Order Number 9405504

Understanding language: A Luddite approach

Brown, Martin Andrew, Ph.D.

City University of New York, 1993

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

A

UNDERSTANDING LANGUAGE: A LUDDITE APPROACH

by

MARTIN BROWN

A dissertation submitted 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.

1993

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.

August 5, 1993 — Date — Richard Mendelsohn — Chair of Examining Committee

August 5, 1993 — Date — Richard Mendelsohn — Executive Officer

Professor Arthur Collins

Professor Jerrold Katz

Professor Richard Mendelsohn

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

To Leigh, all the credit and none of the blame.

CONTENTS

Chapter 1	CHOMSKY'S THEORY: KNOWLEDGE OF GRAMMAR AS AN EXPLANATION OF LANGUAGE USE AND UNDERSTANDING	
1.1	Introduction	1
1.2	Competence and Performance	11
1.3	Tacit Knowledge	22
1.4	The Thesis of Competence and Chomsky's Conceptual Scheme	40
Chapter 2	UNDERSTANDING A SENTENCE IN CHOMSKY'S THEORY	
2.1	Introduction: How to Understand a Sentence	55
2.2	Explicit Representation	60
2.3	Mental Representation	64
2.4	Representing Rules	67
2.5	Physicalism	74
2.6	Computer Model of Mind	75
2.7	Conclusion	82
Chapter 3	FUNCTIONALISM AND THE COMPUTER MODEL OF THE MIND	
3.1	Introduction	84
3.2	Hornsby's Criticism of Functionalism	87
3.3	The Idea of Functionalism	91
3.4	Functionalism and Theories	95
3.5	What Psychological Theory?	98
3.6	Functionalism and Physicalism	100
3.7	Behaviour	107
3.8	Output	117
3.9	Behaviour and Output Compared	127
3.10	Actions and Bodily Movements are not 'Token Identical'	130
3.11	Perception	134
3.12	Input	140
3.13	Input and Perception Compared	154
3.14	Refutation of Functionalism	156
3.15	Functionalism and the Computational Theory of Mind	158
3.16	Scientific Psychology	163
Chapter 4	MEANWHILE, INSIDE THE HEAD	
4.1	Computers, Symbols and Brains	167
4.2	Haugeland on Formal Systems	170
4.3	What "Symbol" Means	174
4.4	Symbols and Representation	177
4.5	Formal Systems and the Possibility of Error	179

4.6	Language as a Formal System	183
4.7	Automatic Formal Systems	185
4.8	Computers and Natural Physical Systems	190
4.9	Computers and Error	191
4.10	The Intentional Character of Rule-Following	194
4.11	The Difference Between Brains and Computers	197
4.12	Biological Standards	198
4.13	Chomsky Redux	201

BIBLIOGRAPHY 206

CHAPTER 1

CHOMSKY'S THEORY: KNOWLEDGE OF GRAMMAR AS AN EXPLANATION OF LANGUAGE USE AND UNDERSTANDING

1.1 INTRODUCTION

> Science explained people, but could not understand them. After long centuries among the bones and muscles, it might be advancing to knowledge of the nerves, but this would never give understanding. (E.M. Forster, Howards End)

> The big thing about language is that we form new sentences with old words. (Wittgenstein, Lectures on Philosophical Psychology p146)

Anyone who speaks a language can use and understand an endless variety of sentences never heard or used before. This ability to use and make sense of novel arrangements of familiar words is mundane but remarkable. How is it possible? Here is a natural and plausible explanation:

Understanding a sentence is grasping its meaning. The meaning of a sentence is a function of the meanings of the words of which it is composed, and their arrangement in the sentence. The way in which the individual words combine to give the meaning of the sentence is determined in part by the grammar of the language. So the ability to understand a sentence consists in knowing (a) the meanings of the words of which it is composed, and (b) the grammatical rules according to which they combine to generate the meaning of the whole sentence.

The process of understanding a novel sentence may then be

explained: When a novel sentence is heard, the hearer uses his knowledge of word-meanings and grammar to interpret the sentence, to work out what it means. When a speaker produces a sentence, she uses her knowledge of grammar and word-meanings to form a sentence which will express what she intends to say. So the hearer's interpretation, if all goes well, matches the speaker's intention; the hearer makes an inference to the thought expressed in the speaker's words. The speaker encodes a thought in speech (or writing), and the hearer (or reader) decodes it. In producing a novel sentence, the speaker (or author) makes use of the same knowledge of grammar and word meaning used by the hearer in understanding, in roughly the reverse procedure. In this picture, 'if anyone utters a sentence and means or understands it, he is operating a calculus according to definite rules' (Wittgenstein 1953 #81).

This is in outline Chomsky's theory of the role of knowledge of grammar in use of language. In his review of Skinner's Verbal Behavior, he wrote:

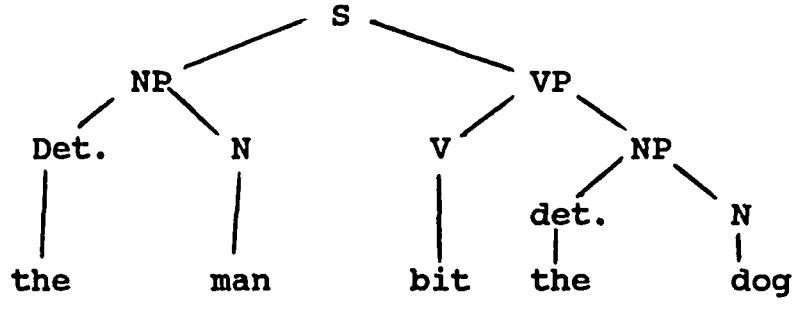
> It appears that we recognize a new item as a sentence not because it matches some familiar item in any simple way, but because it is generated by the grammar that each individual has somehow and in some form internalized. And we understand a new sentence, in part, because we are somehow capable of determining the process by which this sentence is derived in this grammar. (Chomsky 1959:59)

And in Aspects of the Theory of Syntax: 'Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his

language.' (Chomsky 1965:8) He explains that 'by a generative grammar I mean simply a system of rules that in some explicit and well-defined way assigns structural descriptions to sentences.' (1965:8)

The speaker-hearer's knowledge of grammar is, in this view, analogous to the linguist's, and interpreting a sentence is an activity similar to the linguist's in performing a grammatical analysis. Part of understanding a sentence is deriving a "structural description", a representation of its grammatical structure. A structural description of a sentence is the kind of representation to be found in linguistic textbooks: an analysis that makes explicit the grammar of the sentence - the grammatical relations among its constituent words. For example, the sentence, 'The man bit the dog', may receive the syntactic analysis [X]:[1]

