaker, to the possible-worlds analysis of propositions.

The **pragmatic theory** views the rational agent as having a stock of desires and beliefs, which combine together to produce a disposition to behave in a way that is likely to fulfill the desires, given the beliefs. A picture one may draw is that of an agent being confronted with a range of

alternative actions he may take towards the realization of his desired state of affairs. These alternative actions have, according to his beliefs, various possible outcomes, towards which he may have pro or con attitudes. The agent chooses that course of action that best fulfills his desires. For example: when desiring to get from a point A to a point B, I consider the alternative transportational devices available to me: a train, a bus, a taxi, my car. Each of those will, very likely, bring me to my destination. I have additional beliefs about the possible outcomes of employing any of these devices, and corresponding pro and con attitudes towards these consequences: I do not wish to stand in a crowded train, neither to take an annoyingly slow bus, nor to pay a lot for a taxi ride. My decision will be made according to my pro and con attitudes towards these possible outcomes. The content of my attitudes--the stock of beliefs and desires relevent to the situation--can be treated as distinctions made between the elements of the range of possible actions and possible outcomes. My choice of action is the decision to try to actualize a certain state of affairs, that is picked out of a collection of possible states of affairs.

The idea that beliefs and desires lead to actions is widely accepted. The pragmatic theory is striving to do much more than reiterate the trivial: it attempts to explains what belief and desire are solely in terms of the

actions they lead to. The explanation it gives is the following:

> to desire that p is to be disposed to act in ways that would tend to bring it about that p in a world in which one's beliefs, whatever they are, are true. To believe that p is to be disposed to act in ways that would tend to satisfy one's desires, whatever they are,in a world in which p, together with one's other beliefs, were true. (p. 15)[1]

The pragmatic theory identifies the beliefs and desires a person has with certain functional states that are responsible for the person's behavior. Both belief and desire are explained in terms of the properties of the dispositions they tend to generate.

There are several problems with this account, the most prominent of which is that of circularity. Reconsider the transportation example discussed above - from the fact that I took the train downtown, one could construct at least two incompatible hypotheses about my beliefs and desires, in order to explain my behavior. As part of one such hypothesis, I desire to go downtown and believe that my car is still in the garage. The other contains my belief that my car is in working condition, that I don't want to go downtown, or anywhere else for that matter, but rather have the perverse desire to stick to a sweaty crowd wherever I can. The problem is not that of lacking the decisive method

---

1. All quotations, unless stated otherwise, are from Stalnaker's book Inquiry. Parenthetical page references are to this book.

for choosing the correct hypothesis -- Stalnaker doesn't claim to provide a behavioristic reduction of mental states to behavior. The difficulty is rather with the limitation of the functional role of these attitudes. Even if the functional role of my beliefs is fixed, the content of my beliefs is not fixed until the content of the desire is. And similarly for desires. If there were a way of pinning down the content of my desires (for instance, the perverse one), one could then fix the content of the beliefs associated with it in explaining the behavior. If we could fix the content of my beliefs, we would be able to do so for the content of the desires that are part of the same hypothesis. But the pragmatic picture doesn't provide us with the means for doing either, but only with a correlation between the two mental states. This correlation is helpful only when either the content of desire or the content of belief is already given. Otherwise the content of my beliefs and desires is left indeterminate.

Stalnaker recognizes that the explanation of the content of belief and desire in terms of the behavior they tend to cause is not sufficient. The pragmatic theory captures only the forward looking aspect of belief and desires, since it explains them in terms of what they tend to cause. It should be supplemented, Stalnaker thinks, by a backward looking aspect which would tie the relevant events and states of the world with the beliefs they tend to cause. By adding this causal aspect, we are equipped with a new

relation between beliefs and what causes them, which provides a way of breaking into the vicious circle. We can fix the content of beliefs first, which allows the determination of the content of the desires, according to the correlation established by the pragmatic theory.

The content of a belief, as a backward looking propositional state, is analyzed in terms of what tends to cause the person to have this belief, under normal conditions. The causal relation between that which causes the belief and the belief is the propositional relation of indication. Indication is a type of correlation that holds between an object and its environment under the following conditions: the object has intrinsic states that tend, under normal or optimal conditions, to correlate with its environment in a systematic way, and the subject tends to be in the state it is in because the environment is the way it is. For instance, the length of a column of mercury in a thermometer tends to vary systematically with the temperature of its environment, thereby indicating this temperature, under normal conditions. The number of rings on a cross section of a tree trunk correlates, under some conditions, with the tree's age, indicating the age according to some formula.

To define the indication relation, one needs to specify a relevant set of alternative states of the object which do the representing (e.g., the tree's having x rings); a one-

to-one function from these states into the corresponding states of the world (e.g., it may take "the tree's having x rings" into "the tree's being x years old"); fidelity (normal, optimal) conditions that should hold for the relation to hold (like health of the tree, lack of distorting influences like certain climatic conditions, etc.). Once these three components are specified, the correlation relation is defined by Stalnaker as follows: let the variable a range over the relevant states of the object and let f be the 1-1 function from the objects' states into the world's states. The correlation holds iff, for any a, if fidelity conditions hold, then

1. the object is in state a iff the environment is in state f(a),

2. the environments' being in state f(a) causes the object to be in state a.

Indication is a relation between objects and propositions. It is schematically defined as follows: "the object indicates that p iff, for some a in the relevant set of alternative states of the object, first the object is in state a, and second, the proposition that the environment is in state f(a) entails that p" (p. 13).

When we apply this definition to belief, we have the following explanation of the relations our beliefs have with the world: "we believe that p just because we are in a

state that, under optimal conditions, we are in only if p, and under optimal conditions, we are in that state because p, or because of something that entails p" (p. 18). The threat of a vicious circle, posed by the interdependency of the explanations of belief and desire as dispositional states, is believed by Stalnaker to be removed by the addition of the indication relation to the theory: the content of belief is fixed by its pattern of causal connection with the world. Beliefs are distinguished from other representational states because of their role (connected with desires) in producing behavior. An example: how does my belief that there is an apple on the table represent some fact about the world, and causes me to act in a certain way, given certain desires? First: the fact that there is an apple on the table causes me, under normal conditions, to have this belief, and second: combined with my desire to eat something, the act of approaching the table, grabbing the apple and sinking my teeth into it explains my relevant beliefs, given my desire. We generally presuppose that people's beliefs correlate with and are caused by their environment. When a deviation occurs, we explain it as absentmindedness, optical illusion, self-deception, incapacity or misremembering, etc. If I were to sink my teeth into a plastic apple, my false belief that there is an edible apple on the table can be explained as visual deception, forgetfulness or the like.

The causal-pragmatic theory postulates two types of relations holding between the agent and the contents of his attitudes. We have seen that the correlation of his desires and beliefs as dispositions to behave in certain ways is explained as "tendency to bring about". A simple example for this relation is an object or a system, that when in equilibrium state, tends to remain there, while when in disequilibrium, it tends to change in ways that would bring it into equilibrium. The second type of relation is the correlation that holds between the agent's beliefs and the state of the world, which Stalnaker calls "the indication relation".

Stalnaker thinks that the causal-pragmatic theory can help solve the problem of intentionality: the problem of understanding the nature of intentional relations, of understanding how the intentional states represent the world. He sees in the causal-pragmatic theory an approach which helps us to understand intentional relations in terms of naturalistic relations.

Naturalistic relations, like the one which holds between heat and the behavior of metals, are relatively unproblematic, and can be fully understood in terms of causal interaction, spatiotemporal relations, physical and chemical laws etc. Intentional relations, on the other hand, are problematic. One can fear things which do not exist, like ghosts and monsters. One can wish to square the

circle, to live forever or to have a perfect child. One can desire food, without there being a particular food one desires, or wish to have a husband without having a specific man in mind. One problem these phenomena bring out is the puzzling character of the representational properties of our mental states: how do they represent, or are about, other things?

In order to solve the problem of intentionality, Stalnaker needs to provide a naturalistic account for the two relations he reduces belief to: the disposition to perform certain actions, and the relations a belief has with the world. We can understand how the number of rings the tree trunk has represents the tree's age, or how the mercury's level in a thermometer represents the temperature of our bodies, but how are we to understand the representation of the world by our beliefs and desires? By showing how intentional relations can be understood in terms of the forward looking relation "tendency to bring about" and the backward looking "indication", Stalnaker lays down a framework which has some promise of explaining the intentional relations in naturalistic terms: we have seen that "indication" and "tendency to bring about" can perhaps be explained physicalistically.

If the causal-pragmatic theory is correct, then the objects of belief ought to be individuated in a coarse way, as in the possible-worlds analysis. The reason is twofold:

(1) Indication, it will be recalled, was defined in terms of causal relations between belief states and states of the world. Since logically equivalent sentences describe the same state of affairs, it follows from the definition of indication that a belief state which indicates p is bound to indicate q as well, if the two are logically equivalent. (2) The pragmatic aspect of the theory leads to the identification of logically equivalent beliefs and desires, since logically equivalent states have the same functional role in causing behavior. Thus the causal pragmatic theory strongly motivates the possible worlds analysis of propositions.

Stalnaker presents two problems about propositional-attitudes, particularly about belief, that he believes the causal-pragmatic theory to provide adequate answers to: 1. How do these attitudes have content? The solution is that the causal aspect of the theory fixes the content of beliefs according to the indication relation they have with the world. 2. How can we distinguish between different propositional-attitudes? The answer is that the way to distinguish between them is by their functional role in producing behavior, according to the pragmatic aspect of the theory. This aspect postulates the correlation between desire and belief as dispositions to act.

Stalnaker explains the content of belief in terms of "indication" which in turn is explained in terms of causal

relations that hold under various "fidelity conditions". One problem with this is that the theory is presented in a very schematic way. It leaves the notion of "fidelity conditions" exceedingly vague, vague to the point that it could be anything. As a result it is difficult to evaluate whether the indication relation fixes the contents of beliefs. Nonetheless, some criticisms apply even to the schematic account.

The notion of the indication relation seems to be too restrictive. It does not allow for imagination, memory, dreams, inventions etc. as sources of identifiable beliefs. Furthermore, it is not clear what state of the world could cause me to have beliefs of ethical, theological or aesthetical nature. Even if we allow a sentence like 'It is wrong to abuse your dog' to be a correct description of some state of affairs, what kind of causal relation can obtain so as to produce the belief in this sentence? We have many beliefs whose content does not seem to have any causal relation with the world, and which indicate at most a temporary state of mind, mental or intellectual. In general, it would appear that "non-factual" beliefs may not have their content fixed by the indication relation.

Another difficulty with the idea is that the indication theory of belief allows one to prove too much. Stalnaker aims to prove that necessarily equivalent propositions are identical. A belief state indicates equally all states of

affairs that correspond to logically equivalent propositions. The problem is that we end up identifying propositions which correspond to physically equivalent states of the worlds. From the definition of indication it follows that if an instrument indicates a state of the world S_1, and S_1 is physically equivalent to state S_2, then the instrument indicates S_2 as well. Do beliefs follow this rule? -- If it is a physical law that a rainbow is visible iff it rains and shines simultaneously, then the content of my belief that a rainbow is visible now is identified with the belief about the simultaneous occurrence of rain and shine. Other examples may make this weak identity condition on propositions far more absurd. Let us assume that as a result of some biological law, I have a headache iff the amount of sugar in my blood and the density of calcium in certain cells drops below level x. It is easy to imagine my ignorance of this biological law, still, by the indication relation, my belief that I experience a headache ends up also being a belief that there are these chemical processes in my body.

It may appear that the difficulty can be overcome by setting the fidelity conditions in such a way that it interferes with identifying physically equivalent propositions. Some fidelity conditions may be of an obvious type: normal lighting conditions, no hidden mirrors, no distorting atmospheric conditions, the agent not being drugged etc. We may even allow restrictions in the form of

fidelity conditions about social and linguistic habits of the believer's community. But Stalnaker needs more than that in order to overcome the problem: he needs to exclude from the range of alternative possibilities the worlds in which the physically equivalent propositions are true, whenever the agent is ignorant of that equivalence. It looks like subjective psychological and intellectual conditions about my personal way of thinking, the scope of my knowledge, perception and other mental capacities should be incorporated, so that the content of my belief about my headache is not identified with a proposition about sugar and calcium level in my body. Such a move would bring Stalnaker straight from the frying pan into the fire. The reason behind introducing the notion of indication to begin with was Stalnaker's desire to reduce the intentional relation of belief to a naturalistic relation, which does not employ intentional terms in its definition.

These criticisms do not show the possible-worlds theory to be wrong. They are aimed only at Stalnaker's attempt to justify the theory through the causal-pragmatic account. We should examine how Stalnaker attempts to cope with the problems created by the identity conditions for propositions in his theory. But before doing this, I would like to mention briefly some of the other arguments he brings in support of the claim that the possible worlds analysis is motivated by a natural notion of **information**, and that it

helps to understand tacit beliefs, belief-attributions to animals, to children and to abnormal agents.

By means of sentences we communicate information and impart news. When do we consider something "news"? An intuitive answer is: when it tells us something about the world; when it enables us to rule out certain possibilities that we would not have ruled out without that information. This natural account leads one to view information in terms of alternative possibilities: as a device given to us for distinguishing some possibilities from the others. If we identify the proposition that a sentence expresses with the information it is used to communicate, then we end up with the possible-worlds theory.

One desirable consequence of this analysis of propositions is that we can fix the content of a sentence independently of language and linguistic behavior. This is desirable, since the notion of content is, intuitively, independent of the particular linguistic form used to communicate it.

Acquiring information, increasing knowledge, amounts to having an increased ability to distinguish between possibilities. This model for the content of knowledge seems suitable for some types of inquiry, namely - those concerning empirical matters. If propositions are contingent contents for all statements - how do we explain the content of mathematical or logical statements? We

consider these necessarily true or false, hence in the possible-worlds semantics they express the universal or the null propositions. What sets are ruled out when we acquire the content of a mathematical statement? None, it seems, since this content is true (or false) in every possibility.

Stalnaker aims at a unified account of all inquiry, deductive inquiry being just a special case. His answer (to be spelled out in the next section) is that mathematical truths/falsehoods statements do indeed express contingent propositions: propositions about the relations between the statements and the one necessarily true or the one necessarily false proposition.

Possible-worlds theory, according to Stalnaker, helps solve a tension in our attribution of beliefs to animals. On the one hand, we have the tendency to explain the behavior of an animal -- of, say, a dog barking at a tree, right after witnessing the squirrel climbing on it -- with the help of notions like desires and beliefs the animal has. If the dog barks at the wrong tree, we tend to say that he has a false belief about the location of the squirrel. On the other hand, we are reluctant to attribute to the dog the ability to entertain well-structured sentences of some full blooded language like English, or to fully comprehend even simple concepts. How can the dog's belief be the same as the human's belief that the squirrel is on the tree?

Stalnaker accommodates our tendency for a common-sense psychological explanation of some animals' behavior in terms of their beliefs and desires, by tracing the difficulty of attributing these beliefs to either one of the following: 1. assumptions about what belief is like (for instance: a disposition to assert or to assent to the proposition when asked about it) 2. assumptions about the complexity of the objects of beliefs (like: structured objects made up of concepts, objects with semantic structure which mirrors the grammatical structure of a sentence, etc). Unstructured propositions (sets of possible worlds) as the objects of belief have the virtue that they are attributable as contents of beliefs to agents who lack a language.

We use sentences of our natural language to ascribe beliefs to animals, but that does not mean that we attribute to them the ability to entertain all possible situations. All we presuppose is that these animals have some mechanism for representing and distinguishing between some alternative possibilities. This mechanism may be very limited and crude, recognizing only gross differences between alternatives, and may not contain any representation of possible situations which are very different from the actual one. Because propositions do not mirror the structure of the sentences which express them, we can use complex semantical expressions to ascribe attitudes to animals. We allow for their limited cognitive capacities by letting the propositions expressed distinguish between fewer

alternatives -- those that the animals represent, without thereby attributing our concepts to them.

Even though I probably have a deeper understanding of concepts like "a tree", "a squirrel" and some spatial relations, my belief that the squirrel is up on that tree is identical to my dog's belief, when we take the content of that belief (the proposition) to be just a way of dividing a particular given set of relevant alternative possible worlds into two parts. The set of relevant possibilities - the set that is relevant to my dog's belief - may not contain an hallucination of a squirrel, a false tree, etc. Since the identity conditions for propositions are relative to a given domain of arguments, the expressions which attribute these beliefs to me and to my dog may determine one proposition relative to one set of possible world (the one limited to what we believe the dog is capable of entertaining) and may determine different propositions relative to larger sets.

Stalnaker uses the same strategy in giving an account of beliefs of children, of abnormal people, of ignorant people or any attribution of belief to an agent who lacks a full understanding of all the concepts involved in the expression used to attribute the belief to him. One of the virtues of the possible-world analysis of propositions is that it allows for "quasi-knowledge", partial understanding, and growing comprehension. As Stalnaker explains, "The child who says his Daddy is a doctor...can divide a certain,

perhaps limited, range of alternative possibilities in the right way and locate the actual world on the right side of the line he draws. As his understanding of what it is for Daddy to be a doctor grows, his capacity will extend to a larger set of alternative possibilities or, rather, the extension of his capacity will be the growth of his understanding." (p.65)

Tacit beliefs are those ascribed even though the believer has never expressed or consciously thought about them. For example, we may ascribe to Bertrand Russell the belief that Big Ben was larger than Frege's left earlobe (Stich's example). But this belief is probably a tacit one. We usually tolerate such belief attributions, e.g., when they are needed to complete the set of premisses from which a person explicitly draws some valid conclusion. If we take belief to be a sentence-like representation, such a tolerance amounts to assuming that we can store away in our memory-bank representations of all the trivial and obvious facts we take for granted. We may want to allow agents to have an infinite number of independent beliefs, but be reluctant from equipping them with a belief state with an infinite number of components. A way out of this problem is suggested by the the possible-worlds account of propositions: a belief state does not have propositions as its components. Individual beliefs are properties of such a belief state: to believe that p is for the proposition that p to be true in all the possible worlds in the belief-state.

This possible world conception of a belief state makes having a certain belief identical to not recognizing a certain possibility (that that belief is false), which makes it easier to explain having an infinite number of beliefs, having tacit ones or previously unrecognized presuppositions.

All of the above arguments in this section are aimed at showing that the analysis of propositions, suggested by the possible-world theory, is not merely a convenient technical device in semantics. Stalnaker believes that the pragmatic picture of our mental attitudes, the conception of the notion of information as increasing our capacity for distinguishing between alternatives, and his accounts for other relevant problems - all of these provide a deep intuitive motivation for accepting the account of propositions as sets of possible worlds. It is a consequence of this account that the objects of belief are identical iff they are necessarily equivalent.

SECTION II

In this section I will explore in more detail the possible-world analysis of propositions, the specific problems it poses for belief contexts, and two of the three devices Stalnaker uses for coping with these problems. The third device will be discussed in a later chapter.

A proposition is a function from possible worlds into truth-values. Since identity conditions for functions are purely extentional (for any functions f,g, if they are defined for the same arguments and have the same value for each argument, then they are identical), a proposition is fully determined relative to a domain of possible worlds by the subset of the domain for which the proposition takes the value 'true'.

How much structure should we impose on propositions? Stalnaker's tendency is to bring it to a minimum: "the aim of the possible world definition of proposition is to assign to the content of representation just the structure that is motivated by the pragmatic account of the functional role of representation" (p.23). Propositions are like sentences, in that they can stand in entailment relations, be contradictories etc. There are, however, on Stalnaker's conception, at least three differences between them: propositions have no constituents; relations between propositions hold relatively to a domain of possible worlds; propositions are not components of a belief state (as would be sentences, if we were to regard a system of beliefs as being a list of sentence-like items). Propositions are rather properties of a belief state.

When propositions are treated this way, some of the problems that the possible-worlds semantics encounters in belief contexts are:

(i) The problem of **equivalence:** since necessarily equivalent propositions are identical, believing one, according to the possible world theory, entails believing the other. But agents do not always stand up to this test. The failure to do this is particularily visible when dealing with mathematical or logical propositions: since all the true ones are necessarily equivalent to each other (and similarly with the false ones), believing that 2+2=4 entails believing a most complicated, perhaps not yet discovered, mathematical truth.

(ii) The problem of **conjunction:** if person a believes that p (Bap), and that same person believes that q (Baq) then it follows by this semantics that a believes that p and q (Ba(p.q)).

(iii) The problem of **deduction:** by the possible worlds semantics, if $Bap_1, Bap_2 \ldots, Bap_n$, and $p_1, \ldots, p_n$ logically imply q, then Baq. But in reality, agents do not always believe all the logical consequences of their beliefs, not even those which take merely a step or two to deduce. The above implication holds because the set of possible-worlds in which the proposition $(p_1 \& p_2 \& \ldots \& p_n)$ is true is identical with the set of possible-worlds in which the proposition $(p_1 \& p_2 \& \ldots \& p_n \& q)$ is true, since q is a logical consequence of $p_1, p_2, \ldots, p_n$. Hence believing that

$(p_1\&...\&p_n)$ entails believing that $(p_1\&...\&p_n\&q)$. If we assume that a conjunction is believed iff all its conjuncts are believed, then it follows that $Bap_1,...,Bap_n$ implies Baq.

(iv) The problem of **contradiction:** since a contradiction entails everything, Bap and Ba-p entail Baq for all q. Thus possible world semantics excludes the realistic possibility of inconsistent beliefs.