[X]



Understanding the sentence therefore means, in part, knowing

[1] I assume only for the purpose of the discussion that this is the correct analysis of the sentence, in the appropriate form. Nothing depends on this assumption. I am concerned with the idea that anything of this kind is part of the ordinary mental life of a person using a language: grammatical analysis of a sentence, in accordance with rules of syntax, semantics, phonetics, etc.

that it has this grammatical structure, which involves "recovering" this structural description, analysing the sentence, in thought, to reveal this underlying grammatical structure. So

> [a] fully adequate grammar must assign to each of an infinite range of sentences a structural description indicating how this sentence is understood by the ideal speaker-hearer. (Chomsky 1965:4-5)

In Knowledge of Language (1986), Chomsky defines the framework of inquiry in this way:

> The three basic questions that arise, then, are these:
> (i) What constitutes knowledge of language?
> (ii) How is knowledge of language acquired?
> (iii) How is knowledge of language put to use?
> The answer to the first question is given by a particular generative grammar, a theory concerned with the state of the mind/brain of the person who knows a particular language. (Chomsky 1986:3)

A generative grammar specifies what a speaker knows by stating the grammatical rules and principles of the language. This is a partial answer to '(i) What constitutes knowledge of language?' Knowledge of grammar is conceived as "mental representation" of these rules. Interpretation (decoding) of a sentence results in "mental representations" of its grammatical structure, which are analyses like [X], in accordance with the rules represented. This is a partial answer (the outline of an answer) to '(iii) How is knowledge of language put to use?' Chomsky holds that 'it is reasonable to suppose that the rules of grammar are mentally represented

and used in thought and behaviour' (1980:129):

> ...I know of no proposed explanation for the fact that our judgements and behaviour accord with certain rule systems other than the assumption that computation involving such rules and the representations they provide takes place in the mind' (1980:130).

The problem for the linguist trying to discover what is known by a person who knows a language is then, according to Chomsky,

> to find the mental representations that underlie the production and perception of speech and the rules that relate these representations to the physical events of speech (1986:41).

This pattern of explanation is commonly assumed by linguists, and by cognitive psychologists. For example, the authors of an introductory textbook in linguistics write:

> What a person knows when he knows a language is how to translate between internal and external representations. (Foss and Hakes:12)
>
> Of course, understanding doesn't "just happen". There is a great deal of evidence which indicates that you have done something quite complex when you understand an utterance. (ibid. 15)

And

> ...each speaker has tacit knowledge of a set of rules for the sentences of that language. The judgment that a particular utterance is or is not a sentence is made by using these tacit rules. (ibid. 17)

In a contribution to a recent anthology called <u>An Invitation to Cognitive Science</u> it is claimed:

> With their grammatical knowledge, human language speakers are able to construct infinite collections

> of well-formed sentences.... Given this creative aspect of syntactic ability, we know that our knowledge of the well-formed sentences of English cannot take the form of a simple list.... On the basis of this we conclude that syntactic knowledge must be encoded within us in the form of a finite set of rules and principles that allow us to generate the sentences of English from smaller, subsentential elements such as words. (Larson 1990:28)

And in the introduction to a collection of philosophical essays on the issue of "mental representation" we read:

> To learn to speak Navajo, one must acquire specific, if tacit, knowledge of the sentence patterns and pronunciation of Navajo. A grammar of Navajo, in that it provides specific rules for the construction of Navajo sentences, represents the kind of knowledge of that language that one must have to speak it. It is such grammars, grammars on the lower level, that we assume will represent the stored knowledge in competence-based models of linguistic performance.' (Bresnan & Kaplan 1982:xvii-xviii)

Knowledge of a generative grammar is thus proposed as a partial explanation of (a) the ability to understand novel sentences (including the ability to recognise sound patterns, ink marks, etc., as sentences in a particular language), (b) the ability to 'produce' novel sentences and (c) the ability to have linguistic intuitions ('judgements') of particular grammatical properties and relations of sentences, such as grammaticality (being grammatical or ungrammatical), and co-reference (eg. of a noun and a pronoun).

These explanations justify the hypothesis of knowledge of grammar. The reason to suppose that speakers have knowledge of generative grammars is the explanation this affords of

their abilities to produce and comprehend novel sentences, and to recognise grammatical properties and relations of sentences and expressions.[2]

Explanation of abilities by reference to knowledge is familiar in philosophically uncontroversial cases. For example, an engineer's ability to design and build a bridge may be explained (in part) by her knowledge of principles of mechanics and mathematical theorems. She uses her knowledge in designing the bridge and directing its construction. Furthermore, this knowledge of principles explains her possession of, or capacity to acquire, knowledge of properties and relations of the bridge, for example that it can withstand such-and-such a load, or that it will oscillate with a certain amplitude and frequency in a particular wind speed (comparable with a speaker's knowledge of particular grammatical properties of sentences).

There is, however, an obvious difference between this case and the explanation of language use and understanding by knowledge of grammar: An engineer is conscious of knowing principles of mechanics and theorems of mathematics; she could state them on request, recognise them and distinguish correct from incorrect formulations (though not infallibly); she makes

---

[2] Fodor (1968b), for example, argues for the thesis of tacit knowledge (of grammar) as the best explanation of ability to use language. Graves et al. argue that explicit knowledge of particular grammatical properties justifies the hypothesis of knowledge of rules of grammar, as opposed to mere 'computational procedures', which would explain the ability to use language.

conscious and explicit use of them in designing a bridge; and so on. An ordinary speaker, by contrast, is not conscious of knowing a generative grammar, is not aware of following rules or applying principles of the kind specified by the grammar, cannot say what any of these rules are, cannot in general confirm (or refute) the linguist's attribution of knowledge, or suggestions of which rules are followed, may not be familiar with concepts employed in the articulation of the grammar (even such concepts as "noun", "verb", and "article"), and may lack the capacity to grasp these concepts or understand any formulation of the grammar. In short, we seem not to know grammar, in the way that the engineer knows mathematics and mechanics. The hypothesis must therefore be qualified:

> Obviously, every speaker of a language has mastered and internalized a generative grammar that expresses his knowledge of his language. This is not to say that he is aware of the rules of the grammar or even that he can become aware of them, or that his statements about his intuitive knowledge of the language are necessarily accurate. Any interesting generative grammar will be dealing, for the most part, with mental processes that are far beyond the level of actual or even potential consciousness; furthermore, it is quite apparent that a speaker's reports and viewpoints about his behavior and his competence may be in error. Thus a generative grammar attempts to specify what the speaker actually knows, not what he may report about his knowledge. (Chomsky 1965:8)

The knowledge of grammar attributed to speakers is therefore said to be <u>tacit</u> or unconscious:

> As I am using the term, knowledge may be unconscious and

> not accessible to consciousness. It may be "implicit" or "tacit". No amount of introspection could tell us that we know, or cognize, or use certain rules or principles of grammar, or that use of language involves mental representations formed by these rules and principles. (Chomsky 1980:128)

And,

> We have (generally tacit) knowledge of the rules of language...(1986:222)

Explicit, conscious knowledge and application of a grammar is of course possible, in principle. This would be demonstrated by a student of linguistics, who might be examined on theoretical knowledge and its application in just the same way as a student of engineering is tested. 'Tacit knowledge' is to be contrasted with this.