Any theory about propositional attitudes has to deal with notorious counterexamples that stem from the resistance of these attitudes to logical laws. Most philosophers who have attempted to deal with these problems have looked for technical solutions to the question of how to correlate a suggested logical form of a belief sentence with the fact that agents may have very little "logic" in their belief systems.

Some solutions claim the objects of attitudes to be things other than propositions, objects such as sentence tokens or more abstract objects. Others avoid the question of fixing the objects of belief by taking a sentence like 'a believes that p' to be of a subject-predicate form, in analogy with 'a is fat'. (They take 'believes that p' to be unanalyzed constituent.) Those who swear loyalty to propositions are led to a fine grained notion of proposition by making their identity conditions stronger and complicated (like intentional isomorphism, synonymity, etc.).

Stalnaker rejects these attempts, even when they seem to overcome some of the problems that the identity conditions of objects of belief encounter. He claims that by their nature, these attempts put stumbling blocks on the route to a larger semantic theory about the content of expressions. In the following pages I will study the major devices Stalnaker uses to explain away the problems mentioned earlier. These devices are:

1. "Expresses" relation, which holds between a sentence and a proposition, is complicated and not always known to the agent. In particular, mathematical ignorance can be explained as ignorance about the multiple ways a sentence may express the one universal proposition.

2. Belief-states (a set of propositions believed) can be grouped in a way that blocks the conclusion Ba(p & -p) from the premisses Bap and Ba-p, when each of the two premises belongs to different belief-states. More generally, the way the notion of a belief state is defined puts restrictions on conjunctions of beliefs.

3. Context relativity is a multiple purpose device, whose major contribution is in blocking some inferences in belief contexts. If p and q belong to different context-sets, then Baq can not be inferred from Bap, even if p is logically equivalent to q or

logically entails it. This device will be discussed in the next chapter.

Stalnaker uses the first device - the "expresses" relation - to deal with the problem of equivalence, i.e., to explain away the fact that the possible-worlds theory appears to make the following argument valid;

a believes that p
p is necessarily equivalent to q
---
a believes that q

Stalnaker points out that this argument does not hold in the above form, due to a confusion between use and mention: what a believes is a proposition, but that which are necessarily equivalent are expressions (sentences which express propositions). It may be objected that the conclusion would still follow, if we take the semantic relations of "S expresses that p" to be necessary. However, Stalnaker stresses that this is implausible: "Whenever the structure of a sentence is complicated, there will be a nontrivial question about the relation between sentences and the propositions they express, and so there will be room for reasonable doubt about what proposition is expresses by a given sentence" (p.72). Apparently valid arguments of the above form, that have true premises and a false conclusion, are explained as the agent's failure to realize that one of the propositions he believes is expressed by that sentence.

Stalnaker's ideas can perhaps be made clear by considering mathematical beliefs. Possible-worlds semantics appears to lead to the most implausible conclusion here, since there are just two of them: the necessarily true and the necessarily false ones. If the object of belief is a mathematical (true or false) proposition, it is hard to avoid the conclusion about believing any other mathematical proposition (true or false respectively), whether or not the believer recognizes all the ways of expressing them. For that particular class of beliefs, Stalnaker provides a new kind of object: "it seems reasonable to take the objects of belief and doubt in mathematics to be propositions about the relations between sentences and what they say" (p.73). These semantical propositions are contingent. The statement "2+2=4" does not mean that two plus two equals four in all possible worlds. If a language is defined not in terms of the semantic properties of its sentences, but rather in terms of the community of its speakers and the physical properties of the linguistic expressions, then we have to allow a possible world in which some epistemic counterpart of the expression "2+2=4" says something different then our standard interpretation of it. O'Leary, who believes that two plus two is four, but either fails to realize that the sentence expressing that belief is necessarily equivalent to "there are an infinite number of primes", or, worse - sincerely dissents from the latter, lacks the information, according to Stalnaker, about the relationship between the

sentence he fails to believe and the one necessarily true proposition.

The difference between a **belief state** and **belief attribution** now becomes important. By Stalnaker's proposal, what we believe, doubt, are ignorant about in mathematical or logical contexts are contingent semantic propositions of the general form 'sentence S expresses the proposition p'. But when we attribute beliefs, we use the sentences which (standardly) express the necessary propositions. We would say about O'Leary: "He believes that 2+2=4, but does not believe that there are an infinite number of primes". In this sentence the complement 'that 2+2=4' is not used to pick out the necessary proposition, on Stalnaker's view, but the contingent proposition 'that the expression '2+2=4' expresses the necessary true proposition'. In order to bridge the gap between the objects of belief-attribution and the objects of belief, we need a semantics for belief attribution sentences that recognizes the fact that the relationship between the sentential complements of sentences attributing beliefs and the contents of the beliefs attributed is far more complicated (mostly due to context dependency) than people tend to take into account. (Stalnaker's proposal for such a semantics will be discussed in the next chapter.)

Stalnaker believes that his suggestion provides a way of reconciling the possible worlds analysis with two obvious

facts: that there is a plurality of mathematical truths (many contingent propositions about what mathematical statements express) and that there can be a nontrivial belief in necessary truths and falsehoods. Even though there are only two mathematical propositions, there are infinitely many contingent propositions relating expressions to them. Since the latter are the objects of belief in mathematics, we obtain plurality of mathematical beliefs; and because of the complexity of the relations between expressions and their content, believing a mathematical proposition may not be trivial.

This proposal, as it stands, looks very implausible: if the objects of mathematical beliefs are propositions about the relations between sentences in a specific language (say English) and the world, how can we explain the mathematical beliefs of agents who do not use this language? Do the Chinese and the Hindu mathematicians believe different things when they believe that there are infinitely many primes? Stalnaker might have anticipated these questions: in one place in the book he weakens his claim. Some kinds of mathematical propositions, he states, convey information not about the relations of sentences to the necessarily true proposition, but rather about the relations of some structure and the one necessarily true proposition: "For these cases we might take the objects that beliefs and doubt are about to be a common structure shared by many, but not all, of the formulation which express the necessarily

true proposition." (p. 74). This new way of looking at the (or just some) objects of mathematical beliefs may answer the problem mentioned above, if it were made more explicit and detailed, which unfortunatelly it is not.

An argument that was presented by Larry Powers (1976) and by Saul Kripke deserves attention, since Stalnaker does not seem to provide a satisfactory answer to it. Consider Alfred, who is familiar with some axiomatic system of first order logic. The facts are that each axiom sentence expresses a necessary truth contingently, and that the application of rules of inference to sentences expressing necessary truth results in sentences that express necessary truth -- but Alfred knows these facts. It follows that Alfred knows (or believes) every theorem sentence of this system, but surely he does not. Stalnaker views this as a special case of the problem of deduction -- if it could be shown that the person who believes $\underline{p}_1,\ldots,\underline{p}_2$ does not necessarily believe the consequence $\underline{q}$, then the objection could be overcome.

Let us now turn to the central problem of deduction. To solve the problem Stalnaker invokes the concept of a "belief state". He thinks that a person can be in several belief states at the same time, and that the cases where the person fails to believe consequence $\underline{C}$ where he believes premises $\underline{p}_1 \ldots \underline{p}_n$ that logically imply $\underline{C}$, are to be understood as those cases where some of the premises $\underline{p}_1$,

$p_2, \ldots, p_n$ belong to different belief states within this person. The process of deduction, according to Stalnaker, is the process of integrating separated belief states.

Stalnaker distinguishes three deductive conditions that we might put on beliefs. Let A be a set of propositions believed, and p, q, be propositions:

1. If $p \in A$ and $p \rightarrow q$ then $q \in A$ (deductive condition)
2. If $p \in A$ and $q \in A$ then $(p \& q) \in A$ (closure condition)
3. If $p \in A$ then $-p \notin A$ (consistency condition)

Stalnaker claims that these conditions hold within each belief state, but not for the totality of all the agents' beliefs. What do we achieve by limiting the application of the closure and the consistency conditions to belief states only? -- Although by this theory an agent believes all the consequences of all of his beliefs that belong to the same belief state, it does not follow that he believes a consequence, if the premisses belong to different belief states.

Moreover, if a person happens to believe that p and that -p at the same time, it does not follow that he believes everything, which would have followed by the deductive condition, if the closure condition is imposed on all the beliefs. The restrictions put on the application of these two conditions seems to solve the problems of conjunction and contradiction in belief context.

The answers to the problems of equivalence and deduction are given, as mentioned, by the device of a complex "expresses" relation between expressions and propositions, and by the device of separating beliefs into belief states. The proposition believed within one belief state constitutes a consistent deductively closed set, but the totality of all beliefs may fail to do this.

Before going any further into this idea, we better have a closer look at the concept of a belief state. Bear in mind that in order for such a central concept to do the work it is intended to do, its definition should incorporate three elements: a) the causal aspect; b) the pragmatic aspect; c) propositions being the objects of belief, defined as sets of possible worlds.

Stalnaker gives several definitions of the concept of a belief-state:

(i) A belief state is a **conception** of how the world is, i.e.: two belief states are the same iff they are the same conception of how the world is.

This definition may be understood as capturing the "backward-looking" aspect of belief, which is explained in terms of indication. A belief state can be viewed as a positive account of information that is accumulated from the outside, that has stood up to whatever fidelity conditions which are implicit to the situation, and has become part of what we actively entertain and understand. In order to be

in two different belief states the agent need not have two radically different conceptions about the world; they may be identical in all aspects except for a possibility recognized in one but not in the other.

(ii) A belief state is characterized by a **set of possible worlds**, those not excluded by the agents' conception of the way things are, i.e.: the proposition believed is true in all the possible worlds in the belief state. Note that this definition has a negative characteristic: "to believe that p is simply to be in a belief state which lacks any possible world in which p is false... having too many beliefs is the same as not recognizing enough possibilities... a person can have beliefs by default, or through lack of imagination..." (p.69).

According to this definition, everything we do not explicitly exclude as possible is contained in some belief state or another. The result of inquiry, addition of information, is "shrinking the size" of belief states, by ruling out possibilities we passively allowed in. This conception of a belief state, though it is flexible enough to permit deductive closure and even incompatible beliefs, does not seem to conform to the way we tend to attribute beliefs. We will not say that a believes that p just because he has never thought about it. Stalnaker would probably answer that he separates the content of beliefs

from the way we attribute them to others. Still another problem remains: if our beliefs have anything to do with our behavior, the negative definition may exclude the possibility of explaining both how we form our beliefs and how they interact with (positively defined) desires to produce actions.

(iii) Beliefs are **behavioral dispositions.** To be in separate belief states is to have separate dispositions, when those dispositions are vaguely individuated according to their "domain of display" i.e., the different kinds of situations in which they are displayed. This definition seems to envelope the pragmatic element of the theory. That an agent a believes that p means that the actions that are appropriate - those he is disposed to perform - are the actions that will tend to serve his interests and desires in situations in which p, together with other relevant beliefs in the belief state, is true.

It is worth stressing that Stalnaker characterizes belief in a negative way: a person believes that p relative to a belief state if and only if there is no world in the set of possible worlds representing that belief state, in which p is false. Thus Stalnaker is able to explain tacit beliefs and presuppositions. For example: William the III of England believed, in 1700, that England could avoid a war with France. Did he also believe that England could avoid a nuclear war with France? Stalnaker's answer is that if this

proposition was not believed, then at least it was tacitly presupposed, in the "negative" sense by which believing a proposition is ignoring the possibility that it is false.

Sometimes the context determines what alternatives are relevant. An example that illustrates this is that of a man standing in front of a zebra cage in the zoo, the sign says "zebras", and the man, who believes he sees zebras, says so to his son. Does he also believe that these are not cleverly disguised mules? - after all, this follows logically from his belief that he sees zebras. Stalnaker's answer, as I understand it, is that in a normal context it is correct to say that the man believes that he sees zebras, since normal context excludes the possibility of the zebras being mules cleverly disguised. But in other contexts, where this possibility is not excluded, the attribution may be problematic.

Stalnaker sees the problem of deduction as posing two questions:

1. What is the nature of information conveyed in a statement about deductive relations?

2. How do we acquire this information?

Any declarative sentence may be viewed as determining two propositions: the proposition that is expressed, according to the standard semantic rules, and the proposition (of the general form 'S expresses that p') which

talks about the relations between the sentence and the proposition it expresses. The information conveyed in a statement about deductive relation is, according to Stalnaker, of the second type: "The subject matter of mathematical propositions is notation or structure exhibited in notations" (p.86). A failure to see that a proposition is necessarily true or necessarily equivalent to another is a failure to see what propositions are expressed by the expressions in question.

The second device, that of belief states, responds to the second question: we acquire the information about what sentences express mathematical truth or falsehood by putting separate belief states together, but doing that may not be a trivial task. Acquiring information amounts to ruling out more possibilities than we did before. Information, a content of a statement, must be contingent according to this account. The objects of deductive inquiry, being semantic propositions about the relations between expressions and the two necessary propositions (rather than the necessary propositions themselves) satisfy this requirement. Making these objects contingent, the theory achieves the following goals: (1) it explains ignorance of necessary propositions and of propositions equivalent to believed ones, and explains the plurality of mathematical truths and the difficulty of acquiring them, without giving up the idea that propositions (set of possible worlds) are objects of belief and inquiry; (2) it gives an account of the

epistemology of deductive knowledge - putting one's separate belief states together - as a special case of inquiry in general. "Inquiry in general is a matter of adjusting one's beliefs in response to new information, but in the case of deductive inquiry, the information that initiated the change is new, not to the agent, but only to one of his belief states. By dividing the agent into separate centers of rationality, we make it possible to see the processing of the information an agent already has as a phenomenon with the same structure as the reception of new information" (p.87).

Stalnaker offers two different solutions to the problem of mathematical ignorance. I. A mistake in calculating the result of 33+46 is due to not integrating separate belief states: one with the correct belief about the result of adding 6 to 3, another contains the correct belief about the result of 30+40. This suggestion does not seem to utilize the idea of the objects of mathematical beliefs being semantic propositions about the relations between sentences and what they express. The ignorance of this mathematical truth (his example) is treated on par with any other failure of combining any two beliefs p, q, into a belief in p&q. II. A mistake in calculating the result of 33+46 is due to not realizing either the relations between one or more of these numerals and the numbers they are intended to denote, or that the operator sign used in the term above denotes addition. Stalnaker admits that this is implausible, but

takes it to show only that his account is incomplete, not that it is mistaken.

The device of separate belief states solves the problem of having to believe a consequence of a several premisses, but a problem still remains: by this account, it follows that the deductive consequence of a single sentence is always believed. For instance, every sentence of the form '-p' is equivalent to the sentence that asserts that p is equivalent to some contradiction. It is easy to imagine that one could believe one of these two equivalent sentences but not the other. We can even imagine the possibility of believing a sentence p without believing that p or q.

Stalnaker's basic idea is to try to understand mathematical and logical beliefs on the model of empirical beliefs. Both of these, according to him, convey contingent information. But - while the latter conveys this contingent information to the belief system as a whole, the process of deduction is viewed by Stalnaker as providing new information to the belief states individually. Deduction becomes the process of integration of belief states. We have very few good theories about mathematical and logical knowledge. Stalnaker's theory, though it has the difficulties mentioned above, is an intriguing new attempt to provide an insight into mathematical and logical inquiry.

Stalnaker's suggested solution to the problem of deduction is not well worked out. The nature of the objects

of beliefs in mathematics is not well defined: are they actually the propositions expressed by the sentences, or propositions about this "expresses" relations, or propositions about the relations of structure to meaning? What belief states do we need to integrate in order to eliminate different kinds of mathematical ignorance (for instance: a mistake in calculating 7+4)? How does the complement of a belief sentence in mathematics pick up a contingent proposition (other than the necessary ones)? Stalnaker's attempt to answer these questions within a pragmatic theory of context is examined in the next chapter.

Chapter 4

The Possible Worlds Theory: Context Dependency of Belief Attributions

Most of the semantical theories about belief sentences follow the strategy of positing one kind of objects of belief or another in order to overcome the logical problems involved. In a sentence of the form Bap, these theories take the complement p to denote a more or less definite content that is built up in accordance with the conventional semantical rules; so the clue for solving the logical problems, according to these theories, lies within the identity conditions for that content. Stalnaker's basic strategy is, on the other hand, to employ a simple theory for the objects of belief and a complex theory about the relations between sentences and their content. So, though the objects of beliefs are the familiar sets of possible worlds, detecting what set of worlds is expressed by a given sentence is a complicated task, for one has to consider not only the semantic features of sentences, but also their context of use.

Stalnaker wishes to draw a distinction, which people may not have been fully aware of, between the **philosophical** theory about belief, and the **semantic** theory about belief attribution. A belief, according to him, is just a two place relation between an agent and a proposition (a set of

possible worlds). A theory of belief attribution, according to Stalnaker, needs to recognize that the question of what particular proposition is related to an agent a in the belief attribution Bap does not have a simple answer, and depends on contextual factors. The context dependency of some sentences is obvious: the proposition that a sentence like 'I am hungry' determines depends crucially on the denotation for 'I' that the relevant context supplies. The novelty of Stalnaker's approach lies in extending the use of context to account for parts of the natural language that seemed free of such dependency. In particular, Stalnaker wants to explain the logical behavior of belief attributions by appealing to various pragmatic factors.

Stalnaker's theory about the nature of belief was laid down in detail in the previous chapter. The "highlights" are that belief state should be characterized both by the way it is caused and by the actions it disposes the believer to perform in relation to his desires. A belief state is represented by the set of possibilities which are not excluded by the agent's conception of the way the world is. The set of propositions believed, relative to a belief state, is just the set of propositions true in all those possibilities. To say that a believes that p is to characterize a's belief state as one included in the set of possible worlds in which p is true.

The notion of context plays a dual role in explaining linguistic phenomena. In its first role, the context provides the denotations of pronouns, the range of the quantifiers, the reference of indexicals etc. In this role the context supplies the information needed to figure out what the sentence literally says. The second role the context plays puts the resulting interpretation into a more complex explanation as to the reasons for performing that speech act and the goals it is supposed to accomplish. As the general setting in which communication takes place, the context fixes the background of presumed beliefs and purposes of the participants. In virtue of these, the meaning that is actually communicated by a sentence may differ from its literal meaning.

Grice (in "Logic and conversation") illustrates this phenomenon and puts forward some principles and maxims to explain it. Among them are instructions like "make your contribution to the common purpose recognized by the participants informative, relevant, as truthful as you believe it to be, nonambiguous, brief, clear, orderly". Under certain contextual conditions, given that these maxims are presumed to be preserved by the speaker, and that a mutual background knowledge and the conventional meanings of the words used are available to all participants, an expression may convey a suggestion, or imply, or altogether mean a proposition which is not conventionally associated with it. In a conversation about a bank teller Mr. B, a

person A says: "B has not been to prison yet". This sentence may make the conversational implicature informing the hearer about A's opinion of B's honesty, or on banks tellers tendency to yield to criminal temptations provided by their occupation. If A never meant to imply that, then a violation of some conversational maxim (relevance, precision in delivering information) occurs, rendering the utterance inappropriate or misleading in its context. A letter of recommendation for a candidate in philosophy which reads only: "my student X has a beautiful hand writing" imparts information about the student's ability as a philosopher, because we assume that the teacher knows that other information is required, that he is reluctant to be clearer about it, and that he trusts the reader to read "between the lines".

The overall role of the context in a theory of conversation is to supply the information needed for a correct interpretation of a linguistic performance. Stalnaker reifies this information as a set of possible worlds -- worlds in which the common presumptions of the participants in the conversation are all true. Stalnaker calls this set the basic context set. This set is the set of live options as to how the world might be, that the participants in the conversation distinguish between, and take each other to do the same. It is continually updated in the course of a conversation, as information exchanged between participants is accepted. In that way, the

linguistic actions that the context is used to interpret also end up modifying the basic context set.

For example: the basic context set will fix Noel as the referent of the pronoun 'he' in the sentence 'he exercises a lot of self control', if every participant in the conversation shares the background knowledge about Noel's unfaithful wife and his "cool" behavior in confronting the evidence. The basic context set will not determine Noel as the referent if there is even one participant, ignorant of the juicy gossip about the couple, who takes this sentence to be about, say, Bob, the hidden alcoholic, who has not touched a drink in two days. In the second case, no proposition is being determined by applying the propositional function 'exercises a lot of self control' to the individual unsuccessfully referred to by "he". The basic context set may expand into the one described in the first case once the relevant story is shared with the ignorant participant.

Stalnaker points out that when belief sentences are used, there is an additional context that comes into play. The proposition expressed by '_a_ believes that _p_' is a function of another proposition, denoted by 'that _p_'. The latter may contain, for example, a definite description which may have one definite referent in the basic context, but another one as far as _a_'s beliefs are concern. The belief attribution 'Phoebe believes that her partner is

stupid' will not fix its proposition properly if we do not consider the fact (known to all participants) that the one she believes to be her partner is just a dummy figure put there by her real, very shrewd partner in business, for tax evasion purpose. In order to give the embedded sentence of the belief attribution a correct interpretation, one needs to consider the presuppositions that the participants in the conversation have about Phoebe's beliefs. These presuppositions form the embedded or the derived context set out of the basic context set in the following way: "For each possible situation in the basic context, Phoebe will be in a definite belief state which is itself defined by a set of possible situations - the ones compatible with what Phoebe believes in that possible situation. The union of all the possible belief states will be the set of all possible situations that might, for all the speaker presupposes, be compatible with Phoebe's beliefs. This set of possible situations is the derived context for interpreting the clauses that are intended to express the content of Phoebe's beliefs" (Stalnaker's "Belief attribution and context" 1987, p.12).