The idea that knowledge of a generative grammar underlies, and explains, the ordinary use of language stands in need, as I shall argue, of an account of the nature of the alleged knowledge, and of its - tacit - use in speaking and understanding. Chomsky claims that understanding a sentence involves an act of interpretation, which he explains in terms of 'mental representations'. I will discuss Chomsky's account of knowledge of grammar in the remainder of this chapter. 'But an interpretation is something that is given in signs' (Wittgenstein 1974:47). For this reason, I shall argue in Chapter 2, the account in terms of 'representations' threatens to generate a regress, as these representations themselves would be grammatical objects which must be understood. The

answer to this, suggested by Chomsky, assumed by many in cognitive psychology, and defended by some in the philosophy of mind, is the idea that the brain is like a computer, mental processes are computational, and psychological explanations - including those of psycholinguistics - are essentially concerned with programs of the mind. In Chapters 3 and 4 I will develop arguments against this model of mind. If I am right, and this model is incoherent, then Chomsky's hypothesis of knowledge of grammar provides no explanation of the ability to speak and understand a language, at least on the 'scientific' assumptions he shares with many theorists in cognitive science.

My arguments are derived from Wittgenstein's later work, especially the Philosophical Investigations. But I do not present them as an explicit interpretation of Wittgenstein, or as claims that Wittgenstein argued such-and-such or so-and-so. It is, I think, generally accepted that the 'rule-following' passages of the Investigations present a cogent critique of phenomenalistic or introspective theories in the philosophy of mind. Thus an imagined colour sample is no more useful in explanation of the meaning of colour-words than a physical sample. No one any longer defends such theories. But it is widely believed that his arguments - if indeed there are arguments in Wittgenstein's late writings - are beside the point for scientific, physicalist theories in philosophical psychology. As far as these are concerned, Wittgenstein

either overlooked their possibility, or dismissed them without argument.[3] This essay is an attempt to extend Wittgenstein's criticisms to these currently fashionable types of explanation, in particular as they offer an account of the nature of knowledge of language, in terms of complex grammars.

1.2 COMPETENCE AND PERFORMANCE

Chomsky distinguishes sharply between knowledge of grammar, which he calls competence, and the application of this knowledge in using and understanding language, which he calls performance:

> We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations)....The problem for the linguist, as well as for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior. (Chomsky 1965:4)

Performance, then, means 'the actual use of language in concrete situations', acts of speaking, hearing, reading and writing, and so on. This is, on the face of it, the everyday, common-sense concept of use of language.

Competence is defined as 'the speaker-hearer's knowledge of his language'. Chomsky then tells us that 'A grammar of

---

[3]See eg. Fodor 1975, McGinn 1984

a language purports to be a description of the ideal speaker-hearer's intrinsic competence' (1965:4). Thus (grammatical) competence is what is described by grammar, the analytical description of a language developed by a grammarian.

In Chomsky's view, competence is quite independent of performance. That is to say, someone could possess knowledge of grammar without ever putting it to any use in speaking, understanding what others say, or judging that particular sentences have certain grammatical properties or relations:

> If, as I am now assuming, to know a language is to be in a certain mental state comprised of a structure of rules and principles (comparably, for certain other aspects of cognition), then in theory one could know a language without having the capacity to use it... (1980:51).

Knowledge of a language, in Chomsky's view, is not to be understood as any kind of ability to use it, 'even in thought' (1980:52). In particular, he rejects the suggestion that knowledge of grammar is a kind of "know how", the ability to speak grammatically, for example.

> The term "competence" entered the technical literature[4] in an effort to avoid entanglement with the slew of problems relating to "knowledge", but it is misleading in that it suggests "ability" - an association that I would like to sever...(1980:59).
>
> Knowledge of language is often characterised as a practical ability to speak and understand, so that questions (i) [What constitutes knowledge of language?] and (iii) [How is knowledge of language put to use?] are closely related, perhaps identified. Ordinary usage makes a much sharper distinction between the two questions, and is right to do so. Two people may share exactly the same

---

[4] In Chomsky's Aspects of Syntax 1965.

> knowledge of language but differ markedly in their ability to put this knowledge to use. Ability to use language may improve or decline without any change in knowledge. This ability may also be impaired, selectively or in general, with no loss of knowledge, a fact that would become clear if injury leading to impairment recedes and lost ability is recovered. Many such considerations support the commonsense assumption that knowledge cannot be properly described as a practical ability. (1986:9)

So a theory of competence is not a theory of performance. To say what someone knows in knowing a language is not to say how they use this knowledge in speaking, etc. To say <u>that</u> a person knows a language, or a grammar, in Chomsky's view, is not to say anything about their use of language or their ability to use it; not even that they possess any such ability.

Nevertheless, 'a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker-hearer's knowledge of the language' (1965:9). Thus a theory of performance will 'incorporate' a theory of competence:

> ...competence, the system of rules and principles that we assume have, in some manner, been internally represented by the person who knows a language and that enable the speaker, in principle, to understand an arbitrary sentence and to produce a sentence expressing his thought... (1980:201)

> ...the goal of the investigator will be to determine the nature of the competence system that expresses what it is that the mature speaker knows, and to develop process models that show how this knowledge is put to use. (1980:203)

> Theories of grammatical and pragmatic competence must find their place in a theory of performance that takes into account the structure of memory,

> our mode of organizing experience, and so on. (Chomsky 1980:225)

This follows from the view that knowledge of grammar is a partial explanation of the abilities to use and understand a language. Although 'a generative grammar is not a model for a speaker or a hearer' (1965:9), explanation of how use and understanding is achieved will include explanation of how this knowledge is put to use. This will only be part of the explanation, because other kinds of knowledge and ability will also be exercised. For example, taking "grammar" to refer only to syntax, knowledge of the meanings of words (semantics) will also be employed; so will knowledge of circumstances, of the identity of the speaker, of the previous course of the conversation, and many other factors relevant to understanding what a speaker's (or author's) utterance means. In Chomsky's terminology, "pragmatic competence" is involved in the use of language, as well as "grammatical competence":