Both the basic context set and the derived context set are useful in guiding the interpretations of the utterances, and in explaining their appropriateness. Stalnaker's example: the statement 'Harry regrets taking the bribe' is inappropriate in a context where the participants do not accept that Harry was offered a bribe and took it.

Similarly, the sentence 'Phoebe believes that Harry regrets taking the bribe' requires a context in which the participants assume that Phoebe's believes that a bribe was offered to Harry, and that he took it. For the belief attribution statements to be appropriate, it is not necessary for these presuppositions about Harry to belong to the basic context set. The explanation of the belief sentence is parallel to that of the subordinate clause: the role that the basic set plays for the first is the role that the derived set plays for the second.

The function of context in Stalnaker's theory is not only to provide the background presuppositions for a conversation, but also to demarcate the possibilities relevant to the conversation, the possibilities that are used to define the propositions expressed. The set of possible worlds which defines a proposition is, according to Stalnaker, "a space of relevant alternative possible states of some limited matter, determined by a context in which some rational activity (deliberation, inquiry, negotiation, conversation) is taking place" (same paper, p.4). The pragmatic aspect of the theory tells us what set of possible worlds are relevant to the explanation of a particular linguistic phenomenon. This aspect of context is useful for explaining certain belief attributions, for instance, when the agents are animals or children (discussed above). This aspect, however, does not play a significant role in

explaining away the logical problems concerning belief sentences.

Let us now review the major problems about belief, and see how Stalnaker may use the pragmatic theory of conversation to solve them. The first and major problem of belief sentences is the problem of equivalence: how are we to explain Alfred's believing that p but not believing that q, even though q is synonymous to, or logically equivalent to, or semantically isomorphic to p? The second problem is associated with attributing to an agent a belief in logical falsity: what content can such a belief have within the possible worlds semantics? The third problem is that of deduction, and its special instance, the problem of conjunction -- under what conditions can we correctly ascribe to an agent the belief in a logical consequence of his beliefs?

If p and q are logically equivalent, then by the possible worlds semantics p and q express identical propositions. The pragmatic element that Stalnaker incorporates into his explanation does not overrule this identity, but rather allows for situations in which the expressions 'that p' and 'that q' denote different propositions (different sets of possible worlds). As mentioned earlier, the relationship between the expression 'that p' and the proposition p is considered complicated

enough to enable the first not to denote the second, under certain contextual conditions.

The theory of conversation provides us with a mechanism in virtue of which equivalent sentences, uttered in a context, convey different information, and thus correspond to different sets of possible worlds. An example from Grice Stalnaker might use is the following: consider two sentences, each of which could be used by a critic to describe some musical performance:

> Miss X sang "Home sweet home".
>
> Miss X produced a series of sounds which correspond closely with the score of "Home sweet home".

It is clear that more information is conveyed by the second sentence, perhaps about the some defects in the performance. We need to consider whether this kind of device helps Stalnaker in dealing with the logical problem at hand.

Let us recall Bigelow's example, considered earlier:

(1) Robin will win.

(2) Everyone who does not compete, or loses, will have done something Robin will not have done.

Since (1) and (2) are necessarily equivalent propositions, they are identical by the possible world semantics. The problem is to explain how in the sentences

(3) Marian believes that Robin will win.

(4) Marian believes that everyone who does not compete...

the complements pick out different sets of possible worlds. The use of the derived context set can explain the difference between (3) and (4) -- if they are associated with different derived sets, then (1) and (2) can express different propositions. But this does not solve our problem, since in a single context the two do pick out the same set of possible worlds by Stalnaker's theory.

The maxim that may be useful here is that one should make one's contribution as brief as possible. Sentence (4), it can be argued, does not comply with that, so it is not conversationally equivalent to sentence (3). It is the very length of the sentence that is the factor in communicating the extra information, just as in the example considered earlier about Ms.X's singing.

The Gricean maneuver is useful for the explanation of contextual interpretation of sentences, but it faces some difficulties when applied to belief attributions. This theory does not yield any principles for determining the proposition corresponding to a sentence in a context. All we may obtain for this example is that (3) and (4), when used in a situation, are assigned different interpretations. The theory does not necessarily assign different objects for Marian's beliefs. For all we know, (3) and (4) may be

conversationally different in that one of them expresses something which is not about Marian's beliefs at all.

Another difficulty: one of the necessary conditions for the occurrence of a conversational implicature is that all participants know the standard proposition that the sentence in question expresses. In the above example, it is in virtue of the fact that the participants recognize that (1) and (2) are equivalent, that the Gricean maxim comes into play and yields the implicature that the speaker is trying to communicate something else with (4) then merely (3). (The same mechanism is involved in the singing example.) However, this does not provide an explanation for the difference between the (3) and (4) in those situations in which the participants do not recognize the equivalence of (1) and (2). In these situations the maxim of brevity does not come into play.

Let us now turn to the second problem listed above: what belief state is characterized by attributing to Sarah the belief that 34+77=101? The proposition she believes is true in no possible world. According to Stalnaker's theory, sometimes the complement of a deductive or mathematical belief attribution is used to pick out not a necessary proposition, but rather a contingent proposition about the relation of the sentence used to the necessary proposition it expresses. This contingent proposition in our case is: 'the sentence '34+77=101' expresses the necessarily true

proposition'. The worlds in which this proposition is true are those in which the sentence '34+77=101' happens to mean that there is no largest number, or that 2+2=4, or that 34+77=111 or any other logical truth. Is that what we, intuitively, take the believer of a miscalculation to believe? Is the proposition Sarah believes metalinguistic or semantical in character? Even if a case could be made for turning mathematical beliefs into semantical ones, there is another difficulty: The belief attributions made by the following two sentences: 'Sarah believes that S' and 'Sarah believes that S expresses the necessary truth' do not necessarily share the same truth value. The first may be true, but the second one is not, if Sarah does not understand some of the notions employed, like that of a necessary truth or its uniqueness. The second may be true, but not the first, if the belief was acquired by accepting that a certain formula S expresses, say, a mathematical axiom, but the notations used in formulating S are foreign to Sarah.

Stalnaker explains the case of O'Leary, who believes that 2+2=4, but fails to believe that there is no largest prime number, by observing that the relations between sentences and the propositions they express are complicated by pragmatic considerations to the point of having to entertain the possibility that the sentence '2+2=4' does not express a necessary truth. A basic difficulty with this idea is the lack of a systematic explanation, in the theory,

for the type of proposition that the complement of a belief attribution picks out. We are not told the conditions under which we shift from considering the p in Bap to stand for the set of possible worlds in which p is true, to the set of worlds in which the proposition "P expresses that p" is true, to the set of worlds in which "the structure exhibited by P expresses that p" is true.

Furthermore - even if we were supplied with an account that would enable us to discern between the options, or were confined somehow to treating all and only contents of mathematical belief attributions as of the second type (sentence S expresses proposition p) - some difficulties remain. We are not told by Stalnaker how to construct (or just understand) a possible situation in which the proposition "that there is no largest prime number expresses a necessary truth" is false, within a context of a specific community of speakers whose basic context set contains all the semantic and pragmatic information needed to give the sentence "there is no largest prime number" a unique interpretation. The proposition expressed by this sentence, for lack of nontrivial pragmatical conditions, is bound to be true in all those worlds in the derived set in which the proposition expressed by '2+2=4' is true. All of Stalnaker's examples for contextual restraints: pronouns, time-space indices, conversational implicatures - are irrelevant for these cases. In the derived set, the proposition "2+2=4" is true in every possible world, once

all participants recognize it to be believed by O'Leary, and it is very hard to be convinced that, for some unexplained mysterious reasons, another sentence, made in the same language, expressing by all (possible worlds') semantic and pragmatic considerations exactly the same proposition, somehow fails to do that.

In his book Stalnaker claims to have solved the problem of conjunction (given that Bap and Baq - does it follow that Ba(p & q)?), and the problem of contradiction (which follows from the previous one by replacing q by -p) with the help of the device of separate belief states. If Bap and Baq belong to different belief states, then, since the closure condition applies only to propositions within a belief state, the conjunction of the contents of those beliefs is blocked. It would appear that the device operates in an ad hoc manner -- whenever we can not safely attribute a belief in a conjunction to an agent, Stalnaker would claim that the conjuncts belong to separate belief states -- unless the device of context relativity determines ahead of time how to separate belief states.

The context relativity of belief attribution is a very useful notion in many aspects. It tells us what possible worlds are contained in the basic context set, and what worlds inhabit its derived context set. A successful attribution of a belief to Tom characterizes a certain belief state of Tom as one that is true in each of the

worlds in the derived context set -- a context set for whose construction we have clear instructions. But the derived set can be used to characterize any number of belief states that Tom may have, and we are given no tools for distinguishing one from the other.

We may have the following options:

1. Every belief attribution ascribes a belief with respect to a different belief state. The problem with this idea is that it undermines the whole causal pragmatic basis of the theory. Tom's disposition to go to the restaurant on A street may be caused by his desire to eat, and by believing the following two propositions: 'The restaurant is either on street A or on street B'; 'The restaurant is not on street B'. A theory that splits the attributions of these beliefs into separate belief states loses its force as a pragmatic explanation of actions in terms of beliefs and desires.

2. In each context, all belief attributions to the same agent will characterize the same belief state of that agent. It is obvious that following this possibility leads to attributions of beliefs the agent in question does not have.

3. The only option left is the one which recommends being selective - some belief attributions to an

> agent belong to one belief state, and some others to another. If we had the principles required for such a demarcation, the problem of belief attribution would have looked solvable. Unfortunately, we do not obtain them from Stalnaker.

The device of separate belief states should have been incorporated into the notion of the derived context set. As it stands, the theory allows indefinitely many derived sets, each corresponding to a belief state, and no clue is provided as to the mechanism for separating any of these from the others. We are left at the starting point of the puzzle: under what conditions can we successfully attribute to Tom a belief in a conjunction, the conjuncts of which he believes?

As flexible as Stalnaker makes his "pragmatic" possible world semantics, it still does not explain deductive ignorance away.

ADDENDUM

After completing my study of Stalnaker, three critical papers on the subject came to my attention: Michael Pendlebury's "Stalnaker on inquiry" Journal of Philosophical Logic 16, August 1987 pp.229-272; Stephen Schiffer's "Stalnaker's problem of intentionality" in Pacific Philosophical Quarterly 67, 1986, pp.87-97; Hartry Field's

"Stalnaker on intentionality" same issue, pp.98-112. Stalnaker's response to the latter two is published in the same issue of the PPQ (pp.113-123). Some of the objections that I have raised above are also put forward by these critics. Below I consider Stalnaker's replies.

Two major objections were brought concerning the causal aspect of Stalnaker's theory. The first deals with the notion of "fidelity conditions", and the second with the notion of "indication relation". Both Field and Schiffer doubt whether the notion of normal or optimal conditions could be spelled out without using intentional or semantical notions. Stalnaker admits that in his presentation the idea of fidelity conditions is not fully spelled out in nonintentional terms. As a matter of fact, we observe, they have not been spelled out at all, in any kind of terms. Stalnaker concedes: "The reference to normal conditions is a kind of promissory note, but I think it is a note to be cashed in future empirical inquiry, not future conceptual analysis" (PPQ. p.118). But even then in the future, "...I do not think it will ever be possible to give a global account of fidelity conditions" (PPQ. p.118).

Stalnaker's definition of "indication" has, as mentioned earlier, the objectionable consequence by which physically or causally equivalent propositions are identical. This point is also observed by Schiffer, who brings the example of a person believing that Sally has the

flu (p), and who ends up being also believing that q, where q is the underlying biological condition that is causally sufficient and necessary for p. Stalnaker admits that this is serious problem for his account. He has no solution for the problem, but he mentions a strategy that may be helpful in solving it. The strategy is essentially the same as the one discussed above for dealing with the problem of logical equivalence. It relies on the by now familiar devices of the context dependency of belief attributions, the complicated relations between sentences and their content, and the ability in his theory to separate belief states. Stalnaker is a long way from making a convincing case for all these devices even for solving the problems they were intended for, so we may be left with the unpleasant feeling that the soldiers he calls to fight his battle are either already dead, or at least in a great need for an overall operation.

The solution Stalnaker offers to the problem of deduction involves two complementary strategies: A. the "metalinguistic" move by which some beliefs are about the relation between a sentence and a proposition it expresses; B. the "compartmentalization" of belief states (the quoted label is Stalnaker's). Field raises the objection, also raised above, about the unintuitive application of the "metalinguistic" strategy to subject matters having no linguistic content. Stalnaker says in reply: "The strategy, generally, is to take a proposition to be about a

relation between a representation, which could be something other than a linguistic representation, and its content." Stalnaker also denies that a proposition about the relation between a sentence and a proposition is about a linguistic or a semantic fact.

Field brings an example of a single inconsistent belief state, and questions each of the strategies Stalnaker would employ to solve the difficulty. Field asks us to consider Cantor's belief state while developing his set theory in which an instance of the naive comprehension schema proved logically inconsistent. Stalnaker could try to ascribe to Cantor a belief state whose content is that these sentences (that we take to express the comprehension axioms) express truths. Field doubts whether the move will work: it succeeds in making the content of the belief state consistent, but creates problems when we consider the metalinguistic beliefs about other relevant facts that Cantor had, like those about the meaning of the symbols (membership, conjunction etc) employed in his set theoretic language. We end up with an inconsistent metalinguistic belief, so the shift from object level propositions to metalinguistic propositions appears to have gained nothing. At this point Stalnaker is likely to resort to the device of separating belief states, since that strategy is used to explain the possibility of having inconsistent beliefs. This move seems to Field plainly inadequate. The belief about the comprehension axioms expressing a truth and his

other metalinguistic belief about the meaning of these axioms obviously are involved together in explaining Cantor's actions. Separating the two is implausible. In relation to this point, Stalnaker brings more examples for fragmented beliefs, an idea no one disputes is applicable to many cases.

Stalnaker attempts to combine a metaphysical theory (the causal-pragmatic one) with a semantical analysis of propositions (in terms of possible worlds) to obtain a theory that attempts to explains the logico-semantic features of belief attribution sentences, and also provides an account of what belief is. The arguments made for adopting the suggested semantics on the basis of the metaphysics are not convincing. The pragmatic aspect of the theory proves to be circular, but the causal one called in to remove the circularity fails to pin down the content of beliefs. The logico-semantic proposal for the analysis of belief attributions resembles a theory of competence, applicable to ideal believers only, and is impotent against a whole stock of counterexamples about beliefs that perfectly rational agents may have. A lot of promissory notes given to the reader of Stalnaker's book and paper are as yet left unpaid. The essential promissory note concerns the context relativity of belief sentences, which is supposed to have the potential power of solving the above difficulties. But, as I have argued in this chapter, the

applications of these devices to belief context face formidable difficulties.

Chapter 5

Structured Meanings[1]

There are, as we have seen, two broad strategies for dealing with the problem of logical opacity of belief attributions. First: individuate meanings so finely that even strongly equivalent sentences are assigned different meanings. This option has its advantages: it preserves the idea that the complement sentence is a genuine semantic constituent of a propositional attitude sentence, and it preserves what is commonly considered as Frege's compositionality principle--the principle that the meaning of a whole is determined by the meanings of its parts. Its disadvantage is that when meanings are individuated so finely, the semantic rules--the rules that tell us how the meaning of a complex expression is built out of the meanings of its parts--reveal little concerning an expression's logical and semantic properties.[2] Second: accept an account of meaning that is more closely tied to truth and inference but deny that sameness of meaning of the complement sentences p and q implies the

---

1. This chapter is an expanded version of a critical study of Cresswell's Structured Meaning (1985), published jointly with Anil Gupta in Journal of Philosophical Logic 16, No. 4 November 1987 pp.395-410. I wish to thank Anil for his helpful suggestions on the first draft of this study.

2. For a theory that follows this strategy see Thomason (1980).

sameness of meaning of, for example, the propositional attitude sentences '_x_ believes that _p_' and '_x_ believes that _q_'. This option does not force us to abandon Frege's compositionality principle. We can keep Frege's principle, and with it semantic theory in its traditional form, but we have to say that the complement sentence _p_ is not a genuine semantic constituent of the sentence '_x_ believes that _p_'. Cresswell, following the second option, constructs in his _Structured Meanings_ a "post-modernistic" solution to the problem of propositional attitudes--a solution that combines some old ideas with some new technology.

SECTION I

Cresswell's general theory of meaning has two components: (1) the familiar possible-worlds account of the meaning of sentences and names and (2) the principle, which we shall call _The Principle of Categorial Semantics_, that the meaning of a whole is obtained from the meanings of its parts through application of functions to arguments.[3] This second component serves as an important guide and constraint in the construction of Cresswell's semantics. The way it works can be made clear through some examples. Let us recall that the meaning of a sentence, according to the possible-worlds theory, is a set of possible worlds, the set whose members

---

3. Cresswell also allows λ-abstraction.

are all and only those worlds in which the sentence is true. (We ignore tense and other indexical phenomena.) The meaning of a proper name, according to the theory, is just the denotation of the name. (This assumes that all names have denotations and are rigid. A more complex account of names, one that is independent of these assumptions, is possible.) The Principle of Categorial Semantics leads naturally to the following account of the meaning of a one-place predicate. Since a one-place predicate F combines with a name a to form a sentence Fa, and since the meaning of Fa is to be obtained by function application, the meaning of F should be a rule or function that when applied to the meaning of the name a gives the meaning of Fa. That is, the meaning of the predicate F should be a function that assigns to each object d the set of worlds in which d is F.[4] The meaning of F, on this account, carries the same information as on the more familiar view that it is a function that assigns to each world w the extension of F in w. However the former is, but the latter is not, in accord with the Principle of Categorial Semantics mentioned above.

Another example: Consider the sentence

(1) Oliphant does not sing.

A plausible semantic analysis of (1) is this:

---

4. Actually, in Cresswell's theory the meaning of F may be a partial function. It may be undefined at some arguments.

(2) **not** (**sing** **Oliphant**).

In this analysis **not** is viewed as a sentence operator--it combines with a sentence to yield a sentence; **sing** is viewed as a one-place predicate--it combines with a name to yield a sentence; and **Oliphant** is viewed as a name. By the Principle of Categorial Semantics, the meaning of **not** is a function that takes as inputs meanings of sentences, i.e., propositions, and yields as outputs propositions. More specifically, it is the function whose value at each set of worlds W is the complement of W.

The general idea behind the illustrations given above is that of the categorial analysis of language initiated by Leśniewski, and further developed by Adjukiewicz, Carnap, Bar-Hillel, Cresswell, and others. In this type of analysis, expressions are divided into various categories according to their logical and semantic roles. There are two basic categories--the category of sentence, represented by 0; and the category of name, represented by 1. The rest of the categories are constructed out of these according to the rule that if τ and $\sigma_1, \ldots, \sigma_n$ are categories then so is $(\tau/\sigma_1, \ldots, \sigma_n)$. An expression belonging to the category $(\tau/\sigma_1, \ldots, \sigma_n)$ combines with n expressions of respective categories $\sigma_1, \ldots, \sigma_n$ to yield an expression of category τ. Its meaning is a function that when applied to the meanings of expressions of respective categories $\sigma_1, \ldots, \sigma_n$ gives as value a meaning of category τ. Once a domain D_0 of sets of

possible worlds and a domain D_1 of objects is specified, the possible meanings for all the categories are determined. For a category $\sigma = (\tau/\sigma_1,\ldots, \sigma_n)$, the set of meanings of category σ, D_σ, contains functions (possibly partial) from $D_{\sigma_1}\times\ldots\times D_{\sigma_n}$ into D_τ.

Here are some examples of expressions belonging to different syntactic categories and their corresponding meanings:

- A one place functor (e.g., **not**) is of category (0/0), and its meaning is a function which takes as input a set of worlds (a proposition) and yields a set of worlds.

- A two place functor (e.g., **and**) is of category (0/0,0), and its meaning is a function which takes two propositions and yields a proposition.

- A one place predicate is of category (0/1), and its meaning is a function which yields a proposition for each object in D_1.

Examples of some more complicated categories and their meanings:

- The category: (0/1,0). The arguments are: an object, a set of worlds. The value: a set of worlds. Examples: --**believes**--, --**said**--.

- The category: (0/1,1). The arguments are two objects. The value is a set of worlds. Example: --**is the father of**--.

- The category: ((0/1)/(0/1)). The argument is a function from an object to a set of worlds, the value is also a function from an object to a set of worlds. Examples: **slowly**, **quickly** (most adjectives and adverbs belong to this category). **Slowly** combines with a one place predicate such as **walks** to yield **walks slowly**, which belongs to the category (0/1).

- The category: (((0/1)/(0/1))/1). The argument is an object; the value is a function whose argument is a function from an object to a set of worlds, and whose value is again a function from an object to a set of worlds. Examples: **in**, **to** (perhaps most prepositions). The word **in** combines with a name, e.g., **Chicago** to yield an adverbial phrase **in Chicago** which is of category ((0/1)/(0/1)). We saw above that such adverbial phrases combine with one place predicates to yield one place predicates. Thus, **in Chicago** can combine with a one place predicate **lives** to yield a new one place predicate **lives in Chicago**.