> For purposes of inquiry and exposition, we may proceed to distinguish "grammatical competence" from "pragmatic competence", restricting the first to the knowledge of form and meaning and the second to knowledge of conditions and manner of appropriate use, in conformity with various purposes. Thus we may think of language as an instrument that can be put to use. (1980:224)[5]

---

[5] This distinction is also a commonplace in the literature. For example: 'Although our linguistic competence lies at the heart of our knowledge of language, it is clear that we know more than just a grammar. A grammar specifies the rules we know, but it does not state how we make use of that knowledge...A theory of the additional knowledge is a theory of linguistic performance.' (Foss & Hakes:18)

"Grammatical competence", knowledge of grammar, explains the ability to make utterances grammatical, and to grasp the grammar of others' utterances; "pragmatic competence" refers to other aspects of knowledge, and perhaps ability, which allow us to produce sentences relevant to the situations in which we utter them; and which enters into the ability to make sense of another's words. So for example, if I say 'That's an error', talking about a move in a game, my grammatical competence explains my ability to put words together in a grammatical, hence in this respect meaningful, sentence. But my knowledge of the rules of the game, of the state of play, of the skill of the player, of the person to whom I address the remark, of social etiquette, and more, may all enter into my ability to make an acceptable, relevant, plausible, perhaps true, utterance. "Pragmatic competence" covers all of this second category of knowledge, understanding and ability (or at least that part of it which can be described as "knowledge"). Similarly, such factors enter into understanding an utterance.

Furthermore, Chomsky recognises a dimension of human behaviour which we may altogether lack the ability to explain. This is roughly the element of free choice in our utterances and other behaviour: how we decide what to say, or do. This may be what Chomsky calls a "mystery", ie. something beyond the capacity of humans to understand, due to limitations inherent in the human mind. So we should not expect a complete explanation of a person's verbal behaviour, if this

means to explain why the person said a particular thing rather than something else, or nothing at all, as well as why the utterance had the syntactic form it had, etc.

But knowledge of grammar is one factor in the production and interpretation of sentences, in Chomsky's model, and in this sense is assumed to be a partial explanation of performance. The case of the engineer provides an analogy. Her knowledge of principles of mechanics and theorems of mathematics is only one factor in her ability to design and supervise the construction of a bridge. She will also use her knowledge of the properties of materials, of environmental and economic circumstances, her ability to draw, to communicate with others (including her ability to use language), and much more. But her "mechanical competence", as we might call this part of her knowledge, is one factor in her overall ability. A "performance model" of the engineer would include a "competence model", a description of her knowledge of mechanics, and explain how she puts this knowledge to use.

The distinction of "grammatical competence" and "pragmatic competence" points up just what aspect of producing or understanding a sentence knowledge of grammar should explain. In the case of comprehension, it is just the grasp of the grammatical relations among the words and expressions of the sentence, which taken together constitute its grammatical structure. In the narrower sense of "grammar", in which it means "syntax", it is the totality of relations such

as "subject of the verb", "object of the verb", "coreferential with", "modifies" (as adjective to noun), etc.[6]

There is another sense in which the theory of "competence" is not a theory of "performance", in Chomsky's view. The generative grammar of a language, as specified by a linguist, according to Chomsky, is a theory of the speaker's competence, of what the speaker knows in knowing the language. This may be specified in terms of rules which 'assign a structural description' to any sentence of the language. That is, applied to a sentence, these rules may be used to derive such an analysis of the sentence. The linguist or student of linguistics may perform a derivation by applying these rules to a sentence. Nevertheless, this method of derivation is not necessarily the same as that actually employed by someone "interpreting" the sentence in ordinary use of language. The "parsing procedure" employed by a speaker-hearer is not necessarily the parsing procedure demonstrated in a textbook of linguistics. The order of applying the rules, or the method of derivation, might be different. This might be compared with different ways of performing multiplications, for example the difference between methods (a) and (b), which could be described as two different applications of the same

---

[6] It can be argued that semantics and syntax are not independent of each other, for example that being a noun is an aspect of the meaning of eg."tree", so that the syntactic relation of verb and object in "the dog climbed the tree" is not independent of the semantics of the words "dog" and "tree" etc. I don't mean to assume any position on such questions.

rule:

```
(a)        18            (b)       126 (x1)
      x   126                    + 126 (x2)
         1800                      252
          360                    + 126 (x3)
          108                      378
        =2268                    + 126 (x4)
                                    .
                                    .
                                  2142
                                 + 126 (x18)
                                 =2268
```

The parsing procedure used by the speaker-hearer is to be found by empirical study of "real-time" language processing, in Chomsky's view, which might require evidence from experiments in psychology laboratories, or in principle from neurophysiological investigation. These would contribute to a "performance model", which would describe the way a person "interprets" or "produces" a sentence.

The theory of competence does, however, state 'what the speaker-hearer knows' about a language, including knowledge of the underlying structure of sentences, the structural descriptions appropriate to particular sentences uttered or perceived. Thus:

> What we are suggesting is that the notion of "understanding a sentence" be explained in part in terms of the notion of "linguistic level". To understand a sentence, then, it is first necessary to reconstruct its analysis on each linguistic level; and we can test the adequacy of a given set of abstract linguistic levels by asking whether or not grammars formulated in terms of these levels enable us to provide a satisfactory analysis of the notion of "understanding"....In general, we cannot understand any sentence fully unless we know at least how it is analyzed on all levels... (Chomsky 1957:87)

This implies that the speaker knows the rules or principles specified in the linguist's grammar; and the result of the speaker's application of these rules in use of language is (something like) the linguist's structural descriptions of sentences; but the method of derivation, the particular way the rules are applied to a perceived sentence, for example, may not be the same as that followed by a linguist demonstrating a derivation on a chalk-board or in a textbook. The steps taken, or the order of application of the rules, may not be the same. (There is also a debate in the literature about the independence of syntactic analysis: whether, in "processing" a sentence, we make use of semantic or pragmatic knowledge in deriving a syntactic analysis of a sentence, eg. in determining the interpretation of a "garden path" sentence, such as 'The horse raced past the barn fell'.[7])

However these questions are decided, the assumption is that a process of syntactic analysis by the application of rules occurs in the mind of the hearer (or speaker), in a way analogous to the linguist's method of analysis. Something of the same kind is assumed to occur in the speaker-hearer's mind. The speaker-hearer undertakes some such derivation of structural descriptions by the application of rules to perceived sentences; and carries out something like the reverse of this procedure in producing a sentence. If this is not so, it makes no sense to claim that "competence",

---

[7] See eg. Garrett, in Osherson & Lasnik eds.

described as knowledge of a generative grammar, explains "performance", described, in part, as knowledge of the grammatical structures of sentences perceived, or knowledge of their structural descriptions.

I will not be concerned with the particular content of theories of either competence or performance. My interest is in the idea that this kind of explanation can be given. Questions of principle about this hypothesis would not be affected by the supposition that performance actually involves procedures different from those employed by a linguist analysing a sentence, as long as these procedures are taken to be the same kinds of operation: rule-governed derivations. And this is a fundamental assumption of Chomsky's theory, and of cognitive theories of language in general. This is the form to be taken by explanations of the use of "competence" in "performance".

The hypothesis of "competence", then, is dependent on the theory of its use in "performance" in the following way. The hypothesis is justified by the claim that it (partly) explains the ability to use and understand language, and to attain explicit knowledge of properties of sentences, etc. It must be intelligible that the hypothesised knowledge of grammar can be used in the production and comprehension of sentences, or in the attainment of explicit grammatical knowledge. The hypothesis of such knowledge must <u>be</u> a possible explanation of the use and understanding of language, in the relevant

respect. So despite Chomsky's contention that question (i) What constitutes knowledge of language? is to be sharply distinguished from question (iii) How is knowledge of language put to use? (Chomsky 1986:3), the answer to (i), namely that it consists, in part, in knowledge of the rules of grammar, is closely connected to the answer to (iii). If "knowledge of grammar", as characterised in the theory of "competence", does not allow an answer to (iii) which amounts to an explanation of how we are able to speak grammatically, understand the grammar of sentences, and come to know grammatical properties and relations of sentences, then there is no basis for the hypothesis of "competence", in Chomsky's sense.

It is natural to assume that knowledge of grammar would (partly) explain use and understanding, and explicit grammatical knowledge, because of the apparent similarity between this explanation and familiar explanations such as that of the engineer's abilities. In the next chapter I will challenge this assumption.

1.3 TACIT KNOWLEDGE

The idea of tacit knowledge of grammar, knowledge of complex, abstract, grammatical rules and principles attributed to every person who speaks a language, has been found problematic by some philosophers. Two aspects of the idea together make it problematic: inaccessibility to consciousness, and the abstract nature of the proposed

knowledge. The hypothesis is that we possess knowledge of a general, abstract and complex nature, and that this explains particular and relatively simple explicit knowledge, as well as use and understanding of language.

The basis of the objections is that ordinary competence in a language, the ordinary ability to communicate, does not seem to involve such knowledge in any ordinary sense. We are not aware of knowing a generative grammar, or conscious of using the knowledge; we cannot say what the rules are, in general, or confirm (or refute) the linguist's hypotheses; a speaker may not possess the concepts used in stating the rules, and so may be unable to understand them; and some people who can speak a language may lack the capacity for this understanding altogether. This is so even in the case of such a simple rule of traditional grammar as 'adjectives agree with the nouns they modify in gender, number and case'. A native speaker of a language for which this rule holds, whose use of the language conforms to the rule, may not understand it, and may be unable to do so. These are just the reasons that the knowledge attributed to the speaker is described as "tacit" or "unconscious".[8] They are also the reasons that the idea of

---

[8] It seems quite possible for a language to lack the vocabulary to refer to verbs, nouns, etc., so that no speaker of that language could express or understand its rules, even at this elementary level. It is plausible that grammatical vocabulary and the articulation of grammar, as a theory of a language, only develop when the culture takes a certain linguistically reflective turn, perhaps with literacy. The kind of analysis proposed in generative grammar is a sophisticated, complex use of language, dependent on a considerable degree of education and training for

such "knowledge" can seem implausible or conceptually puzzling. Searle expressed this point of view:

> One of the chief difficulties of Chomsky's theory is that no clear and precise answer has ever been given to the question of how the grammarian's account of the construction of sentences is supposed to represent the speaker's ability to speak and understand sentences, and in precisely what sense of "know" the speaker is supposed to know the grammar. (Searle 1972:11)

In Chomsky's view, as we have seen, the grammarian's account is <u>not</u> supposed to 'represent the speaker's ability to speak and understand' (it is a theory of "competence", not a theory of "performance"); but the second half of Searle's challenge requires an answer: in what sense of "know" does a speaker know a generative grammar?

Chomsky alleges that the ordinary notion of knowledge is not very clear, and it is unimportant if the concept of "tacit knowledge" is not the same as the ordinary concept:

> One might ask whether it is proper to use the ordinary language term "knowledge" in this connection. Is it, for example, proper to say that a person who knows a language in the ordinary sense "knows the rules of the language" (the I-language) in the technical sense? In part, the answer is certainly negative, because I-language, like other technical notions of scientific approaches, is not language in the pretheoretic sense, for reasons discussed earlier. It is not clear that much is at stake here; the intuitive concept of knowledge becomes hazy and perhaps misleading at certain crucial points, and ordinary usage in fact differs from language to language; one does not speak of "knowing a language" but rather of "speaking" or

---

its <u>conscious</u> or explicit exercise.

> "understanding" it in languages very similar to English, although this does not affect our concern to discover the cognitive system - whether we call it "knowledge of language" or something else - that enters into our particular knowledge of facts [about the grammatical structures of particular sentences]. (1986:265).

But if "tacit knowledge" is not knowledge, in the sense in which the engineer has knowledge - or in some similar sense - then it's not clear that any explanation of abilities and explicit grammatical knowledge is given by the hypothesis, in which case the reason to think people have such "tacit knowledge" is undermined. Chomsky tacitly acknowledges this when he introduces terminology intended to replace talk of "knowing" grammar, but explains the new terminology in terms of "knowledge". For example:

> To avoid terminological confusion, let me introduce a technical term devised for the purpose [to explain knowledge of language], namely "cognize", with the following properties. The particular thing we know, we also cognize.... Furthermore, we cognize the system of mentally-represented rules from which the facts follow. That is, we cognize the grammar that constitutes the current state of our language faculty and the rules of this system as well as the principles that govern their operation.....
>
> In fact, I don't think that "cognize" is very far from "know" where the latter term is moderately clear, but this seems to me a relatively minor issue....Thus "cognizing" is tacit or implicit knowledge, a concept that seems to me unobjectionable.....I will return to the terms "know" and "knowledge", but now using them in the sense of "cognize" - that is, admitting both conscious and tacit knowledge - and hoping that possible confusion will have been allayed. The fundamental cognitive relation is knowing a grammar; knowing the language determined by it is derivative. (Chomsky 1980:69-70)

This obviously leaves unanswered the question whether "tacit knowledge" is any kind of knowledge. If it is not, then "cognize" has been given an equivocal use, and the relation of a speaker to grammar, hence the role of this in explaining use of language etc., has yet to be explained. Similar remarks apply to Chomsky's use of the expressions "competence" and "I-language" (Chomsky:1986).