In general, this type of categorial analysis exploits higher-order set-theoretic entities to yield a particularly simple and attractive semantic theory of language.

The idea behind Cresswell's solution is that words such as 'believes' do not always express a relation between an individual and the proposition expressed by the complement sentence. Sometimes, on his view, they express a relation between an individual and the meanings of the parts of the complement sentence. For instance, by the possible-worlds theory the proposition expressed by (1) is the same as that expressed by the logically equivalent

(3) Oliphant sings if and only if there is a barber who shaves all and only those barbers who do not shave themselves.

However this does not imply, according to Cresswell, that the belief sentences

(4) Natasha believes that Oliphant does not sing,

(5) Natasha believes that Oliphant sings if and only if there is a barber...,

have the same truth value. The reason is that (4) can be read as asserting a relationship not between the proposition that Oliphant does not sing and the believer Natasha but rather a relationship between the meanings of 'not', 'sings', 'Oliphant', and Natasha. For (4) to be true, according to Cresswell, some quite complex relationship has to hold between these meanings and Natasha. That this relationship holds does not imply that the quite different relationship asserted by (5) -- between a different list of meanings and Natasha -- obtains. This idea is reminiscent

of Russell's multiple relation theory of judgment, but there are important differences between the two.

On Russell's theory of judgment there are numerous relations of belief, not just one. This is a consequence of Russell's unwillingness to countenance 'that p' in 'x believes that p' as a denoting term. Cresswell, on the other hand, does view 'that p' as a denoting term. As a result he can, and does, treat belief as a two-place relation. But the relata of this relation are not always persons and propositions. They are sometimes persons and structured meanings, where the latter are complexed n-tuples built out of meanings. For example, (4) can be analyzed as asserting that the two-place belief relation holds between Natasha and the structured meaning

(6) $\langle \omega_{\underline{\text{not}}}, \omega_{\underline{\text{sing}}}, \omega_{\underline{\text{Oliphant}}} \rangle$

built out of the meanings of 'not', 'sing', and 'Oliphant'. Note that on this theory even though 'that p' is a constituent of 'x believes that p', p may not be a (semantically significant) constituent of 'that p'. In particular, if 'that Oliphant does not sing' is to denote the triple displayed above it must have the analysis

(7) **that** (**not**, **sing**, **Oliphant**).

Here **that** belongs to the category (1/(0/0),(0/1),1). Its meaning is the function $\omega_{\underline{\text{that}}}$ which takes as inputs the meanings of a sentence operator, a one-place predicate, and

a name, and gives as output a member of D_1, i.e., the ordered triple consisting of the three meanings.

One respect in which Cresswell's theory differs from Russell's, then, is that it does not treat predicates such as 'believe' as ambiguous. There is another related difference, that concerns the expression 'that'. Once that-clauses are analyzed on the model of (7) within the framework of categorial semantics, one is forced to conclude that 'that' is ambiguous. For on the most natural analysis, the 'that' in (5) cannot be the same as the one in (7): it takes several more meanings as input and yields a more complex n-tuple as output. Following Cresswell, let us distinguish the various analyses of 'that' by subscripting **that** with the categories that it takes as inputs. (The output is always of category 1.) So the **that** in (7) is really $\textbf{that}_{((0/0),(0/1),1)}$. Some other possible **that**'s are $\textbf{that}_0$ and $\textbf{that}_{((0/0),0)}$. The meaning of $\textbf{that}_0$ is just the identity function: its value for any argument is the argument itself. (But the argument and the value belong to different categories. The argument belongs to 0, the value to 1.) The meaning of $\textbf{that}_{((0/0),0)}$ takes meanings from $D_{(0/0)}$ and D_0 and yields ordered pairs of them.[5] I will not

5. The semantic interpretation of a 'that' clause may depend on the propositional attitude verb it follows. For instance: a that clause, which is the complement of a verb like 'says' or 'is given', picks out a more finely individuated object than a complement of verbs like 'knows', 'believes'.

attempt to spell out the meaning of 'that' on the most natural analysis of (5).[6]

There is another way of looking at Cresswell's theory: we distinguish between the sense and reference of a sentence and take the reference of a sentence to be the set of possible worlds in which the sentence is true. That is, we identify the reference with what was called above the meaning of the sentence, and we take the sense of a sentence such as (1) to be the structured meaning (6).[7] Propositional attitudes, on this reading of Cresswell's theory, express a relation between a person and the sense of the complement sentence, not its reference.[8] Viewed this way, Cresswell's approach appears as a combination of the Carnap-Lewis account of sense in terms of intensional isomorphism[9] and the Fregean idea that propositional

---

6. Cresswell's theory of that-clauses puts strong closure conditions on the domain D_1. These conditions entail that copies of the domains D_σ, for all categories σ, must be present in D_1. If $\sigma = (\tau/\sigma_1,\ldots,\sigma_n)$, it follows that D_σ cannot consist of all the functions from $D_{\sigma_1}\times\ldots\times D_{\sigma_n}$ into D_τ. This raises two concerns. First, the sentences that require higher-order quantification might not receive the right interpretation. Second, some intuitively meaningful λ-terms might be ruled meaningless.

7. This view of the sense and reference of a sentence is not strictly Fregean. Even the idea that the sense of a sentence is a structure built out of the references of the parts is entirely foreign to Frege.

8. Ignoring for the moment the point made in the paragraph after next.

9. See Carnap(1947) secs.13-15 and Lewis (1970) secs.V.

attitude contexts are oblique: the reference of a sentence in such contexts is its usual sense.[10]

Cresswell's theory, then, agrees with Frege's--and disagrees with Russell's--in treating the that-clause as a genuine semantic constituent of the propositional attitude sentence. When sense is understood in the Carnap-Lewis way, Cresswell's theory can even be read as saying that the denotation of the that-clause is the sense of the complement sentence. However, Cresswell agrees with Russell, and disagrees with Frege, in not viewing the complement sentence as a genuine constituent of the propositional attitude sentence.

There is one respect in which Cresswell disagrees with both Frege and Russell. It is a consequence of Cresswell's theory that all propositional attitude sentences are structurally ambiguous. Consider (4). Its that-clause is open to at least these analyses:

(8) $\underline{\text{that}}_{((0/0),(0/1),1)}$ (**not**, **sing**, **Oliphant**)

(9) $\underline{\text{that}}_{((0/0),0)}$ (**not**, (**sing** **Oliphant**))

(10) $\underline{\text{that}}_{0}$ (**not** (**sing** **Oliphant**)).

---

10. Cresswell himself makes use of the terminology of sense and reference in the informal presentation of his ideas. His formal semantic theory, however, does not use the distinction. The semantic rules work on one level only, the level of reference.

These analyses generate different interpretations of (4). If the that-clause is given the analysis (10) then (4) expresses a relation between Natasha and the proposition that Oliphant does not sing. (This analysis makes (1) a genuine semantic constituent of (4).) If the that-clause is given the analysis (9), then (4) expresses a relation between Natasha and the structured meaning

(11) $\langle \omega_{\underline{\text{not}}}, \omega_{\underline{\text{sing}}}(\omega_{\underline{\text{Oliphant}}}) \rangle$.

Note that the second element of this ordered pair is the proposition that Oliphant sings. Note also that under this analysis (1) is not, but 'Oliphant sings' is, a genuine constituent of (4). That all propositional attitude sentences are structurally ambiguous in the way just illustrated is a new and intriguing consequence of Cresswell's theory.

SECTION II

This section of the chapter examines Cresswell's applications of his ideas in dealing with three problems that confront his core proposal: the problems of structural ambiguity (in 2.1), the problem of names (in 2.2), and the problem of iterated attitudes (in 2.3).

2.1 The Problem of Structural Ambiguity

We have seen that it is a distinctive consequence of Cresswell's theory that propositional attitude sentences are structurally ambiguous. The claim of ambiguity as presented in Part 1, however, is very strong and has some undesirable consequences. It implies, for instance, that there are readings of (4) and (5) on which they say the same thing. (In these readings 'that' is taken to mean $\omega_{\underline{\text{that}}_0}$.) To avoid such consequences Cresswell introduces what he calls the Macrostructure Constraint:

> ...in nontechnical discourse most propositional attitude sentences are to be construed in such a way as to make the attitude verb sensitive to...the macrostructure of the complement sentences. I am deliberately leaving this a little vague. In fact, it has to be, because one of my principal claims is that there is a large amount of ambiguity in propositional attitude sentences and so the theory cannot tie down at all precisely what the limits are (p. 80).[11]

The macrostructure constraint is aimed at establishing a correlation between the syntactic complexity of the complement sentence and the desirable complexity of the content of the attitude. This correlation is not made precise. The examples Cresswell brings to illustrate the point involve propositional attitude verbs like "is given", "deduce", where the sensitivity required for the analysis of

---

11. The logic behind the 'so' in the last sentence is not very clear to me.

the embedded sentence is of the very fine level of intensional isomorphism. For example, the sentence

> Mary is given that someone either steals spoons and forks or steals spoons but does not steal forks.

does not have the same meaning as the sentence

> Mary is given that someone steals spoons.

But having to take, as a rule, practically every word in the complement sentence as a genuine constituent of its semantic structure may be more than Cresswell is willing to bite on, due to the obvious consequence that even the simlpest logical transformation does not hold. At another point Cresswell says concerning the Macrostructure constraint that the meaning of a propositional sentence "almost always takes account of" those parts that are "very easily identifiable subsentences or subclauses" (p.80). Applied to the example above, the constraint ensures that "whereas 'steals the spoons' is taken as a basic unit, 'steals the spoons and forks' is not" (p.80). No principle for easy identification of subsentences is provided, leaving us to wonder whether the constraint is aimed at blocking some undesirable substitutions of logically equivalent atomic propositions, or of cointensional predicates or of coreferrential names.

In summary, Cresswell does not spell out the notion of macrostructure, but the idea underlying the Constraint is that the use of a more complex sentence in the complement

generally indicates a more complex content.[12] Cresswell's Macrostructure Constraint weakens the claim of structural ambiguity and makes it somewhat vague, but does not cancel it altogether. The question arises whether this kind of ambiguity is actually found in natural languages. Cresswell gives, in Chapter 4 of his book, the following example, that is meant to show that it is.

Cresswell asks us to consider the following situation: he is reading a map in which sectional distances are given along the route that he and his wife are travelling. Say that they are travelling from a point A to a point C and the route takes them through B. Suppose also that according to the map the distance between A and B is 5 kilometers and the distance between B and C is 7 kilometers. Now Cresswell notes, in our opinion correctly, that both of the following responses to the query "how far is it from A to C?" would be appropriate and true.

(12) Well, the map indicates that it is 5 + 7 kilometers from A to C.

(13) Well, the map indicates that it is 7 + 5 kilometers from A to C.

---

12. It is unclear whether the Constraint is meant to work at the level of semantics or of pragmatics. In the latter case problems similar to the one about the equivalence of (4) and (5) could perhaps arise for the theory. For the constraint, it would appear, could be overridden by other pragmatic considerations allowing the 'that' in (5) to be read as $\underline{\text{that}}_0$.

Cresswell continues, "Contrast this with a situation in which I am putting the map to a different kind of use. My wife wants to know whether the long stretch or the short stretch comes first, so she asks me to tell her what the map indicates about distances in the order in which we are driving (pp. 34-35)." In this situation, Cresswell claims, (12) would be a correct response but not (13). Cresswell concludes that (13) is ambiguous: in one reading, the one deployed in the first situation, the relative order of 7 and 5 is not important; in the other reading, the one deployed in the second situation, the relative order is important. Cresswell solidifies his argument noting, "The interpretation of (13) in which the relative order of 7 and 5 is important can be clinched by adding 'in that order':

(14) Well, the map indicates that it is 7 + 5 kilometers from A to C, in that order."[13]

One obvious problem with this example is that it is not clear that the behavior of (13) in the two situations is to be explained by positing a semantic ambiguity in (13). It may be that (13) is used elliptically in the second situation. Or it may be that the different uses of (13) can be given some pragmatic explanation. However, this is not the main reason for our dissatisfaction with Cresswell's example. It seems to us that there is a problem here even if it is granted that (13) is semantically ambiguous. The

---

13. Page 35. The numbers within parentheses have been changed to fit with the numbering used in this chapter.

problem is that the example does not show that (13) exemplifies the specific kind of ambiguity predicted by Cresswell's theory. Cresswell needs to show that even when the complement sentence is unambiguous, the propositional attitude sentence can be ambiguous. But Cresswell's example illustrates at best only the familiar fact that a propositional attitude sentence is ambiguous when the complement is. For in Cresswell's example if we say that (13) is semantically ambiguous then we should say equally that the complement

(15) It is 7 + 5 kilometers from A to C

is semantically ambiguous also. It can be put to the same kinds of uses as (13) is in the situations described by Cresswell.[14] This suggests that if there is semantic ambiguity in (13), its source is the ambiguity of the complement, not the sensitivity of 'indicates' to different levels of analyses of the complement. Another way of putting the point is this. The two uses of (13) relate the map to different propositions (i.e., different sets of possible worlds). What is wanted, however, is a case of ambiguity in which the contents of the attitudes are different even though the propositions expressed by the complements are the same.

---

14. This is evidence for not positing a semantic ambiguity to explain Cresswell's phenomenon.

2.2 The Problem of Names

The structured meaning approach to the problem of propositional attitudes has resources to disallow substitutivity *salva significatione* of strongly equivalent sentences: it can rule as inadmissible even the commutativity of conjunction, the transformation from active to passive, and other such transformations that preserve many semantic properties. As long as sentences are structurally different or, if structurally similar, they are built out of intensionally nonequivalent components--that is, as long as sentences are not intensionally isomorphic--the structured meaning approach can assign to them different senses, and bar their intersubstitutability. The individuation of the objects of attitudes on the structured meaning approach is, thus, very fine.

However, it may not be fine enough. The approach implies that intensionally isomorphic sentences have the same senses. But such sentences are sometimes not intersubstitutable within propositional attitude contexts. It appears, for example, that the following two sentences can differ in their truth value:

(16) Quentin says that Hesperus is Phosphorus,

(17) Quentin says that Hesperus is Hesperus.

The complements of (16) and (17), however, are intensionally isomorphic: they are built out of intensionally equivalent components in the same way. Intuitively it seems that the objects of (16) and (17) should be distinct. But according to the structured meaning view they are the same.

Cresswell attempts to solve this problem by appealing to the description theory of names. He thinks that in (16) and (17) the names 'Hesperus' and 'Phosphorus' can be replaced by distinct definite descriptions. Cresswell does not make a specific proposal about what descriptions these names should be replaced with. One possibility that he mentions, and looks favorably upon, is that the name 'Hesperus' can sometimes mean 'the planet called "Hesperus"' (and similarly for 'Phosphorus'). Cresswell is aware of the criticisms that the description theory has received in the recent years. Though he does not attempt to show this, he thinks that these criticisms can be met.

It seems that the following example, modeled on a standard criticism of the description theory, creates a problem for Cresswell's theory. Intuitively, the following can be observed about these two sentences:

(18) Even though it is a necessary truth, Quentin does not know that Hesperus is Phosphorus,

(19) Even though it is a necessary truth, Quentin does not know that Hesperus is Hesperus,

To begin with, these two sentences relate Quentin to different objects of ignorance; the 'it' in these sentences refers to what Quentin is said to be ignorant about; and, finally, (18) could be true. Let us now consider the sentence obtained by replacing the names 'Hesperus' and 'Phosphorus' in (18) by the corresponding descriptions:

(20) Even though it is a necessary truth, Quentin does not know that the planet called 'Hesperus' is the planet called 'Phosphorus'.[15]

There seems to be no reading of (20) which has the same properties as (18). If we give the descriptions narrow scope[16] then (20) must be false irrespective of Quentin's ignorance because it is a contingent truth that the planet called 'Hesperus' is the planet called 'Phosphorus'. On the other hand, if we give the descriptions wide scope[17] then what Quentin is said to be ignorant about turns out to be the same as that in (19).

More fundamentally, the appeal to the description theory of names does not solve the main problem here but

---

15. The descriptions used to replace the names are immaterial to the argument so long as (logically) distinct descriptions replace distinct names and they pick out Venus contingently.

16. In which case the first conjunct of (20) says that it is necessary that the planet called 'Hesperus' is identical to the planet called 'Phosphorus'.

17. In which case the first conjunct of (20) says that there is a unique planet called 'Hesperus' and there is a unique planet called 'Phosphorus', and it is necessary that they are identical.

only shifts it to a different place. The main problem is that on the structured meaning approach, even though the substitution of intensionally equivalent sentences does not preserve meaning, the substitution of intensionally equivalent (atomic) terms does. So the problem stated above arises not only for names but for terms of all categories. Even if we succeed in barring the intersubstitutability of coreferential names, the intersubstitutability of intensionally equivalent (atomic) predicates (for example) remains and causes similar problems.

2.3 The Problem of Iterated Attitudes

It was first observed, as far as I know, by Cresswell himself that iterated attitudes raise a problem for theories that are based on Carnap's idea of intensional isomorphism. The problem arises as follows. Consider a sentence such as

(21) Mortimer believes that Natasha believes that Oliphant does not sing.

It would appear that one possible semantic analysis of this sentence is

(22) **believes** (**Mortimer**, **that**$_{((0/1,1),1,1)}$ (**believes**, **Natasha**, **that**$_0$ (**Oliphant does not sing**))).

Here **believes** is treated as belonging to the category (0/1,1). Its meaning, say ω, is a function (possibly partial) from $D_1 \times D_1$ into D_0. The meaning of (22), then, is

$\omega(m, x)$, where m is the meaning of **Mortimer** and x is the meaning of the that-clause

(23) **that**$_{((0/1,1),1,1)}$ (**believes**, **Natasha**, **that**$_0$ (**Oliphant does not sing**))).

Letting n be the meaning of **Natasha** and W be the proposition that Oliphant does not sing, the semantics of **that**$_{((0/1,1),1,1)}$ yields that the meaning of (23), x, is

$\langle\omega, n, W\rangle$.

And so the meaning of (22) is

$\omega(m, \langle\omega, n, W\rangle)$.

Now suppose, for reductio, that this is well-defined and yields the value z. Since

$\omega(m, \langle\omega, n, W\rangle) = z$ iff $\langle\langle m, \langle\omega, n, W\rangle\rangle, z\rangle \in \omega$,

we conclude,

$\langle\langle m, \langle\omega, n, W\rangle\rangle, z\rangle \in \omega$.

This implies that there is a sequence of sets $Y_1, \ldots, Y_n$ such that

$\omega \in Y_1 \in Y_2 \ldots Y_n \in \omega$.

More particularly, we have,

$\omega \in \{\omega\}$

$\in \langle\omega, n\rangle$

$\in \{\langle\omega, n\rangle\}$

$\in \langle\omega, n, W\rangle$

$\in \{m, \langle\omega, n, W\rangle\}$

$\in \langle m, \langle\omega, n, W\rangle\rangle$

$\in \{\langle m, \langle\omega, n, W\rangle\rangle\}$

$\in \langle\langle m, \langle\omega, n, W\rangle\rangle, z\rangle$

$\in \omega$.

This violates the Axiom of Foundation. Consequently, on Cresswell's theory, (22) is ruled meaningless. Clearly the same problem arises for any analysis of (21) in which **believes** is an independent constituent of the main that-clause. Whether the problem arises for analyses that satisfy the Macrostructure Constraint is unclear due to the unclarity and imprecision surrounding the Constraint.

Cresswell's response to the problem is to accept the apparently undesirable consequence of his theory and to argue that the consequence is not as bad as it appears to be. He points out in the first place that the problem does not arise for various alternative analyses of (21). Here are two such:

(24) **believes** (**Mortimer**, **that**$_0$ (**believes** (**Natasha**, **that**$_0$ (**Oliphant does not sing**)))).

(25) **believes** (**Mortimer** **that**$_{((0/1),1)}$ (**believes** **that**$_0$ **Oliphant does not sing**, **Natasha**)).

Note that in the second of these analyses the expression

(26) **believes that$_0$ Oliphant does not sing**

is viewed as a one-place predicate.[18] So the first occurrence of **that** in (25) is not sensitive to the internal structure of (26), and this prevents the problem discussed above from arising. (Similar considerations hold for (24).) Cresswell suggests that analyses such as (24) and (25) may capture well enough the meanings of iterated attitude sentences like (21).

Another option that is available in interpreting (21) is to introduce a hierarchy of belief predicates and to view the two occurrences of 'believes' in (21) as belonging to different levels. This would allow (21) to have the analysis (22) and it would get around the set-theoretical problem; but it introduces several others that are even more intractable. (See Gupta (1982) for a discussion of some of these in the context of a theory of truth.) Cresswell's own preference is to argue that there is something intuitively wrong with analyses such as (22). His reason: "...it seems, if a sentence like [(22)] is to be admitted, that we must already know what **believe** means in order to say what it means. It seems to me better just to say that [(22)] is semantically anomalous and expresses no proposition...(p. 90)."

---

18. This is possible in virtue of λ-abstaction.

In my opinion the problem of iterated attitudes is a purely technical problem, one that deserves a purely technical solution. Whether there is anything intuitively wrong with sentences like (22) and, in particular, whether they creat a circularity in the meaning of **believe** can only be determined once we have some account of the conditions under which persons bear the belief relation to structured meanings. These substantive matters cannot be decided on the basis of formal considerations alone. An example may make the point clearer.

Consider the meaning of the **is** of identity in Cresswell's scheme. A problem parallel to the one for **believe** arises for it. Let us take **is** to be of category (0/1,1). Then its meaning, ι, is a function from $D_1 \times D_1$ into D_0. Now consider the that-clause

(27) **that**$_{((0/1,1),1,1)}$ (**is**, **Natasha**, **Natasha**).

Its meaning is

(28) $\langle \iota, n, n \rangle$,

and it belongs to D_1. Foundation dictates, as before, that the function ι must be undefined whenever (28) is one of its arguments. Hence, contrary to expectation, ι is not defined over all of $D_1 \times D_1$. But there is no intuitive problem about the meaning of **is** as it applies to (28).[19] The problem is

19. Similarly, there is no intuitive problem in the meaning of **say** being well defined at structures that contain

purely technical, being due to the set theory used. It can be solved by moving to a set theory in which the axiom of foundation is denied.[20] There may be special considerations that make (22) an intuitively unacceptable analysis of (21). The problem of iterated attitudes described above does not show that it is.

---

itself as a constituent, especially if say is given the analysis favored by Cresswell in Chapter 13 of his book.

20. See Aczel 1988 and Barwise and Etchemendy 1987.

Chapter 6

The Sentential Theories of Belief

It was argued in the last three chapters that the intensional opacity of belief sentences cannot be accounted for by theories that take objects of belief to be sets of possible worlds or "structured meanings". These theories are sometimes said to be following the "indirect approach" to objects of belief, since they regard these objects as determined indirectly from the complement sentence of the belief attribution.

In this chapter I discuss theories that follow the "direct" approach. These theories take the objects of belief to be furnished by the complement sentence itself, thus they postulate an even finer individuation of the objects of belief, in terms of the syntactic structure of the attributing clause. A major motivation for these theories, aside from the logical problem of opacity, is ontological: in general, these theories do not appeal to intentional entities like propositions in their explanation of belief and other attitudes. Several types of sentential theories are found in the literature. They differ in the linguistic entity they take the object of belief to be (e.g., sentence type, sentence token, eternal sentence, utterance, sentence in a natural or in a mental language),

and sometimes in the relativity (e.g.,to language, meaning, context) that they ascribe to belief.

This chapter is divided into four sections. The first one displays **"naive"** quotational theories, that take the object of belief to simply be the sentence used to ascribe that belief. The second section discusses Scheffler's **"sentence-ascription"** theory. The third section is devoted to Davidson's **"paratactic"** theory of indirect discourse and its possible extension to the treatment of belief sentences. The fourth section deals with the **"mental-sentence"** theory, whose main advocate is Fodor.

I hope to demonstrate in this chapter that although the need for a finer individuation of the objects of belief is legitimate, linguistic entities, however they are constructed, will not do. The resulting identity conditions for these objects are bound to cut too thin, and to create additional difficulties (which will be discussed below).

SECTION I

In its simplest form, a sentential theory takes the complement sentence p in the belief attribution:

(1) a believes that p

as mentioned, not used. 'That p' is the **quotational name** of a sentence type; i.e.: 'that ...' denotes the sequence of words that follow the word 'that'. Thus, (1) is analyzed, on this theory, as containing 'believes', which is viewed as a two place predicate; a term a which denotes an agent; and another term 'that p' which denotes a sentence in English.

This theory posits strong identity conditions on the objects of belief, and thereby gains several important advantages. To begin with, it accounts for the opacity of belief: if the denotation of the term 'that p' is simply the sentence p, then substituting a different sentence in the complement, even one logically equivalent to p, results in a term that denotes the new sentence, not the original one. Such a substitution alters the argument provided for the value of 'believes', hence the value of the whole belief attribution may change. Thus on this theory the inference

David believes that John is a spy
John is David's father

David believes that his father is a spy

is parallel to the inference:

'John is a spy' is written on the blackboard
John is David's father

'David's father is a spy' is written on the blackboard.

The invalidity of the first argument is as blatant as that of the second. Another advantage is that the theory yields an account of quantified sentences. Unlike Quine's theory, it allows us to infer from (1) that there is something that a believes.

The most obvious problem with the above naive theory is that the object of belief, being a sentence type, may be context dependent (e.g., when it contain indexicals, tenses etc.), or ambiguous (e.g., as in 'visiting relatives may be boring'). In either case the sentence type denoted by 'that p' fails to pin down the content of the belief attributed. Other difficulties for the theory stem from the language dependency of sentences:[1] the same sentence might exist, by coincidence, as a linguistic form in two languages, carrying a different meaning in each. An example (with sentences understood in phonetic terms): 'Empedokles liebt' means, in German, 'Empedokles loved', while in English it means 'Empedokles jumped.'[2] Another problem concerns the extension of 'believes': by the theory, the extension of

---

1. Some critics understood the language dependency of sentences to mean that we cannot attribute to Galileo the belief that the earth moves, unless we say the sentence that follows 'that' in Italian. But the existence of a believer who does not speak English is an objection to the theory only if beliefs are identified with disposition to assent to that same sentence. A quotational theorist may well hold that even if we take assent to S to be evidence for believing S, still it is neither sufficient nor necessary for a correct attribution of the belief that S.

2. This example was introduced by Davidson, 1968.

`believes' is a set of ordered pairs of believers and sentences in English. Symmetry dictates that the extension of the German word 'glaubt' would contain ordered pairs of believers and sentences in German. But this violates a plausible principle of translation of words by which translation preserves extension.

There are two major strategies that are taken by the sentential theorists in response to the objections to the "naive" theory:

A. **Enrich** the relation of belief into a three or more place predicate. There are essentially three options available for such "enrichment": maintain that 'believes' holds for an agent, a sentence and (i) a language; (ii) a context; (iii) an interpretation.

B. **Redefine** the objects of belief, while preserving their linguistic characteristic. E.g., these objects may be identified with sentence-tokens, eternal sentences, sentence-tokens grouped by intensional isomorphism, sentence tokens grouped by inferential role, utterances or sentences in an internal language.

Some of the major theories that follow strategy B will be considered in later sections. Of the theories that follow strategy A, the first option mentioned above takes 'believes' as a three place predicate that holds for believers, sentences and **languages**. Some difficulties --

those related to language dependency -- are eliminated on this proposal, but problems of context dependency and ambiguity remain. A new objection arises, pertaining to the individuation of languages, which may be considered as problematic as the individuation of intensional entities like propositions. The second option argues for relativity to **context** in the third argument place.[3] Context fixes the content of indexical sentences, but may fail to do so when an ambiguous sentence occurs in the complement. This option also confronts objections relating to inexpressible beliefs, to language dependency, and to the analysis of belief attributions which employ quantifiers (e.g., '$\underline{a}$ believes something that $\underline{b}$ does not believe').

Another attempt to answer some of these objections along the first strategy is to render 'believes' a three-place predicate, whose third argument aims at capturing the **meaning** of the complement sentence. So, a sentence like (1) has the logical form of B($\underline{x}$,$\underline{s}$,$\underline{m}$), where $\underline{x}$ is the believer, $\underline{s}$ is the sentence-type that follows, and $\underline{m}$ is the interpretation of s. The sentence B($\underline{x}$,$\underline{s}$,$\underline{m}$) is true iff $\underline{x}$ believes the sentence $\underline{s}$ when it has the meaning (interpretation) $\underline{m}$.

---

3. Such an analysis may be considered equivalent to taking the objects of belief as eternal sentences (i.e., sentences which are equipped whenever needed with explicit mention of time, place, persons etc., to remove the effect of indexicals).

At first glance, this proposal seems to be a promising start: when the appropriate method for determining the interpretation of a sentence are spelled out, relativity to context and language dependency may not be necessary. A second glance reveals, however, that by smuggling the notion of interpretation, this proposal begs the question of fixing the object of belief. Furthermore, as Cresswell observed, it creates the suspicion of redundancy: if the notion of "interpretation" is to do the job assigned to it, it is also powerful enough to provide, on its own, entities which can be the objects of belief. In other words, if the interpretation m of the sentence s is sufficiently finely discriminating to pick out just those cases in which s has the interpretation by virtue of which x believes s, then m alone is sufficient as the object of x's belief, rendering the sentential component in the schema irrelevant to the analysis of the belief sentence. The role played by "quotation" in this approach is at best a minor one. From the sententialist's point of view, quotational theories which involve considerations of interpretation of the embedded sentence cease to be genuinely quotational, thereby lose the benefits resulting from ontological economy and from the syntactic criterion for individuating beliefs.

SECTION II

One of the most prominent theories along strategy B -- the Inscriptional Theory -- is advocated by Scheffler.[4] Scheffler offers what he calls "a nominalistic analysis" of indirect discourse -- an analysis which treats them as structurally analogous to direct discourse. Scheffler actually proposes a non-standard but interesting account of direct discourse. He views inscriptions framed by quotes as concrete general terms (not abstract singular ones). If each of two people writes 'pigs fly', then, on Scheffler's theory, the inscriptions produced by the two fall under the general term ''pigs fly''; as also do all inscriptions that are **replicas** of them. (''Pigs fly'' does not denote an abstract entity.) The conditions any two sentence-inscriptions have to meet in order to be considered "replicas" of each other are:

> ...two sentence-inscriptions represent the same sentence if and only if they are replicas of each other (i.e., are spelled exactly alike), have similar language affiliation (i.e., both are French, both Italian etc.), and lack indicator terms (i.e., term-inscriptions which are replicas, though one appear in one of the sentence-inscriptions with one denotation, and another appears in another sentence-inscription with different denotation).[5]

---

4. 1954 and 1963.

5. 1963, p.103.

The reference of 'That (-)' includes all and only those sentence-inscriptions which are **rephrasals** (i.e., replicas) of the sentence-inscription occurring between its parenthesis. In general, 'That (-)' is a predicate forming operator, such that insertion of a sentence-inscription between its parenthesis yields a predicate inscription applying to rephrasals of the insert.

The function of 'that p' in indirect discourse contexts is, by this account, exactly like the function of quotation in direct discourse: it is a one-place predicate of concrete inscriptions that represent p in the sense explained above. Thus, the sentence

(2) John writes that pigs fly

is analyzed as:

(∃y)(John writes y & y is an inscription of 'pigs fly')

Scheffler does not analyze the relation 'writes', but what does not follow from the theory is that it is necessary for John to produce the inscription above in order for that relation to hold. Scheffler's analysis concerns the logical form and the ontological character of indirect quotation contexts, and not the substantive analysis of the conditions under which such sentences are true. Concerning belief attributions, Scheffler states explicitly that "...when the 'belief true' relation holds, it need not be expected that the agent produce, be aware of, or even understand the

inscription believed true."[6] The construal of a belief attribution such as (1) is presumed true just under those conditions in which ordinary 'believes that' statements are considered true, no matter what these conditions may be.

A belief attribution is treated on par with indirect quotations: the logical form of (1) involves a term 'a', a two place relation 'believes', and a predicate forming operator 'that p', which is true of concrete sentence inscriptions or utterances that represent p.

Scheffler points out the following advantages for positing concrete sentence-inscriptions (or utterances) as the objects of belief: being concrete renders the objects of belief clearer elements for any theory than abstract entities such as possible states, both because their identity conditions are simpler, and because the existence of inscriptions and utterances is not jeopardized by their own falsity or even self-contradiction (in contrast to that of possible states). Since inscriptions must be acknowledged to exist, there is no ontological extravagance in using them in the analysis of belief, and, furthermore, they are likely to already belong to any ontology associated with inquiry into human conduct.

There are two major motivations behind the inscriptional theory. The standard one, shared with most

---

6. 1963, p.104.

other sententialists, is to find objects of belief that are both economical and noncommittal ontologically, and also individuated uncoarsly. The inscriptional theory gains the effect of states individuated by the clear criterion of sentence identity. The other motivation behind Scheffler's proposal is to provide a teleological explanation for belief, desire and performance, following Ducasse's model of goal idea. Ducasse postulated that if an agent believes that y is contingent upon x, and if he desires that y, then he is likely to do x. This schema can be manipulated to make the relation of "contingent upon" logically precise, and to replace the "likely to" by a generalization, roughly presented as

> $(\forall x)(\forall y)(\forall z)\{[(x \text{ desires that } y) \ \& \ (x \text{ believes that if } z \text{ then } y)] \text{ then } (x \text{ performs } z)\}$[7]

The wish to incorporate the analysis of belief sentences into teleological generalizations forces, on this view, the treatment of "Propositional Attitudes" as relations to some clearly individuated objects, and rules out a theory (like the fusion theory) by which we cannot dissect 'p' out of 'a believes\desires that p'.

Quine, in response to the heavy artillery aimed at him by Scheffler, asks how would the inscriptional theory cope with belief ascriptions like in 'Paul believes something

---

7. The variable x ranges over agents, z ranges over acts, and y ranges over the objects of belief and desire.

that Elmer does not', where that which 'something' denotes may not be a concrete inscription or utterance. Scheffler recognizes the defect, but finds it insignificant. In his opinion, this shortcoming does not touch the explicit idiom 'a believes that p', and is limited to quantifications which are "expendable without serious loss". Conveniently, Scheffler partly agree with Quine: we can give up some quantification over belief objects (as in the above example and in constructions like 'Eisenhower and Stevenson agree on something'), but not all such quantifications (which may lead to total dismissal of objects of belief), since they are needed for teleological arguments. One may still wonder how could the incorporation of the inscriptional approach in the psychological model explain the fact that Paul and Elmer perform the same acts, given that they have the same beliefs and desires, whatever these may be.

Is Scheffler's analysis of belief sentences, even when restricted to the "explicit idiom", successful in pinning down the objects of belief via sentence-inscriptions? I believe not, first because the identity conditions the theory imposes on these objects are too strong. Common usage (and sense) seems to dictate that, for instance, the belief that Mary loves Jim be identical to the belief that Jim is loved by Mary; that the belief that Sarah and Tom are married be identical to the belief that Tom and Sarah are married; that believing that all cats are black is the same as believing that every cat is black. Scheffler's approach

is forced to assign a distinct object of belief to each one of these sentences. Even if we don't talk in terms of meanings or interchangeability salva veritate, it is obvious that the ascription of the second belief in each pair above, on top of the ascription of the first, adds absolutely no information. It is also hard to conceive of normal circumstances in which a person's belief can be correctly ascribed to him by the sentence 'all cats are black', but cannot be equally attributed by the sentence 'every cat is black'. In that sense, the very strong identity conditions on the objects of belief are doing too much - we can not identify beliefs that are identical by intuitive account.

For another class of belief sentences, the theory is doing too little: if the complement of the belief attribution contains indexical terms, then the theory fails to give an account for it. A sentence such as 'John believes that he is smart' can not be analyzed as asserting some relations between John and all inscriptions 'he is smart', because in different inscriptions the term 'he' may well denote different people. This objection can perhaps be overcome by some device for "eternalizing" the complement sentence, but the inscriptional approach, which relies on concrete physical objects of belief, is barred from using any device that involves abstract entities like eternal sentences.

The scope of sentences treatable by Scheffler's theory is extremely limited: indexical and quantified sentences as those mentioned above are outside that scope, as well as implicit and inexpressible beliefs. Worse than that -- the analysis fails to account for unexpressed beliefs as well. A theory of belief attribution can hardly be a significant part of a teleological explanation if it leaves sentences expressing these beliefs unaccounted for. Scheffler's motivation behind rendering belief a relation to inscriptions is "...to gain the effect of states individuated by sentence identity."[8] This effect has not been gained for belief contexts (as mentioned, intuitively identical states are not identified by the theory). The dismissal of the "substantive analysis" of belief should not go unpunished in an analysis that in effect puts the burden of the essential problem on the relation of believing. It is not at all clear how can we understand the notion of belief, when limited to concrete expressed inscriptions, without involving the notion of content.

SECTION III

Davidson's theory of "saying that" and other Propositional Attitudes is an improvement on quotational and inscriptional theories considered above. Quotational theories view the

---

8. 1963, p.103.

words that follow 'said that' as operating within concealed quotation marks. Sealed within quotations, the complement sentence's sole function is to refer to a sentence by mentioning (not using) it. One motivation for such theories was to overcome the problem of being forced to endorse unwarranted logical relations, which seems to arise if we attribute semantic structure to the complement sentence. By assigning no semantic structure to the problematic component of the belief sentence, quotational theories pay a price on a few scores: since there are infinitely many different quotational primitive names, no language which contains them can have, as Davidson points out, a recursively defined truth predicate. All such languages are, as a result, unlearnable. Moreover - by giving the following correct report 'Pierre said that Paris is a crazy city' one does not say (directly) anything remotely like what Pierre (who does not speak English) is claimed to have said. Davidson feels that an indirect report should not have such language dependency for its correctness, nor rely on problematic definitions of synonymity, translation or meaning.

In order to remedy these flaws, Davidson proposes to analyze the sentence

(3) Galileo said that the earth moves

as containing a singular term 'Galileo', a two place predicate 'said' which is true of a speaker (denoted by the previous singular term) and an utterance, to which the

demonstrative 'that' refers. Davidson's view is similar to that of Scheffler's in that the object of saying is a particular utterance (or inscription). In this respect both inscriptional and paratactic theories differ from quotational theories, in which that object is a sentence type. Davidson's account of of the complement sentence of (3) is different from both quotational and inscriptional theories in two respects: (i) the complement sentence in (3) is used, not mentioned; (ii) the complement sentence of (3) is not a genuine logical component of the whole sentence. Davidson views (3) as consisting of two sentences, paratactically joined:

(4) Galileo said that.

(5) The earth moves.

This is further explicated in this way:

(6) Galileo made an utterance which, with my next utterance, puts him and me in the samesaying relation. The earth moves.

Sentence (3) is true in the same circumstances as sentence (6). Clause (5) gives the content of the speaker's saying, but has no logical or semantic connection with the original attribution of saying (i.e., with clause (4)), except for being an utterance referred to by 'that' in (4). Note that (5), as the relata of 'that' in (4), is neither a sentence nor a proposition -- it is a concrete utterance. One

advantage gained by this move is that the problem of context dependency, which effects the former theories discussed, is overcome, simply because the complement sentence, on Davidson's theory, is used, not mentioned. Further, there is no commitment to intensional entities beyond the desirable (for Davidson) boundaries of extensional logic.

From the logical point of view, the complement sentence in indirect discourse is not contained, according to Davidson's analysis, in the reporting sentence. It is not a semantic component of the utterance, therefore there are no inferential implications. (3) is true just in case Galileo stands in the samesaying relation to the referent of 'that', and he stands in this relation if he makes an utterance that matches in content the attributor's utterance of (5). This analysis accounts for the usual failure of substitutivity in attribution of the "saying" attitude, without invoking a nonextensional semantics, since the reference of 'that' changes with any change in the sentence which follows it.

Davidson's analysis of indirect discourse is made relevant to our subject matter by his ambitious remark, to the effect that his analysis "opens a lead to an analysis of psychological sentences generally (sentences about propositional attitudes, so called)...".[9] His analysis hinges on two major notions, both of which should be (in my opinion) crucially revised in order to even attempt the

---

9. 1968, p.143.

extension of his treatment to belief sentences. The relation of "saying" holds between agents and concrete utterances, but not so the relation of "believing"; the relation of samesaying holds between the one who utters a sentence and the one who reports that utterance, but it does not seem to hold between the one who has a certain belief and the one who attributes that belief to him.

Let us start with the first problem: the complement sentence which is denoted by 'that' has, on Davidson's account, the ontological status of a concrete utterance. Obviously, the analysis of belief sentences cannot be restricted to reported beliefs. From (3) we can infer that there is something (a concrete utterance) that Galileo produced, but from

(7) Galileo believed that the earth moves

we cannot. How can we extended Davidson's analysis to unreported beliefs?

One alternative that suggests itself is to take 'that' in (7) to refer to kinds of utterances. In that case, sentence (7) receives the following Davidsonian analysis:

(5) The earth moves.

(8) Galileo believes that kind of utterance.

But, as Schiffer[10] points out, this interpretation renders 'that' a nonprimitive demonstrative: utterance-kinds can be individuated only by the content of the utterances they subsume. That implies that we do not know the reference of 'that' in (8) until we know what notion of content is intended by the classification of "utterances kinds". Furthermore, since quantification over utterance-kinds, which are universals, does not fit with Davidson's extensional program, he is likely to object to this way of extending his "saying" account to belief.

Another option may be to take 'that' in (7) to refer to potential or possible utterances. But, this alternative is bound to be rejected by Davidson, on the same grounds that the previous one is: by admitting "possible utterances" into the theory, we are crossing the boundaries of extensional logic. I can think of no other possible revision within the desiderata Davidson poses on the theory. It seems that any divergence from concrete utterances, which is necessary in interpreting belief sentences, creates a serious problem of the individuation of these nonactual utterances, especially if one is determined not to talk in intentional terms. In summary: the objects of "saying", in any "extensional dresses", will not do for belief contexts.

The second difficulty in applying the paratactic account to belief sentences concerns the conversion of the

---

10. 1987.

notion of samesaying to fit the new context. Samesaying relation is a relation which utterances, as concrete particular events, cause to hold between speakers, if they are "the same in import".[11] Samesaying can perhaps be modified to include possible utterances, by rendering it partially counterfactual. First, utterances may be construed as ordered triples of sentences, persons and times -- thereby they become abstract objects, whose existence can be established by the existence of the relevant sentence, person and time. If the relevant sentences are actually uttered by the relevant people at the relevant times, these people are samesayers if their utterances carry the same import. If either or both sentences are not uttered, the samesaying relation holds iff, had the sentences been uttered by the people at that time, they would have been producing utterances that would have stand to each other in the appropriate sameness of import.