Furthermore, the argument for the hypothesis of knowledge of grammar derives from the explanation it is supposed to afford of a speaker's ability to use and have intuitions about language in the pretheoretic sense. What we want to know is whether knowledge of grammar (partly) explains the ability to use language, where "knowledge" and "language" are the ordinary concepts. The whole interest of the theory derives from its claims about the ordinary ability to speak a language, ie. what is meant by the ordinary word "language". And the basic explanatory hypothesis depends on the ordinary concept of "knowledge", as exemplified by the case of the engineer. If the theory is that a speaker *knows a *language, where "*knows" and "*language" are technical terms with meanings different from the meanings of "knows" and "language", then the question of how a speaker is able to use and understand language has not been answered.

On this question, some discussion of Chomsky has referred

to Ryle's distinction of knowing how and knowing that.[9] It is clear that Chomsky denies that the basic relation to language is a case of knowing how. The concept of knowing how is just that of possessing an ability, being able to do something, and we have seen that Chomsky is opposed to this way of thinking about language. But the question is vexed, whether "knowledge of language" is a species of "knowing that". Chomsky attributes to the speaker "knowledge of" the language and its grammar (primarily the latter), "mastery of" rules and principles of grammar; we are said to "know" the rules that constitute the grammar, to have "internalised" the grammar. He acknowledges that we do not "believe" the rules of grammar, and that we may have no "justification" for them: hence the classic modern analysis of knowledge, as justified true belief, is inapplicable.[10] Furthermore, of course, we are typically unable to report the rules, or even recognize them as the rules we know and follow. This is part of what makes this knowledge "tacit". Nevertheless, he says,

> I will continue to use the term "know" in the sense of "cognize", in the belief that we thus move towards an appropriate concept for the study of the nature and origins of what all agree to be knowledge that or knowledge of. It seems doubtful that there is a useful sense of such notions as "justification", "grounding", "reasons", etc., in which justification, having good reasons, etc., is in general a necessary condition for knowledge (including knowing that) - nor a sufficient

---

[9] See for example Harman (1967): 'Psychological Aspects of the Theory of Syntax'.

[10] See Chomsky (1980) Rules and Representations p93-4.

> condition for true belief to be knowledge, as the Gettier examples show. Rather, having knowledge should be analyzed at least in part in quite different terms, in terms of possession of certain mental structures, I believe. (Chomsky 1980:99)

Chomsky rejects the analysis of knowledge into Ryle's two categories, and suggests that knowledge of rules, for example, may be a distinct kind of knowledge. So 'A knows the rule R' is distinguished from 'A knows how to act in accord with R' (or 'A is able to...'), and also from 'A knows that the rule (eg. of passive-formation) is R'. 'A knows the rule R', in this view, describes a state of knowledge of A which <u>explains</u> A's ability to form passive sentences. On the other hand, 'A knows the rule R' does not attribute the knowledge <u>that</u> the rule of passive-formation is such-and-such. The latter knowledge would presumably be the kind of knowledge of the rule possessed by a linguist, and so would meet the conditions of ability to state the rule, comprehend a statement of it, and so on.

It seems to me, however, that the forms of expression 'A knows the rule...' and 'A knows that the rule is...' are more or less interchangeable, even when the person cannot state the rule. For example, if a child in kindergarten acts in a way that would make us say 'He knows the hand-raising rule' or 'He knows the rule that you must raise your hand before you speak', we could just as well say 'He knows that the rule is that you must raise your hand before you speak', or 'He knows that raising your hand before you speak is a rule of this

classroom'. This suggests that "tacit knowledge" of language should be assimilated to ordinary cases of knowing that, and contrasted with knowing how to do something - or just doing it, eg. using language one way rather than another, in accord with one rule rather than another. Tacit knowledge is a kind of propositional knowledge, or is something akin or analogous to propositional knowledge. This interpretation is reinforced by the theory that tacit knowledge of rules is the basis of a kind of inferential processing or derivation of structural descriptions, in the mental activity underlying use of language. The idea of representation of rules also suggests assimilation of tacit knowledge to propositional knowledge ('knowledge that').

Stich summarised the objections to the hypothesis of tacit knowledge of grammar ('What Every Speaker Knows'):

> Are there any striking dissimilarities with unproblematic cases of knowledge? Clearly there are many. Commonly when a person knows that p he has occasionally reflected that p or has been aware that p; he will, if inclined to be truthful and otherwise psychologically normal, assert that p if asked. More basic still, he is capable of understanding some statement which expresses what he knows. Yet for the propositions of linguistic theory none of this need be true. (Stich 1971:485-6)

Nagel (1968) proposed a somewhat weaker condition, a condition of 'recognisability' of the attributed knowledge. He compared the hypothesis of tacit knowledge of rules of grammar with Freud's theory of the unconscious. He claimed that the attribution of unconscious knowledge (or belief, motivation,

etc.) in psychoanalysis depends on a 'condition of recognisability' (Nagel 223): the possibility, in principle, of the person to whom the knowledge is attributed recognising the correctness of the attribution, when suitably formulated and presented, 'from the inside'. The subject must be able to agree, 'Yes, that's what I believe', not on the basis of evidence which another person might have, but simply because it is his or her belief (knowledge, motive, etc.). Nagel suggests that some grammatical knowledge may meet this condition, for example knowledge of the rule of English that the plurals of nouns are formed by adding "s", with certain exceptions; but that it would not be met by the more abstract rules or principles of linguistic theory.

Are these necessary conditions of knowing something? Graves et al., replying to Stich, propose a counter-example :

> For instance, someone who knows that his breakfast is on the table downstairs may not be able to understand any sentence that expresses what he knows because he is suffering from aphasia for written language and speech. And an animal, such as a dog, can know where its meal is located even though it is unable to understand any sentence expressing what it knows. (Graves, et al 1973:321.)

A dog knows, we might say, that breakfast is in the kitchen, although of course it's able neither to <u>say</u> so, nor to understand and agree if we say that it knows. A dog cannot say what it knows, or be brought to recognise that the attribution is correct, or understand any statement of the attributed knowledge. So the ability to state or understand

some expression of what's known is not a necessary condition of knowledge, in general.