When applying this modification to the analysis of belief sentences, we obtain the following: samesaying relations holds between the attributor and the believer iff the utterance of the complement of the belief sentence by the attributor makes him a samesayer as the believer. This definition of the samesaying relation for the analysis of belief sentences has some disadvantages: (i) it is restricted to believers who use a language and to utterable

---

11. Davidson 1969, p.169.

beliefs; (ii) the counterfactual terms involved in it call for a nonextensional semantics; (iii) perhaps worst of all - it is not at all clear what is the nature of the relation of equivalent import is, that two possible utterances should have in order to render their producers samesayers. Are those just utterances of sentences that have the same meaning? Or those that give the same information? It seems that any attempt to redefine Davidson's "samesaying" as a relation between a believer and the attributor of the belief cannot really avoid making reference to contents.

Within the boundaries set by "sententialists" on a theory of Propositional Attitudes, Davidson's account of indirect discourse employs notions that restrict its application to attitudes like that of "saying". Any expansion of the analysis to belief sentences as normally conceived in common practice violates desiderata concerning the employment of extensional notions only, and loses the advantage of fine syntactic individuation of the objects of belief.

Davidson's conception of belief as essentially connected to speech suggests that he is willing to bite the bullet and accept the unintuitive notion of belief that results from the expansion of his theory of "saying that" to "belief that". According to him, "...making detailed sense of a person's intentions and beliefs cannot be independent

of making sense of his utterances".[12] The correctness of the belief attribution 'John believes that pigs can fly' depends entirely on John's disposition to assert or to assent to the sentence 'pigs fly' (or to a sentence with the same semantic structure). This implies that non-speakers (e.g., animals, babies) cannot have beliefs, and furthermore, that any attribution of belief by means of a sentence S is false if the believer is not disposed, for one reason or another[13] to express, communicate or assent to it. (If Davidson's view is generally accepted, most "head-shrinkers", who are essentially paid for making correct belief attributions which their patients would not admit of, will be out of work.) Davidson is not concerned with the intuitive properties of belief that his strategy leaves out, e.g., the folk psychological role beliefs play in the explanation of behavior. From his point of view "We have the idea of belief only from the role of belief in the interpretation of language, for as a private attitude it is not intelligible except as an adjustment to the public role provided by language. And given the dependence of other attitudes on belief, we can say more generally that only a creature that can interpret speech can have the concept of belief."[14] Whether only speakers can have the concept of

---

12. 1974, reprinted in Inquiries into Truth and Interpretation, p.144.

13. The belief indicated might be unconscious, or a tacit presupposition.

14. 1975, reprinted in Truth and Interpretation p.170.

belief may or may not be disputed, but Davidson's conclusion -- that only speakers have beliefs -- diverges radically from what the common practice of belief attribution suggest, which is the subject of my investigation.

SECTION IV

Mental Sentence Theories attempt to explain facts about our commonsense notion of belief and belief attribution by taking belief to be a relation between an organism and an internally represented sentence token. The most well developed theory of this kind is put forward by Jerry Fodor. Unlike most other theorists of propositional attitudes, Fodor is concerned not only with the semantic problems associated with belief attribution, but also with a theory that fits with the assumptions and the findings of empirical cognitive psychology.

Fodor's account makes two major claims: (i) there is a language of thought; (ii) 'that p' in (1) denotes a sentence in this language. Consequently, the objects of belief in this theory are mental sentences. At the heart of Fodor's theory of mental representation is the hypothesis that we represent the world to ourselves by storing sentences of a mental language (also called Mentalese, or the Language of Thought, or the Language of Internal Representations). This language has, according to Fodor, the following properties:

it is innate;[15] it is universal; and it is a computational system, with syntax and semantics. These general assumptions are supported, in his view, by the "only psychology we have", and obtain their theoretical justification from being crucial to the program of explaining behavior by reference to mental acts.

A correct belief attribution accomplishes, according to Fodor, in effect two things: it uses a declarative sentence in a natural language (i.e., the complement sentence) to pick out a sentence of the language of internal representations, and it also "locates" a token of the latter in a particular "box" -- the one marked "beliefs". Fodor says:

> To believe that such and such is to have a mental symbol that means that such and such tokened in your head in a certain way; it's to have such a token in your belief box...correspondingly, to hope that such and such is to have a token inscribed in your 'hope' box (1987, p.17).

For a mental state to belong to the "belief box" this state must play a specific causal role (explicated in terms of causal interactions between other mental states, stimuli and behavior) in the mental dynamics of the organism. To have a belief is simply to have a sentence token inscribed in the brain that exhibits the causal interactions appropriate to belief.

---

15. This assumption helps in explaining learning of a first language via the already available mental representations.

Fodor develops his theory of belief attributions within the context of a functionalist theory of the mind and a defense of folk psychology. The former leads him to individuate beliefs according to their functional role in the mental life of the organism; the latter dictates their individuation according to their content. There are, in his account, three separate ways of individuating beliefs:

1. **syntax** of the language of thought

2. **functional role** of the belief state

3. **content** of the belief state

The assumption that the first two criteria coincide in their applications may not be problematic. As fodor says: "The syntax of a symbol might determine the causes and effects of its tokenings much the way that the geometry of a key determines which locks it will open (1987, p.17)". It is difficult, however, to see how the functional role of a belief state and its content generate the same identity conditions for belief states. Fodor has developed a sophisticated account of content, based on David Kaplan's notion of character, to explain how the last two criteria individuate beliefs in the same way.

One of Fodor's main concerns is to construct a theory of propositional attitudes that accords with cognitive theories about the workings of the mind. Fodor's theory provides an ingenious defense of folk psychology in the

context of methodological solipsism,[16] which many have considered impossible. The issue raises many interesting questions as to how mental representations refer to the world; whether their meanings are a matter of their role in inference, decision making and other mental processes etc. As the primary concern of this chapter is with the semantics of belief, this aspect of Fodor's theory will not be further developed.

The opacity of belief context can be accounted for on this theory in the fashion furnished by sententialism. A belief attribution asserts a relation between a person and a mental sentence. Since belief attributions with different complement sentences p, q, may refer to different mental sentences, their substitution does not necessarily preserve the truth values of the original attribution, regardless of whether p and q are semantically equivalent.

Fodor's mental sentence theory seems to be the best of the sentential theories discussed in this chapter. As in the quotational theories, the opacity of belief is explained because the 'that' clause is seen as denoting a sentence. As in the paratactic theory, the problems of ambiguity and indexicality are overcome by taking the complement sentence of the belief attribution as used. The problems related to

---

16. Methodological solipsism is empirical theory about the mind, which takes mental processes to be computational, hence syntactic. Psychological states are individuated without respect to their semantic evaluation, i.e., to how the state corresponds to the world.

language dependency are avoided with the help of the assumption of the universality of the mental language. Inexpressible beliefs, which posed a problem earlier, can be accounted for because not all sentences in the language of thought need be translatable into a natural language. Finally, belief attributions to animals (or any organisms that don't use a language) do not create a problem for Fodor's theory. The fact that a dog can be correctly described as believing that it is raining may cause difficulties for other sentential theories, but not for a theory that employs the notion of mental representations that all believers share.

One difficulty with Fodor's theory concerns implicit beliefs.[17] Fodor appears to be committed to the view that only representations in the belief box constitute beliefs. This implies that all beliefs are explicit. The problem is that the propositional attitudes we have far outstrip those that we (in some sense) actively entertain, for instance the belief that zebras in the wild don't wear overcoats. Even if Fodor distinguishes somehow between explicit and implicit beliefs, the problem is that there are infinitely many implicit beliefs. It is implausible to assume that we have infinitely many mental sentences in our belief boxes. It is specially important for Fodor to account for the phenomenon of implicit beliefs, because implicit beliefs play a crucial

17. This difficulty was first put forward by Daniel Dennett, 1981 p.74.

role in the folk psychological explanations he aims at preserving.

Another important aspect in which Fodor's theory is incomplete concerns the question of how can we **recover the object of a belief** from the complement sentence that attributes it. Sometimes Fodor appears to think that the answer is simple. He suggests that one can recover the object of belief by asking oneself what belief he would be expressing by using that complement sentence. Fodor tells us that an internal formula `p' could be implicated "as the object of the communicative intentions that utterances of 'p' standardly function to express" (1987, p.169). The complement sentence 'pigs fly' in 'Bill believes that pigs fly' is a sentence English speakers would use to express a certain internal formula, were they to have a 'belief-making' relation to it and wish to convey it in words. This internal formula is identified by its overall functional role in our mental system (including the role it has in regulating the production of the sentence 'pigs fly'). It seems to me that this solution is too simple. We do a lot more than this in the attempt to pin down a belief attributed: we often take into account background information about the believer, contextual elements relating to the participants in the conversation, etc. In the appropriate context, the sentences 'Quine believes that Plato is smart' and 'Edith Bunker believes that Plato is smart' (where the last occurrence of 'Plato' refers to a dog

in Disney's cartoons, and that fact is known to the hearers of the belief attributions) will not be used to attribute the same belief to these two agents. To my opinion, many important pragmatic and contextual factors have to be considered in belief attribution, if we are to obtain a philosophical picture of the objects of belief and their semantic relations to sentences ascribing them.

By identifying the object of belief with a mental sentence, Fodor is identifying one theoretical construct (a construct that captures the semantic referent of a 'that' clause in a belief ascription) with an empirical construct that is embedded in empirical cognitive theories. While some questions about objects of belief may belong to the empirical realm, the above identification is problematic. There is a dilemma here: either one maintains that the concept of belief implies the existence of a language of thought, or one views the existence of this language as an empirical (hence contingent) assumption. The former hypothesis is implausible, because of the additional assumptions, e.g., of universality and innateness, Fodor makes. These are obviously not implied by the folk concept of belief. On the other hand, if the existence of the language of thought is an empirical hypothesis, then identifying objects of belief with mental sentences faces the following problem. If there were no language of thought, then belief attributions are rendered, on Fodor's account, meaningless or truth-valueless. That is so because

on this account belief attributions would belong to the same category as 'the present king of France'. I.e., the complement of any belief attribution becomes, in the absence of a language of thought, a denotationless term.

Chapter 7

Informational Contents

The main reason why the problem of belief sentences seems so intractable can be traced, I will argue, to the underlying philosophical framework that many (if not most) theorists share. This framework is analyzed and rejected in the first section of this chapter. In the second section I sketch a different approach to objects of belief, and indicate how they are picked out in belief attribution. In the last section I attempt a semantic solution to the problem of deductive closure.

SECTION I

The framework that many theorists of belief sentences share consists of four theses, two of which I accept:

1. 'Believes' is a two-place relation.

2. The objects of belief are not linguistic entities. I.e., they are not utterances, inscriptions, sentence-types, etc.

The first of these is discussed and defended in Chapter 2, and the second in Chapter 6. The remaining two theses, however, I wish to challenge:

3. Objects of belief are identical to the meanings of sentences.

4. The object of belief can be determined as a function solely of the constituents of the complement of the belief sentence.

Before calling these theses into question, let us examine the intuitions behind identifying objects of belief with meanings:

A. Objects of belief, like sentences, have truth values. Furthermore, the same semantic coordinates can be used to determine the truth or the falsity both of a sentence and of an object of belief.

B. Both objects of belief and meanings are abstract entities that can be conveyed by sentences.

C. Laws of rationality, that are implicit in the practice of belief attribution, are defined on sentences (whether they are laws of deductive logic or of probability).

D. External factors that are involved in determining meanings and truth conditions are also involved in the relations between the believer and his environment. For instance, we expect one to acquire the belief that the kitchen is on fire, when that is the case, and the person is appropriately situated with respect to the situation. We explain why he believes that the kitchen is on fire by the way the world is and by his opportunity to observe it.

Under suitable circumstances, we also obtain information about the world from beliefs people express.

As a result of these intuitions, most accounts view belief as a relation between an agent and a meaning (under one definition or another). There is traditional agreement to the idea that concepts that are necessary to describe a language are also necessary to describe mental states and attitudes. The recent philosophical framework, derived from Frege, has a strong flavor of a semantic prejudice: linguistic meaning is directly connected to the concept of belief. This connection can be explicated in terms of truth conditions, or in terms of assent conditions, or in terms of intentions of speakers, etc. The point I wish to emphasis is that all challenges to Frege's conception are still embedded in the tradition of linking linguistic content with belief content. Within this framework, mental attitudes such as belief are explained as attitudes to some semantic content. This content is identical to (or is of the same nature as) the semantic content of the attributing clause of the belief sentence. The object of belief, i.e., the object denoted by the belief complement, is assumed by the "semantic doctrine" to be (or to have) both a semantic content, and the content of the belief ascribed. That means that the properties of belief, like those mentioned above, are identical to semantic properties of sentences. The fact that beliefs resist semantic and logical rules usually associated with meanings is to be accounted for, on this

framework, by rendering the semantic interpretation of the complement sentence finer than propositions.

Concerning the third thesis above, it is my opinion that the sort of thing that is denoted by that-clauses of belief attributions is not meaning. (I will argue in the next section that it is not usually determined directly from the constituents of the attributing sentence.) I wish to present three arguments for this claim: a conceptual argument, a pragmatic argument and an argument from functionality.

The **conceptual argument** aims to show that the role played by the concept of meaning in a semantic theory is distinct from the role played by the objects of belief in our folk psychology. The notion of meaning is standardly perceived as a theoretical device for explicating semantic properties of the language. We expect this notion to play a central role in accounting for compositionality, for semantic relations between sentences (such as synonymy and translation), for logical relations between sentences (such as consistency and implication), and for the informational and communicative features of languages.

In a simplified model, the meaning of a compound linguistic expression is calculated on the basis of the meanings of the expressions' components and their syntactic arrangements. An ambiguous sentences is defined as one that has (or can be used to convey) at least two distinct

meanings. Synonymy of expressions can be explained in terms of substitution (in intensionally transparent contexts) that preserves the meaning of the original expression. The incompatibility of 'all soldiers are brave' with 'some soldiers are cowards' is traced to the meanings of 'all', 'some', 'brave' and 'coward'. A sentence implies another sentence in virtue of the meanings of the logical constants involved. Communication owes its success to shared meanings of the concepts communicated.

The notion "an object of belief", on the other hand, has two complementary roles. In semantic theory, this term labels the denotation of that-clauses of belief attributions. As such, it should be so articulated that it accounts for the semantic anomalies that belief contexts generate, and does so in a compositional manner. As a theoretical construct within a cognitive psychological theory, the notion of the object of belief participates in theories about how agents acquire beliefs, how bits of information that beliefs carry interact with other bits of information to generate new beliefs, and how beliefs function in the production and explanation of behavior.

Clearly, the roles of meaning and of "objects of belief" are distinct. There is no a priori reason to think that the same notion can play both roles. That by itself does not prove, however, that such a notion could not possibly be articulated. The **pragmatic argument**, given

next, provides some reasons to divorce the two concepts. This argument points to the existence of cases in which belief sentences that contain distinct, even extensionally non-equivalent, complements ascribe the same belief. Furthermore, there are cases of belief attributions with identical, non-indexical, complements that attribute different beliefs in different contexts.

An example of the former phenomenon is provided by a context, in which it is clear that Archie Bunker understands 'fortnight' to mean a period of ten days. In this context, the complement sentence 'Michael takes a shower every fortnight' ascribes to Archie the same belief as does the complement 'Michael takes a shower every ten days'. The two complements have different meanings, and may even have different truth values. Still they characterize in the context described the same belief state. Another example concerns one who does not distinguish between necessary and sufficient conditions (any logic teacher can testify to the alarming extent of this confusion among college students). For such a person, a complement of the form 'A if B' ascribes the same belief as does a complement of the form 'A only if B'.[1] That is so regardless of the fact that unless A and B stand for logically equivalent sentences, every pair of sentences of these forms have distinct meanings on any reasonable account of meaning.

---

1. At least this is so in a speech context in which the person's confusion is common knowledge.

This same context can provide an example for the latter type of phenomenon. The complement sentence 'Michael takes a shower every fortnight' ascribes different beliefs to the Archie and to his wife, about whom we have no indications of conceptual confusions regarding 'fortnight'. A sentence of the form 'A only if B' may ascribe two distinct beliefs to my average student -- one before having some training in reasoning, and one afterwards.

The above identification of beliefs does not assume any explicit criterion of individuation. If we follow a weak functionalist thesis, according to which a difference in functional-inferential role is sufficient (though not necessary) to distinguish between beliefs, many more examples are readily available. The sentence 'John is a doctor' picks out different beliefs for an average child and for a normal adult. The sentences 'the cat is dead' and 'the cat is asleep' may be used to attribute the same belief to a child and to an adult respectively. (Only the second of the two corresponds to a fact in the world, so these sentences differ not only in their truth conditions, but also in their truth value. This does not effect the functional role of the child's belief, if he does not distinguish between the above attributes of the cat.)

The pragmatic argument shows that the object of belief cannot be identified with the meaning of the belief-complement. But that does not establish that there is a

distinction in kind between the two concepts: one may argue (as does Stalnaker) that the object of belief is a meaning of some sentence, not necessarily of the one used in the attribution. I do not have a conclusive argument for the claim that there is a distinction in kind between the two. While I cannot prove that it is logically impossible to construct a semantic theory in which the notion of meaning can also be used for the denotation of belief clauses, plausible semantic theories will have to distinguish the two, in my opinion. The next argument, the **argument from functionality**, is meant to lend support to this claim.

The conceptual argument shows that the roles played by meanings and by objects of belief are different in their respective theories. The argument from functionality starts from the fact that the essential properties of beliefs -- their causal, functional workings in the organism's psychology -- have a subjective, private and less stable nature. The semantic properties of languages, on the other hand, are objective, public and stable. Given the role meanings plays, they are best understood when characterized in external (word to world) terms. That is why the possible-worlds theory of meaning is considered the best we have in semantics. The functional role of beliefs (that the objects of belief should capture) is different for different kinds of believers. That role depends on the conceptual and logical abilities of the believer, among other things. The internal relations of beliefs to each other and to other

mental states seem to be independent, for all psychological explanatory purposes, from the external properties beliefs have, such as their truth value.

Let me illustrate the tension between the "external" and the "internal" properties of belief with the help of two familiar examples from the literature. Pierre (in Kripke's puzzle) believes that London is ugly, but he also believes that Londres est jolie (which means that London is pretty). Pierre does not realize that the two names above refer to the same city. If his beliefs are identified by the standardly assigned semantic content of the sentences ascribing them,[2] we end up assigning to him contradictory beliefs. For reasons stronger than those that motivate the principle of charity, we are reluctant to attribute contradictory beliefs to one who has not committed any logical error. In this case, the external elements involved in identifying Pierre's beliefs impose a too coarse-grained condition on beliefs. In the twin-earth examples (originated by Putnam) the same external elements seem to impose an unnecessarily fine grained condition: the belief of the earthling that water is wet cannot be identified with the belief of his psychologically identical twin that water is wet, since 'water' on earth denotes a different object. External elements, that are standardly used to fix the meaning of an expression, render the twins beliefs

---

2. Assuming referential theory of names.

different, when considered constitutives of beliefs. But -- all relevant psychological facts seem to identify them.

In short, the argument from functionality purports to show that meaning is best understood primarily in external terms, but the objects of belief must be sensitive to internal aspects. Hence, it is implausible that adequate theories of meanings and beliefs would identify the two.

In summary, any notion of meaning that can be useful in explicating semantic properties of expressions cannot be adequate for the object of belief. The semantic content of a sentence often fails to capture the content of a belief, e.g., by failing to capture the way beliefs participate in inferences (theoretical or practical) that the agent may or may not draw. Traditional attempts to identify the two result in a notion that does neither job adequately.

SECTION II

Beliefs are generally considered to be intentional, i.e., representative, mental states. They represent or are about other things, such as objects, properties, states of affairs. These representative states have specific functional roles in the believer's cognitive life. Little is known as yet about the nature of these representations: they may turn out to be mental pictures, syntactic formulae,

electo-chemical processes or what have you. Nor do we have a clear understanding about how these representations function in the production of other representations, nor about their contributions to a wide range of mental, physical and behavioral phenomena, such as blushing, having hives, generating psychosomatic symptoms, acting towards the fulfillment of desires or having a heart attack. I do not presume to make any specific empirical assumptions concerning either one of these questions. My primary concern is not with questions about belief formation, representation or functionality, nor with the metaphysical nature of beliefs. I am concerned rather with how the belief complement relates to the belief ascribed. Whether beliefs are physical neurological states or irreducible mental states has little to do with the semantic analysis of belief attributions.

Beliefs, just like pain, love, desire and the like, are internal, in the head. We express and ascribe them with the help of a public language. The semantic factors used to characterize properties of a language are external. The semantics of belief attributions must do justice both to the internal aspect of beliefs and to the external aspects of the language used to describe and express them. I wish to outline an account of the workings of belief sentences in a natural language, an account that will harmonize with the vast range of the relevant observable phenomena. Such an account will help us understand the great success we

normally enjoy in expressing and attributing beliefs linguistically. This success is manifested, among other things, in understanding the believer's state of mind, in predicting his behavior and other beliefs he may generate, and in learning from him about the world. All this is achieved by a process of specification of a belief via the complement of the belief attribution, a process that is intuitive, fast and effective to a large extent in practice, but very complicated and intriguing to theorize about.

We have argued above that an adequate theory of the objects of belief cannot identify them with meanings of sentences. One central task before us is to articulate a new conception of objects of belief. I propose that we identify them with what I will call **informational contents**. We need to address the following two questions:

1. What are informational contents?

2. How are they picked out by the complement of the belief attribution?

Let us begin with the first issue. By the arguments given in earlier chapters, it is clear that informational contents are abstract entities; they are not concrete entities such as utterances or inscriptions. Nor are they linguistic entities such as sentences. On the positive side, informational contents have representational properties: they can be true or false. Further, they serve to

characterize belief states. For example, they can be used to describe the state's functional role in the organism's psychology.

In their representational aspects, informational contents capture both semantic and contextual information. Semantic information of an expression is sensitive to syntactic and semantic constraints, and usually coincides with the meaning of that expression. But the semantic evaluation of a sentence does not specify the information it conveys on each occasion. When embedded within a context of exchange, the information communicated by a sentence is not necessarily identical to its meaning. The contextual information conveyed by an utterance of a sentence in a conversation is certainly influenced by the semantic parameters of that sentence, but is not determined by it. Pragmatic facts such as background knowledge,[3] linguistic practices of the participants and the purpose of conveying that information shape the character of the competing alternatives (usually distinct from that of logically possible worlds) that are entertained, as well as those alternatives (not always logically equivalent ones) that are eliminated by the contextual information. The semantic, syntactic and contextual properties of the informational aspect of the objects of belief are used to connect beliefs

3. The term 'knowledge' is used here loosely, and ranges also over beliefs, e.g., those shared by the participants, in the context of deliverance of information.

as mental states with their actual semantic target in terms of reference, social conventions and the like. As a result, beliefs can be evaluated as reliable, true or false.

Informational contents characterize belief states. Belief states have a cognitive role in the organism's psychology due to the relations they bear to at least three types of objects. The relations belief states have to the world are determined by the believer's response to sensory input, and they establish the specific intentional aspect of beliefs (i.e., their semantic aboutness). The relations with other belief states are determined by psychological facts about memory, processes of reasoning employed by believers etc., and they result in the formation of new beliefs on the basis of other beliefs. The relations between beliefs and other mental states are determined by personal characteristics, feelings, propensities and the like, and they establish tendencies to act (verbally or otherwise) or to respond (as in blushing).

There is one aspect of my conception of the functional role of belief that deserves special emphasis. (It distinguishes my conception from those found in the literature, e.g., in Stalnaker's work.) I consider the role of belief in reasoning to be a crucial part of the functional role of beliefs. To use an earlier example, the functional role of beliefs attributed by the use of the complements

1. Robin will win.

2. Anyone who does not compete, or loses, will have done something Robin will not have done.

are distinct because of their different positions in the cognitive process of reasoning. This difference in positions is reflected in my conception of informational contents. The above two sentences represent the same state of affairs in the world, but, I suggest, they do not carry the same informational content. A person who is given (2) is not necessarily given the information contained in (1), though he may be able to obtain it through a logical process. Informational contents, as I understand them, not only bear logical relationships (e.g., implication) to each other, they also capture the logical and conceptual "closeness" and "distance" between the bits of information conveyed by propositions. The reasoning process by which we obtain a belief that is logically equivalent to a belief we have depends on the "informational distance" between them. Logically equivalent propositions whose informational contents are "closer" to each other (e.g., 'p or q' and 'q or p') require less calculation (or less computational time) than do more "remote" ones, such as those conveyed in the Robin example. I will return to the question of how closeness should be defined after considering the second question raised above: how are informational contents picked out by the belief complement.

Though the semantic content of the attributing clause is important in determining the informational content it denotes, it does not, by itself, determine it. The pragmatic argument given in the first section above shows that the same complement sentence, even one without any indexical items, can be used to attribute different beliefs. Hence, the **pragmatic factors** have a crucial role in determining the informational content denoted by the complement of a belief attribution in a given context. The relevant pragmatic factors can be divided into the following four categories:

A. Factors that are used to determine the **denotation** of indexical items such as 'I', 'now', 'here'. For instance, what belief is attributed by the sentence 'Saul believes that I'm hungry' depends on who utters the sentence.

B. Factors that establish the **preferred interpretation** of the complement sentence, subject to conversational constraints. For example, the sentence 'Joe believes that Mary did not drink today' may be interpreted, due to conversational implicatures, as attributing to Joe a belief that is much richer than the semantic content of the above complement indicates.

C. Factors that reflect the **purpose** of the belief attribution in a given context.

D. Factors that capture the **presuppositions** the participants in the conversation have about the believer's **background knowledge, semantic** and **conceptual abilities,** and **logical acumen.**

The first two categories are relatively clear and well recognized. The latter two require further discussion. The aim of belief attribution is to characterize belief states. There are multiple ways of classifying belief states, relative to their multiple properties. A full characterization of belief states involves all of their facets. On most occasions, belief attributions do not attempt such full characterization, but only a specification of the belief state in terms of some of its properties -- those relevant to the purpose of the belief attribution.

The idea that a belief state can be characterized in different ways can be supported by an analogy from physical theories. Physical events can be described at different aspects or levels of abstraction. In the nervous system, for instance, one can discuss the passage of an impulse across the synaptic conjunction between one nerve cell and another. Alternatively, one can look into the chemical process involved and see how the transmitter substance is broken down by the appropriate enzyme. On a different level

yet, one can look at this process by considering the behavior of atoms.[4]

I am not suggesting that the various purposes of belief attribution differ in this sort of complexity, but that the contextual goal is sometimes relevant to determining the aspects of the belief state that the informational content aims to characterize. Sometimes we aim at predicting the believer's assent to a sentence; at other times we attempt to anticipate his non-verbal behavior. Sometimes our interest is not to predict the agent's behavior, but to find out what his belief states indicates about the world.[5] In a context in which the established goal of the belief attribution is to increase the participants knowledge about the world, the informational content denoted by the belief attribution aims to characterize the semantic or intentional aspect of the belief state. The intentional properties of belief states dictate a classification of states that distinguishes between representations of different things in the world. Thus, in the twin-earth example, the informational contents of the earthling's beliefs about water are not identified with the informational contents of his twin doppelgenger's beliefs about water (assuming that the context demands the intentional or semantical taxonomy

---

4. This example is gratefully borrowed from Broadbent's "Non corporeal explanation in psychology" 1981, printed in (ed.) Heath Scientific Explanation pp. 76-98.

5. There are also other possible contextual purposes of belief attribution.

of beliefs). If, on the other hand, the purpose of the belief ascription is to predict the two agents' behavior, their belief states can perhaps be characterized by the same informational content.

The final factor relevant to determining informational contents is the participant's "picture" of the believer in question. I believe that we construct and interpret belief sentences in accordance with a **general empirical theory** we have about believers (human and animals).[6] The general theory of believers is used by us as a schema for dividing believers into large groups, according to their assumed positions on different scales. Some such scales classify believers according to the logical and conceptual abilities we attribute to them. We tend to think that dogs are smarter than cats (e.g., they are capable of producing more, and more complex beliefs on the basis of input information and other beliefs). Children are considered smarter than dogs, and people who have some training in reasoning procedures are apt to draw correct conclusions, and more of them, than people who have not. Piaget writes of a conversation with a five years old, who tells that he has a brother named Jim. When asked whether Jim has a brother,

---

6. For instance, we employ general assumptions about how beliefs are formed relative to sensory input (e.g., gap filling, causal connections making). We assume some habitual modes of thinking, some level of rationality and certain conceptual abilities. We take for granted the storage of some beliefs in the memory, the existence of tacit beliefs and some common background knowledge.

the child asserted "no". Piaget explains this phenomenon in his theory of stages of "pre-operational thought", a theory that utilizes developmental notions to classify people according to their logical and conceptual abilities. The fact that we normally don't condemn the child's belief system as inconsistent can be explained by the hypothesis that we specify his belief state with the help of informational content that is sensitive to the child's asymmetric conception of the concept 'brother'. That sensitivity is a consequence of the "conceptual position" we normally assign to five-year olds, a consequence that constrains the informational contents we take correct belief complements to denote.

There are many experiments in the psychology of reasoning whose results are not totally surprising, since they reflect our assumptions about believers. The interesting details have to be left for a later investigation, but some of them can be mentioned to help support my claim about the theory we have (and can't help employing) in interpreting belief sentences. Some studies expose the fact that given that p implies q, people tend to infer that q from p, but seldom do they infer that -p from -q. Other experiments confirm the hypothesis that people adopt and infer beliefs that are consistent with commonly held beliefs, but reject or err in reasoning about those inconsistent with what is generally accepted. We learn (what we implicitly always knew) that people reason better

with something close to everyday experience than with hypothetical abstract problems.[7]

The classification of agents according to their presumed semantic competence can be done by grouping believers by their projected ability to recognize the relation between a sentence and the world (not confined to speakers) and by their known or predicted judgments about properties like synonymy, contradictoriness, ambiguity etc. The sentence 'my cat believes that there is a prime even number of dishes in front of her' is funny because of the gap between the theoretical complexity of the concepts used in the belief attribution and the presumed conceptual ability of a cat. The informational contents ascribed to the child and to the adult, both of whose beliefs are conveyed by the sentence 'Bill is a doctor' are distinct because their different conceptual abilities affect the extent and the function of information their respective beliefs have.

The theory we have about believers and their assumed capacities is used by us in the individuation of informational contents, and in defining the logical distance between them. Let me illustrate the point with the help of a study about how people reason, done by Wason (1981). In a particular experiment, subjects were shown cards on a table.

---

7. The latter observation is discussed in Wason's "Understanding and the limits of formal thinking", 1981.

They know that each card has a letter on one side and a number on the other side. They were shown four cards which displayed the following on the side facing up:

Card #1: E

Card #2: D

Card #3: 4

Card #4: 7

Subjects were asked to point to all and only those cards that they need to turn over in order to evaluate the sentence:

If a card has a vowel on one side, then it has an even number on the other side.[8]

Not very surprisingly, most subjects pointed to Card #1, but only very few of them pointed also to Card #4. One of the possible lessons one can draw (not one Wason explicitly draws) is that this experiment supports the hypothesis that most normal people are apt to employ (implicitly or explicitly) the rule of Modus Ponens, but significantly fewer people employ the rule of Modus Tollens. This may explain the fact that we feel more justified, generally, in assuming that people follow consequences obtainable by the first rule, and less justified in assuming the application of the second.

---

8. The subjects understood the concepts used in this sentence.

Wason's experiment suggests (to my understanding) that informational contents related by Modus Ponens are logically "closer" together than those related by Modus Tollens. More generally, I am suggesting that logical distance between informational contents depends greatly on whether our believer is a baby, a child, a "layman" or a trained "reasoner". Pragmatic factors are relevant in their individuation.

Relativizing the informational content a belief complement denotes to contextual and internal factors should not result in declaring its individuation highly individualistic or ad hoc, which is the conclusion many (e.g., Bogdan, Dennett, Stich) draw. Such a result is incompatible with the phenomenon of communicative success we normally achieve in conveying and in attributing belief, a phenomenon that a semantic theory of belief sentences should explain. Informational contents, as I understand them, are objectively individuated inspite of the psychological factors that participate in their individuation.

I have not supplied a detailed theory of objects of belief in relation to belief attributions, but only sketched the general form that such a theory should take, and some of the necessary concepts it should utilize. The schema proposed above reflects a conception of objects of belief, that is not very different from many conceptions prevailing in recent publications (e.g., Stalnaker's papers); the

difference lies rather in the strategy I propose for explicating that conception.

Let us examine how some major problems concerning the logic of belief attribution fare on this account. First, the **problem of equivalence** (or logical opacity) does not arise in my account. Suppose that

a believes that p
p is logically equivalent to q.

We cannot always deduce that a believes that q, for several reasons. If q is logically remote from p, then the informational content denoted by 'that p' will obviously be different from the one denoted by 'that q'. If q is logically close to p, the purpose of the belief attribution, and the presuppositions about the believer may also block the inference. If the purpose is to predict what the agent will say, for instance, then the inference is clearly invalid.

The **problem of contradictory beliefs** does not arise in this account either. Attributing a belief in logical falsity (such as that 34 + 77 = 101 or that Cantor's naive set theory is consistent) does not result, on my account, in having to attribute the belief in everything, since the relations between beliefs are not governed by logical rules alone. In any case, this problem is intimately connected to the **problem of deduction**, which is perhaps the most

intriguing of all the problems surrounding belief sentences. A solution to this problem is outlined in the next section.

SECTION III

The assumption that agents draw inferences from their beliefs is crucial for psychological explanations and predictions (as well as in conceptions of legal and moral responsibility, historic explanations, economic theories and so on). The more general assumption of rationality of believers is well entrenched in our folk psychology. Some evidence for this is that when our predictions about an agent prove false, we tend to first cast doubt on the information that was available to the agent (e.g., 'he must have not heard that... or seen that...') before questioning the degree of rationality we normally ascribe to him.

The question of how rational should one be in order to be a believer has received answers that range from the strongest to the weakest demands on inferential ability and performance. Some think that a believer should draw all and only valid inferences (we limit the discussion to deductive reasoning). Weaker positions do not demand full deductive closure, but only "reliability" closure or closure under "feasible inferences", for which the believer draws only some (but not necessarily all) valid inferences.

Whatever the details of a theory of rationality of believers may turn out to be, it is clear that some level of deductive closure is presupposed as part of the general practice of belief attribution. If we attribute to Paul the belief that if Dukakis did not win the elections then Bush did, and that Dukakis did not win, then it does not make sense to withhold the belief attribution that Bush won. Such a context violates our very concept of belief, and calls for an explanation. Belief attributions presuppose a certain degree of rationality in the believer. In other words, the concept of belief generally requires its objects to have some deductive closure: under normal conditions, if a person believes that p and also believes that if p then q, then he believes that q.

The problem is that by standard logical rules, allowing even this minimal deductive closure to an agent results in too much deductive closure. Consider an argument in which the conclusion C is drawn from the premises P_1, P_2, ..., P_n. Suppose the argument has the form;

P_1 & P_2 & ... & P_n

A_1

A_2

.

.

.

A_k

C

where A_1, . . ., A_k are inferential steps, each of which is derived by using Modus Ponens. Minimal deductive closure implies that if Paul believes (P_1 & P_2 & . . . & P_n) then he believes A_1. By the same argument, if he believes A_1, then he believes A_2, and so on. We end up having to attribute to Paul the belief in C, no matter how long and complicated the above proof is. If Paul believes the axioms of first-order predicate logic and the axioms of Peano arithmetic, then the above argument can be used to show that he believes all theorems of Peano Arithmetic. (There are axiomatizations of first-order logic whose only rule of inference is Modus Ponens. These axiomatizations are based on axiom schemes). The problem is to account simultaneously for the logical opacity of belief and for the link belief has to rationality.

The solution I wish to put forward is based on a parallel that obtains between the above argument and the argument in the Sorites Paradox.[9] In the latter we begin with a pile of salt large enough to be described as a heap. The subtraction of a single grain of salt does not make a heap into a non-heap. Hence we can take one grain of salt from a heap, and it will still remain a heap. But then we can take another grain of salt, and then another. So by

---

9. This argument is generally attributed to Eubulides, the Megarian. Another version uses the predicate 'bald'.

continuing this process we obtain the conclusion that zero grains of salt constitute a heap. This result is absurd. The argument that creates the problem of deductive closure, as I see it, is of the same sort as the Sorites paradox, and ought to be handled in the same way.

We take predicates such as 'heap' and 'bald' to be only partially defined. That is, it is not always determinate whether or not they can be truly ascribed to an object. The correct application of these predicates survives a transition from one small change (subtraction or addition of one grain or one hair) to another. Similarly, 'believes' can be treated as a **vague two-place predicate**, i.e., as a relation that admits of degrees of application. The predicate 'believes' seems to behave, logically, in the same way as 'bald' and 'heap' do, which suggests that it is tolerant to marginal changes (in the logical and conceptual complexity of the objects it applies to), but not necessarily to large changes.[10] Furthermore, there is no objective criterion for the demarcation of the significant from the insignificant changes in the objects that 'heap' or 'believe' apply to. The criterion I believe we do use in our practice is subjective to the extent that we take the ascription of belief in the consequences of other beliefs to

---

10. Distinctions of degree of application of 'believes' to objects does not mean that small changes in such objects are insufficient to effect the truth with which the predicate 'believes' applies, but rather that there are small changes in the degree of truth with which the predicate applies.

be a function of our picture of the believer's relevant capacities, together with our implicit theory about the degrees of difficulties in reasoning for different groups of believers.

We can account for the practice of assigning partial deductive closure to believers by assigning progressively lower degrees of belief to progressively remote or complicated consequences from the agents' other beliefs. The rate of "deterioration" of these degrees is influenced by the "boundaries" assigned to the group the agent belongs to, according to our classification of agents by their logical abilities.[11]

As a result of treating 'believes' as a vague predicate, any belief attribution of the form 'a believes that p' expresses a vague proposition.[12] That means, in terms of truth conditions, that there are possible worlds in which the truth value of such a proposition is indeterminate. Though the idea involves considerable

---

11. As mentioned, my program contains the assumption that we employ (unconsciously or as background assumption) an intuitive theory about the difficulty of different patterns of inferences for different classes of agents. Patterns of inferences are ordered with respect to their relative difficulty. Agents are grouped also according to the inferences they are normally expected to employ in their reasoning, both in drawing the correct conclusions, and in avoiding prevalent fallacies (such as concluding '-q' from 'if p then q' and '-p'; or concluding 'all B are C' from 'all A are B' and 'all A are C').

12. This is not identical to a treatment by which the objects of belief themselves are vague propositions.

idealization, I will follow standard theories of vagueness, and view vague propositions as having degrees of truth. These degrees lie anywhere in the interval [0,1].[13] The degree of truth a belief attribution has is not determined, I should emphasize, entirely by the believer himself (e.g., it cannot be identified with a degree of confidence). The degree of truth depends also on the speaker-hearer's assessment of the relevant knowledge, conceptual, semantic and logical abilities of the believer in question.

Formally, 'believes' can be interpreted semantically by a function that assigns to each ordered pair <a,p> (where a is a believer and p is an informational content) a value in [0,1]. The function, thus, tells for every agent and every informational content the degree to which the first believes the latter. At the compositional level, the semantics explains how to calculate the truth-value of a compound expression from the degrees of truth assigned to its components. There are several approaches for the treatment of vague predicates on this level. For simplicity sake, I will choose the multi-valued approach to illustrate how my strategy solves the problem of deduction. (I will confine the discussion to sentential connectives only.) We determine the degree of truth of a compound expression (built out of conjunction (&), negation (-) and implication

---

13. $[0,1] = \{x : 0 \leq x \leq 1\}$, i.e., = the set of real numbers between zero and one.

(→)) by the following rules. Let v(A) be the degree of truth of A. Then,

$$v(-A) = 1 - v(A)$$

$$v(A \ \& \ B) = \text{the smaller of } v(A) \text{ and } v(B)$$

$$v(A \rightarrow B) = 1 \text{ if } v(B) \geq v(A);$$

$$= 1 - [v(A) - v(B)], \text{ otherwise.}^{14}$$

Notice that if the atomic sentences in a formula have the classical values 0, 1, then the above rules assign the usual value to the compound formula. Consider a nonclassical example: suppose that the atomic sentences P, Q, R, have respectively the degrees of truth .9, .5, .2, then the degree of truth of

$$-(P \ \& \ R) \rightarrow Q$$

is determined as follows:

$$v(P \ \& \ R) = \text{the minimum of .9 and .2}$$

$$= .2$$

So,

$$v[-(P \ \& \ R) = 1 - .2$$

$$= .8$$

We have, therefore,

$$v[-(P \ \& \ R) \rightarrow Q] = [1 - (.8 - .5)] = .7$$

---

14. There are other possible ways of evaluating compounds. I've chosen a way that seems to me especially natural.

The degree of truth of any compound sentence can be determined in the above manner, once we are given the degrees of truth of the atoms.

Let us apply the above semantic rules to the case of belief. Let us assume that Paul believes that $\underline{A}_0$. Suppose that

$$\underline{A}_0, \cdot \cdot \cdot, \underline{A}_{100}$$

are sentences, where the inference from $\underline{A}_i$ to $\underline{A}_{i+1}$ is immediate (i.e., the informational contents expressed by $\underline{A}_i$ and $\underline{A}_{i+1}$ are "close" to each other), and that the logical "remoteness" between $\underline{A}_i$ and $\underline{A}_j$ increases as i - j increases. In the framework sketched above, the belief relation holds between Paul and the informational content expressed by each $\underline{A}_i$ to various degrees. Let us use $B(P,\underline{A}_i)$ to abbreviate the sentence 'Paul believes that $\underline{A}_i$'. Suppose that

$$v[B(P,\underline{A}_i)] = (100 - i)/100).$$

Thus, we have

$$v[B(P,\underline{A}_0)] = [(100 - 0)/100] = 1$$
$$v[B(P,\underline{A}_1)] = [(100 - .01)/100] = .99$$
$$v[B(P,\underline{A}_{49})] = [(100 - .49)/100] = .51$$
$$v[B(P,\underline{A}_{100})] = [(100 - 100)/100] = 0.$$

Observe that

$$v[B(P,\underline{A}_0) \rightarrow B(P,\underline{A}_1)] = 1 - \{v[B(P,\underline{A}_0) - v[B(P,\underline{A}_1)]$$

$$= 1 - (1 - .99)$$

$$= (1 - .01)$$

$$= .99$$

More generally, for each i we have

$$v[B(P,\underline{A}_i) \rightarrow B(P,\underline{A}_{i+1})]$$

$$= 1 - [(100 - i)/100 - (100 - (i+1))/100]$$

$$= 1 - [(100 - i) - (100 - (i+1))/100]$$

$$= 1 - [(100 - i - 100 + i + 1) / 100]$$

$$= 1 - (1/100)$$

$$= 1 - .01$$

$$= .99$$

If follows that the conjunction of the sentences

$$B(P,\underline{A}_i) \rightarrow B(P,\underline{A}_{i+1}),$$

for all i, also has the degree of truth .99. Hence, we can assert the following with a high degree of truth

(i) If Paul believes one informational content, then he believes others that it implies "closely".