But there is a difference between the case of a dog - or a pre-verbal infant - and the hypothesis of tacit knowledge of generative grammar. We don't attribute to a dog complex, abstract knowledge. The knowledge attributed to a dog is simple and concrete, whereas a generative grammar consists of highly complex and abstract rules. Of course the behaviour to be explained is much more complex in one case than the other. In philosophically uncontroversial cases of knowledge of something as abstract and complex as the rules of a generative grammar, however, the verbal conditions hold: the person who has the knowledge is normally capable, at least, of understanding some expression of it, and possesses the concepts with which it is expressed. The engineer, for example, grasps the concepts employed in the principles of mechanics and mathematics she knows and uses; she can understand expressions of these principles, can probably state them herself, and make explicit use of them. Similarly, the student of linguistics who knows the rules of grammar is able to understand and use the concepts employed in their articulation. This, of course, is just why knowledge of grammar is described as "tacit" or "unconscious", so it won't suffice to refute the hypothesis. But it shows that the hypothesis of knowledge of grammar is significantly different from attributions of knowledge which are (a) comparable in the

complexity and theoretical nature of what is known, but (b) not philosophically controversial. The counter-examples proposed by Graves et al. do not answer this objection, because they do not meet condition (a). That is to say, it is the intellectualism[11] of the hypothesis of competence that makes the tacit knowledge thesis problematic.

The aphasic is not a counter-example for a similar reason. We wouldn't say such a person had complex, abstract knowledge he had never been able to express. We might attribute knowledge of set theory, for example, if we knew that he had possessed this knowledge in the normal way, ie.had been able to express it, understand its expression, and so on, before suffering from aphasia. But surely not if he'd never had such abilities.

What would we say about a person who was, so to speak, aphasic for life? There are people who achieve little or no verbal expression or comprehension. We attribute beliefs, and knowledge, to them; certainly far more than to animals. The difference between such individuals and non-human creatures is, among other things, that their behaviour is likely to be much more complex; and the belief we attribute will be appropriately complex. So we can easily imagine what kind of behaviour would lead us to attribute to an individual, for example, the belief that the bus is about to arrive, or that

---

[11] I discuss the "intellectualism" of the theory below (pp35-36, 48-50).

there are cookies in the jar - gesture will play a part, which is neither a'natural sign', like clouds on the horizon or tears on the face, nor genuinely language, though there will be intermediate cases. We could imagine behaviour that would express the intention to leave home, or the belief that a vacation is imminent; even the belief that the date of departure is three days off (if, for example, he puts aside three sets of clothing and packs the rest). We would also say that such a person knows breakfast is ready, if he shows that he distinguishes this meal from others, for example by setting out cereals in the morning and making salad in the evening.

There would be limits, determined by the expressive limitations of non-verbal behaviour. We could attribute the belief that the milk is bad; but not that it has been sitting out since the day before yesterday. That it is cold outside, but not that it would be cold if there had not been a heat wave. We would not, in general, attribute a belief or intention the individual could not express. Counterfactual beliefs, for example, do not seem attributable to a person who cannot give them verbal expression.

The thesis of tacit knowledge of grammar attributes to speakers complex, abstract knowledge, from which deductions or inferences are made. Thus Chomsky claims that

> it seems reasonably clear, both in principle and in many specific cases, how unconscious knowledge issues in conscious knowledge ...a person has unconscious knowledge of the principles of binding theory, and from these and other discussed, it follows by computations similar to straight

> deduction that in (9i) [I wonder who the men expected to see them] the pronoun them may be referentially dependent on the men whereas in (9ii) [the men expected to see them] it may not....It does not seem problematic to entertain the hypothesis that the mechanisms of mind permit something akin to deduction as part of their computational character. (1986:270)

In so far as we attribute knowledge to dogs and non- or pre-verbal humans, it is not complex or abstract, and in particular not such as to be the basis of rational inference or quasi-logical derivation. Knowledge of set theory or logic, or such beliefs as that the circle can't be squared, that electrons are negatively charged, that the objects of art are abstract, that existence precedes essence (or vice versa), or that all humans (or dogs) are mortal - these are not normally attributed in the absence of the ability to express them verbally, or at least to understand their expression.

Explanation of use of language and of linguistic intuitions, in Chomsky's model, implies that a kind of deduction or inference from the abstract rules of grammar takes place in the speaker's or hearer's mind. A general rule of sentence-construction, for example, is applied to a particular case, in the way that a general rule of deduction is applied by a student working out a problem in logic. The ability demonstrated in the particular case is supposed to be explained by knowledge of the general rule. Knowledge from which an inference can be made in this way is

characteristically propositional knowledge, "knowledge that". Dogs and preverbal infants are not normally said to know anything at all abstract or complex, nor to make inferences or deductions from general knowledge or belief. But these are marked characteristics of the knowledge attributed in the theory of tacit knowledge of grammar. It does not seem problematic, or contrary to normal usage, however, to say that someone knows a simple rule of language. For example, a Spanish speaker knows that the stress is on the penultimate syllable of words that end in a vowel or -n or -s, otherwise on the final syllable (with certain exceptions). This is unproblematic in the case of a student of Spanish who has memorised this rule, and pronounces Spanish words accordingly. It seems acceptable, also, in the case of a native speaker who has never learned it or consciously thought of it, but who speaks in accord with it. It would be one way of expressing the difference between the native speaker and the learner who does not yet stress words correctly.

It seems likely that anyone who can speak the language can, in principle, be brought to understand this rule and recognise it as one they follow. But as more abstract, technical rules are in question, it becomes less clear that ordinary usage warrants the attribution of knowledge of the rule to the ordinary speaker. In particular, as the expression of the rule goes beyond the comprehension of an ordinary speaker not trained in linguistics, the ordinary

concept of knowledge is increasingly strained. As Chomsky acknowledges, 'there is something strange about' the statement that a speaker knows 'the principle that pronominals cannot c-command their antecedents' (1986:268).

Chomsky asserts that the difference between 'the familiarity of the notions verb, adverb, and object, which enter into' an elementary rule we would not hesitate to attribute to a speaker, and 'the unfamiliarity of the notions Case assignment and adjacency parameter, which enter into' a more general rule of 'strict adjacency', which the grammatical theory of English attributes to the speaker, 'is irrelevant to [a speaker's] knowledge: These states [ie. of knowledge] are what they are, independently of our knowledge of linguistic theory.' (1986:266-7)

This suggests that whether or not a speaker can be said to know a rule is independent of 'our' knowledge of a theory in which the terms of the rule are used - our familiarity with the notions that enter into the rule. It's of course trivial that we can't attribute the knowledge if we're not familiar with the terms in which it must be expressed; and it's also trivial that it makes no difference to what the speaker knows, whether others are familiar with the notions, and therefore whether able to make the attribution. But the question is really whether the <u>speaker's</u> knowledge of linguistic theory is relevant to the attribution to him of knowledge of the rule. Can we say that he knows the rule, if he has no understanding,

and no prospect of understanding, the notions that enter into it? Chomsky's theory must attribute understanding of the concepts - at some level.