Attributing to Paul a belief in a remote consequence of his beliefs receives, however, a proportionately lower degree of truth as the "remoteness" of that consequence increases. For instance, we have

$$v[B(P,\underline{A}_0) \rightarrow B(P,\underline{A}_{100})] = 1 - (1 - 0) = 0.$$

So, we cannot say truly of Paul that

(ii) If Paul believes $\underline{A}_0$ then Paul believes $\underline{A}_{100}$.

This shows that we are not forced to attribute to Paul's beliefs complete deductive closure. We have seen in (i) above that we can attribute to Paul the belief in all "close" consequences of his beliefs, so we can have, in short, some deductive closure without having complete deductive closure.

My aim in the above is not to argue that the many-valued treatment of vagueness is <u>the</u> correct one, but simply to illustrate how the problem of deductive closure can be solved by taking 'believes' to be a vague predicate. There are other treatments of vagueness--such as the supervaluation approach--and these may for some purposes be superior.

(The above framework may also be useful for solving the difficulties concerning the attribution of **tacit beliefs.** Tacit beliefs are generally defined as "obvious" consequences of the explicit beliefs the agent has, or as beliefs that follow "trivially" from the explicit ones. These beliefs, such as the belief that 23,569 > 22,487, are intuitively attributable in many contexts, even if it is clear that they have never been entertained by the agent. If our theory of believers is equipped with some measurement of the "obviousness" or "triviality" by which tacit beliefs follow from others for each group of believers, the treatment of 'believes' as a vague predicate may enable one

to assimilate degrees of "obviousness" to degrees of believing.)

Some may claim that the treatment of 'believes' as a vague predicate adds unnecessary complication to a semantic theory of belief attribution. Why not refine the sense of 'believes' so as to eliminate the indeterminacy of its application? Perhaps one could provide a general explanation of where, in a series of the relevant sort, 'believes' may be applied correctly for the last time. This suggestion echoes the Frege-Russell view, by which the vagueness of ordinary language is considered a flaw that should be eliminated. On the surface it may seem as if this view has been abandoned, but I believe that it creeps into the semantic treatment of natural language. It shows in the tendency to limit the recognizable class of vague predicates to some small number of observational predicates (like colors), while ignoring the fact that other semantic incoherences, such as those involved in belief attributions, may have their source in vagueness also. Most, if not all, of the theories that try to cope with the semantic problems associated with belief sentences take 'believes' to be a precise predicate whose application for any agent and any "object of belief" is definitely true or false. The burden of "precisifing away" the indeterminacy of belief attributions falls, then, on the way the objects of belief are individuated. As we have seen, these theories fail to

account for partial deductive closure of our beliefs, either by allowing too much or too little of it.

Perhaps the tolerance of 'believes' to marginal changes in its objects could be stipulated away with the help of a substantive theory of rationality. But -- should it? In my opinion, belief is essentially a coarse predicate, and we would have little use for a precisely demarcated analogue for 'believes' in the contexts in which the term is typically employed. The utility and explanatory role of 'believes' imposes upon its semantics this tolerance with respect to marginal changes in various aspects of the objects of belief.[15]

In summary, I have argued that the problematic logical issues concerning belief attributions can be overcome by taking the objects of belief to be informational contents, and by taking 'believes' to be a vague predicate.

---

15. This claim may be supported by an interesting analogy, concerning the suggestion to "precisify" color predicates. It has been estimated that a trained normal observer can, under optimal conditions, discriminate something in the order of ten million surface colors. There are, however, about half a million colors that are considered commercially different. There are, on the other end, about 7,500 color names, that can be reduced to 267 equivalent classes. (These are results of research done by the Inter Color Council, National Bureau of Standard Methods of Designating Colors.) Out of those, only eleven were found in a cross culture study (in studies done by Berlin and Kay, 1969) to be basic color terms. The question, then, is how much refinement should be done with the meaning of 'red' to eliminate its vagueness without destroying its use in communication?

-- Bibliography --

Aczel, P. 1988: Non-Well-Founded Sets, CSLI, Lecture Notes 14.

Anderson, A.R. and Belnap, N. 1975: Entailment: the Logic of Relevance and Necessity, Princeton University Press, Princeton.

Anderson, A. 1983: "The paradox of the Knower", Journal of Philosophy 80, pp.338-355.

Barwise, J. 1978: "Scenes and other situations", Journal of Philosophy 78, pp.369-397.

Barwise, J. and Etchemendy, J. 1987: The Liar: An Essay on Truth and Circularity, Oxford University Press, Oxford.

Bealer, G. 1982: Quality and Concept, Clarendon Press, Oxford.

Bennett, J. 1982: "Psychology and semantics: Comments on Schiffer's 'Intention-based semantics'", Notre Dame Journal of Formal Logic 23, pp.258-262.

Bigelow, J. 1978 (a): "Believing in semantics", Linguistic and Philosophy 2, pp.101-144.

Bigelow, J. 1978 (b): "Semantics of thinking, speaking and translation", in (eds.) Guenthner and Guenthner-Reutter, Meaning and Translation, Duckworth, London, pp.108-136.

Binkely, R. 1968: "The surprise examination in modal logic" Journal of Philosophy 65, pp.127-136.

Blackburn, S. 1975: "The identity of propositions" in (ed.) Blackburn, Meaning, Reference and Necessity, Cambridge University Press, Cambridge, pp.182-205.

Bogdan, R. 1986 (a): "The importance of belief", in (ed.) Bogdan, Belief: Form, Content and Function, Clarendon Press, Oxford, pp. 1-16.

Bogdan, R. 1986 (b): "The manufacture of belief", ibid, pp.149-184.

Bogdan, R. 1988: "Mental attitudes and common sense psychology: a case against elimination" Nous 22, pp.369-399.

Burdick, H. 1982: "A logical form for the propositional attitudes", Synthese 52, pp.185-230.

Burge, T. 1978: "Buridan and epistemic paradox", Philosophical Studies 34, pp.21-35.

Burge, T. 1984: "Epistemic paradox", Journal of Philosophy 81, pp.5-28.

Burns, L. 1986: "Vagueness and coherence", Synthese 68, pp.487-513. Reprinted in Philosophers Annual 9, pp.25-51.

Carnap, R. 1947: Meaning and Necessity, The University of Chicago Press, Chicago.

Cherniak, C. 1986: Minimal Rationality, The MIT Press, Cambridge, MA.

Church, A. 1950: "On Carnap's analysis of statements of assertion and belief", Analysis 10, pp. 97-99. Reprinted in (eds.) Davidson and Harman, 1975, The Logic of Grammar, Dickenson Publishing Co., California, pp.129-130.

Church, A. 1954: "Intensional isomorphism and the identity of belief", Philosophical Studies 5, pp.65-73. Reprinted in (eds.) Salmon and Soames, 1988, Propositions and Attitudes, Oxford University Press, pp.159-168.

Cresswell, M. J. 1972: "Intensional logic and logical truth", Journal of Philosophical Logic 1, pp.2-15.

Cresswell, M. J. 1973: Logics and Languages, Methuen, London.

Cresswell, M. J. 1975: "Hyperintensional logic", Studia Logica 34, pp. 25-38.

Cresswell, M. J. 1978: "Semantic competence", in (eds.) Guenthner and Guenthner-Reutter, Meaning and Translation, Duckworth, London, pp.9-28.

Cresswell, M. J. 1980: "Quotational theories of propositional attitudes", Journal of Philosophical Logic 9, pp.17-40.

Cresswell, M. J. 1985: Structured Meanings: The Semantics of Propositional Attitudes, The MIT Press, Cambridge, MA.

Davidson, D. 1968: "On saying that", Synthese 19, pp.130-146. Reprinted in Davidson, 1984, Inquiries into Truth and Interpretation, Oxford University Press, Oxford, pp.93-108.

Davidson, D. 1974: "Belief and the basis of meaning", Synthese 27, pp.303-324. Reprinted in Davidson, 1984, Inquiries into Truth and Interpretation, Oxford University Press, Oxford, pp.141-154.

Davidson, D. 1975: "Thought and talk", in (ed.) Guttenplan Mind and Language, Oxford University Press, pp.7-24. Reprinted in Davidson, 1984, Inquiries into Truth and Interpretation, Oxford University Press, Oxford, pp. 155-170.

Dennett, D. 1977: "A cure for the common code", Mind 86, pp.265-280. Reprinted in (ed.) Block, 1981, Reading in the Philosophy of Psychology vol. II, Harvard University Press. pp.64-77.

Dennett, D. 1978: Brainstorms, Bradford Books, Montgomery, Vermont.

Dennett, D. 1981: "Three kinds of intentional psychology", in (ed.) Healey, Reduction, Time and Reality, Cambridge University Press, Cambridge, pp.37-62.

Dennett, D. 1981: "True believers: the intentional strategy and why it works", in (ed.) Heath, Scientific Explanation, Clarendon Press, Oxford, pp.53-75.

Dennett, D. 1982: "Beyond belief", in (ed.) Woodfield, Thought and Object, Clarendon University Press, Oxford, pp.1-96.

Donnellan, K. 1966: "Reference and definite description", Philosophical Review 75, pp.281-304.

Dretske, F. 1982: Knowledge and the Flow of Information, The MIT Press, Cambridge, MA.

Ducasse, C.J. 1940: "Propositions, opinions, sentences and facts", Journal of Philosophy 37, pp.701-711.

Field, H. 1978: "Mental Representations", Erkenntnis 13, pp.9-61. Reprinted in (ed.) Block, 1981, Readings in the Philosophy of Psychology vol.II, Harvard University Press, Cambridge, MA, pp.78-116.

Field H. 1986: "Stalnaker on intentionality", Pacific Philosophical Quarterly 67, pp.89-112.

Fitch, G. 1985: "On the logic of belief" Nous 19, p.205-228.

Fodor, J. 1978: "Propositional attitudes", Monist 61, pp.501-523. Reprinted in (ed.) Block, 1981, Readings in the Philosophy of Psychology vol. II, Harvard University Press, Cambridge, MA, pp.45-63.

Fodor, J. 1981: Representations, The MIT Press, Cambridge, MA.

Fodor, J. 1982: "Cognitive science and the twin earth problem", Notre Dame Journal of Formal Logic 23, p.98-118.

Fodor, J. 1987: Psychosemantics, The MIT Press, Cambridge, MA.

Frege, G. 1892: "On sense and reference", in (eds.) Geach and Black, 1952, Translations from the Philosophical Writings of Frege, pp.56-78.

Grice, H. 1975: "Logic and conversation", in (eds.) Davidson and Harman, The Logic of Grammar, Dickenson Publishing Co., California, pp.64-75.

Gupta, A. 1982: "Truth and paradox", Journal of Philosophical Logic 11, pp. 1-60. Reprinted in (ed.) Martin, 1984, Recent Essays on Truth and the Liar Paradox, Oxford University Press, Oxford.

Halpern, J. and Moses, Y. 1986: "Taken by surprise", Journal of Philosophical Logic 15, pp.281-304.

Harman, G. 1972: "Logical form", Foundations of Language 9, p.38-65. Reprinted in (eds.) Davidson and Harman, 1975, The Logic of Grammar, Dickenson Publishing Co., California, pp.289-307.

Harman, G. 1975: "Language, thought and communication", in (ed.) Gunderson, Minnesotta Studies in the Philosophy of Science vol. III, pp.270-298.

Harman, G. 1982: "Conceptual role semantics", Notre Dame Journal of Formal Logic 23, pp.242-256. Reprinted in (ed.) Lepore, 1987, New Directions in Semantics, Academic Press, London, pp.55-82.

Hill, D. 1981: "Mental representations and the languages of thought" in (ed.) Block, 1981, Readings in the Philosophy of Psychology vol.II, Harvard University Press, Cambridge, MA, pp.11-20.

Hintikka, J. 1962: Knowledge and Belief, Cornell University Press, Ithaca.

Israel, D. 1987: "Review of Cresswell's Structured Meanings", Journal of Symbolic Logic 52, pp.878-882.

Kaplan, D. 1968: "Quantifying in", Synthese 19, pp.178-214. Reprinted in (eds.) Davidson and Harman, 1975, The

Logic of Grammar, Dickenson Publishing Co., California, pp.160-183.

Kaplan, D. 1975: "How to Russell a Frege-Church", Journal of Philosophy 72, p.716-729. Reprinted in (ed.) Loux, 1979, The Possible and the Actual, Cornell University Press, Ithaca and London, pp.210-224.

Kripke, S. 1976: "A puzzle about belief", in (ed.) Margalit, Meaning and Use, D. Reidel Publishing Co., Dordrecht, pp.239-283.

Landesman, C. 1972: Discourse and its Presuppositions, Yale University Press, New Haven, CT.

Lepore, E. and Loewer, B. 1987: "Dual aspect semantics", in (ed.) Lepore, 1987, New Directions in Semantics, Academic Press, London, pp.83-112.

Lewis, D. K. 1970: "General semantics", Synthese 22, pp. 18-67. Also reprinted in Lewis, 1983, Philosophical Papers vol. I, Oxford University Press, Oxford.

Lewis, D. K. 1981: "What puzzling Pierre does not believe", Australasian Journal of Philosophy 59, pp.283-289.

Loar, B. 1982: "Conceptual role and truth condition", Notre Dame Journal of Formal Logic 23, pp.272-283.

Loar, B. 1987: "Subjective intentionality", Philosophical Topics 15, pp.89-124.

Loewer, B. 1982: "The role of conceptual role semantics", Notre Dame Journal of Formal Logic 23, pp.305-315.

Lycan, W. 1972: "Davidson on saying that", Analysis 33, pp.138-139.

Lycan, W. 1986: "Tacit beliefs", in (ed.) Bogdan, Belief: Form, Content, and Function, Clarendon Press, Oxford, pp.61-82.

Machina, K. 1976: "Truth, belief and vagueness", Journal of Philosophical Logic 5, pp. 47-78.

Martin, R. M. 1969: Belief, Existence and Meaning, New York University Press, New York, NY.

Mates, B. 1952: "Synonymity", in (ed.) Linsky, Semantics and the Philosophy of language, University of Illinois Press, Urbana, IL, pp.111-138.

Mates, B. 1981: Skeptical Essays, University of Chicago Press, Chicago.

McDowell, J. 1980: "Quotation and saying that", in (ed.) Platt, Reference, truth and reality, Routledge & Kegan Paul, London, pp.206-237.

Montague, R. and Kaplan, D. 1960: "A paradox regained", Notre Dame Journal of Formal Logic 1, pp.79-90. Reprintrd in 1974, (ed.) Thomason, Formal Philosophy, Yale University Press, pp.271-285.

Montague, R. 1963: "Syntactic treatments of modality, with corollaries on reflextion principles and finite axiomatizability", Acta Philosophica Fennica 16, pp.153-167. Reprintrd in 1974, (ed.) Thomason, Formal Philosophy, Yale University Press, pp.286-302.

Partee, B. 1973: "The semantics of belief-sentences", in (eds.) Hintikka, Moravcsik and Suppes, Approaches to Natural Language, D. Reidel, Dordrecht, pp. 309-336.

Partee, B. 1977: "Possible-worlds semantics and linguistic theory", The Monist 60, pp.303-325.

Peacocke, C. 1981: "Are vague predicates incoherent?", Synthese 48, pp.121-141.

Pendlebury, M. 1987: "Stalnaker on inquiry" Journal of Philosophical Logic 16, pp.229-272.

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

Pollock, J. 1982: Language and Thought, Princeton University Press, Princeton, NJ.

Pollock, J. 1988: "My brother, the machine", Nous 22, pp.173-212.

Powers, L. 1976: "Comments on Stalnaker's 'Propositions'", in (eds.) MacKay and Merrill, Issues in the Philosophy of Language, Yale University Press, New Haven, pp.93-103.

Prior, A. 1961: "On a family of paradoxes", Notre Dame Journal of Formal Logic 2, pp.16-32.

Putnam, H. 1954: "Synonymy and the analysis of belief sentences", Analysis 13-14, pp. 114-121.

Putnam, H. 1975: "The meaning of 'meaning'", in (ed.) Gunderson, Minnesotta Studies in the Philosophy of Science vol. III, University of Minnesotta Press, pp.131-193.

Putnam, H. 1978: "Meaning, reference and stereotypes" in (eds.) Guenthner and Guenthner-Reutter, Meaning and Translation, Duckworth, London, pp.61-82.

Putnam, H. 1982: "Comments on Fodor's cognitive psychology", Notre Dame Journal of Formal Logic 23, pp.294-295.

Quine, W. 1952: "On a supposed antinomy", Mind 62. Reprinted in Quine, 1966, The Ways of paradox, Harvard University Press, Cambridge, MA, pp.19-21.

Quine, W. 1953: "Three grades of modal involvement", Proceedings of the 11th International Congress of Philosophy, Brussels. Reprinted in The Ways of Paradox, ibid, pp.158-176.

Quine, W. 1955: "Quantifiers and propositional attitudes", Journal of Philosophy 53. Reprinted in The Ways of Paradox, ibid, pp.185-196.

Quine, W. 1960: Word and Object, The M.I.T. Press, Cambridge, MA.

Quine, W. 1977: "Intensions revisited", Midwest Studies in Philosophy 2. Reprinted in Quine, 1981, Theories and Things, Harvard University Press, Cambridge, MA, pp.113-123.

Russell, B. 1910: "On the nature of truth and falsehood", in Philosophical Essays, Simon & Schuster, New York, 1966, pp. 147-159. Originally published by George Allen & Unwin in 1910.

Russell, B. 1918: "The philosophy of logical atomism", in (ed.) Marsh, 1956, Logic and Knowledge, George Allen and Unwin, London.

Sainsbury, R.M. 1988: Paradoxes, Cambridge University Press, Cambridge.

Scheffler, I. 1954: "An inscriptional approach to indirect quotation", Analysis 14, pp.83-90.

Scheffler, I. 1963: The Anatomy of Inquiry, Bobbs-Merrill, Indianapolis, IN.

Schiffer, S. 1981: "Truth and the theory of content", in (eds.) Parret and Bouveresse, Meaning and Understanding, Walter de Gruyter, Berlin and New York, pp.204-224.

Schiffer, S. 1982: "Intention based semantics", Notre Dame Journal of Formal Logic 23, pp.119-156.

Schiffer, S. 1986: "The real trouble with propositions", in (ed.) Bogdan, Belief: Form, Content and Function, Clarendon Press, Oxford, pp.83-118.

Schiffer, S. 1986: "Stalnaker's problem of intentionality", Pacific Philosophical Quarterly 67, pp.87-97.

Schiffer, S. 1987 (a): Remnants of Meaning, The MIT Press, Cambridge MA.

Schiffer, S. 1987 (b): "Extensional semantics and sentential theories of belief", in (ed.) Lepore, 1987, New Directions in Semantics, Academic Press, London, pp.113-142.

Skyrms, B. 1978: "An immaculate conception of modality", Journal of Philosophy 75, pp.368-387.

Soames, S. 1987: "Direct reference, propositional attitudes and semantic content", Philosophical Topics 15, pp.47-88.

Stalnaker, R. 1976: "Propositions", in (eds.) MacKay and Merrill, Issues in the Philosophy of Language, Yale University Press, New Haven, pp.79-91.

Stalnaker, R. 1977: "Complex predicates", The Monist 60, pp.327-339.

Stalnaker, R. 1984: Inquiry, The MIT Press, Cambridge MA.

Stalnaker, R. 1986: "Replies to Schiffer and Field", Pacific Philosophical Quarterly 67, pp.113-123.

Stalnaker, R. 1987: "Semantics for belief", Philosophical Topics 15, pp.177-190.

Stalnaker, R. 1989: "Mental content and linguistic form" (forthcoming).

Stich S. 1982: "On the ascription of content", in (ed.) Woodfield, Thought and Object, Clarendon Press, Oxford, pp.153-206.

Stich S. 1983: From Folk Psychology to Cognitive Science: The Case Against Belief, The MIT Press, Cambridge MA.

Stich S. 1986: "Are belief predicates systematically ambiguous?", in (ed.) Bogdan, Belief: Form, Content and Function, Clarendon Press, Oxford, pp.119-148.

Thomason, R. 1980: "A model theory for propositional attitudes", Linguistics and Philosophy 4, pp. 47-70.

Thomason, R. 1980: "A note on the syntactic treatment of modality", Synthese 44, pp.391-395.

Wason, P. 1981: "Understanding and the limits of formal thinking", in (eds.) Parret and Bouveresse, Meaning and Understanding, Walter de Gruyter, Berlin and New York, pp.411-422.

Wright, C. 1975: "On the coherence of vague predicates", Synthese 30, pp.325-365.