No doubt there is some sense in which a speaker knows all the grammatical rules that his use of the language respects. Perhaps the truth of this is merely that he is able to - or does - speak accordingly; or that he does so and discriminates grammatical and ungrammatical ('correct' and 'incorrect') utterances. What is problematic, and really at issue, is whether he possesses something akin to propositional knowledge, which could be the basis of derivations, or computations, giving rise to 'representations' of the grammatical structure of sentences. I think that the considerations I have adduced here, indicating the anomolous character of attributions of knowledge of grammar, in the generative linguist's sense, to ordinary speakers, demonstrate the need for a theory, or some philosophical account, of the nature of this alleged knowledge: What is meant by saying that speakers know a generative grammar?

The problem can be put like this: If we say that the speaker knows a rule, though he can't state it and can't be brought to recognise that it's a rule he knows and follows, or even to understand it, and if ordinary usage would not warrant the attribution, what exactly is being said of the speaker? The real issue is whether there is an underlying 'mental state' or 'mental structure' or 'representation', which may be

referred to as 'knowledge of a rule', and which plays a role in some kind of inferential or deductive or computational process underlying ordinary use of language. This is the basic question concerning the knowledge of elementary rules of grammar which ordinary usage would condone, as well as the hypothetical "knowledge" of abstract principles of Universal Grammar.

Wittgenstein gives examples of things one can know and not be able to say:

> Compare knowing and saying:
> - how high Mt Blanc is
> - how the word "game" is used
> -how a clarinet sounds.
>
> (1953:#78)

The ability to say how high Mt Blanc is may be a necessary condition of knowing it (unless exceptional conditions obtain); at least, it is knowledge that characteristically can be expressed by someone who possesses it. Knowing how a clarinet sounds certainly does not imply the ability to say how it sounds. Knowing how the word "game" is used is the problematic kind of case. We would normally attribute this knowledge to someone who uses the word correctly. Is that what the knowledge consists in? If the case is analogous to knowledge of grammar, the Chomskyan picture would be that underlying, and explaining, the ability to use the word is a state of knowledge of how the word is used: something like a formulated description of the use of the word, general rules from which particular uses may be inferred; and use involves

such inference. (Chomsky's theory is of course not a theory of word-use in this sense. The case is analogous to Chomsky's theory of competence, not an instance of it.)

If we say that someone knows a rule of language, are we saying that they are in possession of something like a formulation of that rule, in some part of their mind inaccessible to conscious thought in any direct way? Does use of language involve an inner activity of applying these rules to derive structural descriptions, and so on? This is the conceptual issue, which cannot be decided merely by consideration of ordinary usage. But this consideration does show that the attribution of knowledge of the kind proposed in Chomsky's theory is anomalous, and hence problematic. It can't just be said that knowledge of abstract grammatical rules is essentially like unproblematic knowledge. This claim stands in need of justification by a theory of the underlying state common to both, if there is any.

Tacit knowledge of grammar, in Chomsky's sense, thus remains anomalous, when compared with ordinary attributions of knowledge. Comparably complex and abstract knowledge is ordinarily attributed only to a person who is able to say what is known, or at least to understand and acknowledge some expression of it. It cannot merely be said that it is knowledge like any other knowledge. This peculiarity of the tacit knowledge hypothesis creates the need for some cogent explanation of what such knowledge consists in: a <u>theory</u> of

tacit knowledge.

1.4 THE THESIS OF COMPETENCE AND CHOMSKY'S CONCEPTUAL SCHEME

Chomsky's fundamental idea, that anyone who speaks a language possesses knowledge of a generative grammar of that language, and that this explains, in part, the ability to use and understand novel sentences in the language, I will call the THESIS OF COMPETENCE. Since we don't seem to possess such knowledge in an ordinary sense (as the engineer knows principles of mechanics, or as the student of linguistics has knowledge of grammar), it is appropriate to ask what is meant by the hypothesis: What does such knowledge, and its use, consist in? Chomsky proposes a conceptual scheme for answering this question, which is widely accepted in cognitive science. It may be characterised by the theses of scientism, individualism, subjectivism, mentalism, intellectualism, representationalism, computationalism, and physicalism.

Scientism: Knowledge of grammar, hence this aspect of the ability to use and understand a language, can be explained in theoretical terms. A systematic theory can explain 'what knowledge of grammar consists in'. Two theses can be distinguished here: (i) There can be a science of linguistics, which develops theories of meaning, or at least of the contribution of grammar to sentence meaning. (ii) There can be a scientific psychology, which develops theories of

"competence". And, in Chomsky's view, these are one and the same, as he takes linguistics to be a branch of psychology. According to this view, it is a theoretical hypothesis, justified by its explanatory value, that a person who speaks a language has tacit knowledge of its generative grammar. While these views are widely held, they have been denied, for example by Wittgenstein (1953).

Individualism: What grammar (or language) a person knows and uses is not to be explained by reference to others who speak the language. Language and grammar are not essentially social phenomena, in Chomsky's view. A community's speaking a language is to be reduced to, or explained in terms of, the individual members' each speaking the same (or a similar) language:

> I am using the term "language" to refer to an individual phenomenon, a system represented in the mind/brain of a particular individual. If we could investigate in sufficient detail, we would find that no two individuals share exactly the same language in this sense, even identical twins who grow up in the same social environment. Two individuals can communicate to the extent that their languages are sufficiently similar. (Chomsky 1988:36)

The argument implicit in this passage is not valid. No two baseball players play exactly alike; if we could investigate in sufficient detail (which is done in this case), we would find that no two individuals share exactly the same game, in this sense. But it does not follow that they do not play the same game. Even if an individual deviates from the rules, or

believes something to be a rule which is not, or vice versa, this does not imply that she plays a different game, in the ordinary sense. Of course, we could decide to individuate games this way, ie. to use "game" such that no two individuals play the same game, only similar games. But this would be a redefinition of "game". In the ordinary sense, all players play the same game, although all play it with individual variations. But in Chomsky's view,

> Speakers of what is loosely called English do not have partial knowledge of some English superlanguage, but rather have knowledge of systems that are similar but in part conflict. (1980:118)
>
> I have argued that the grammar represented in the mind is a "real object", indeed that a person's language should be defined in terms of this grammar, and that the vague everyday notion of language, if one wants to try to reconstruct it for some purpose, should be explained in terms of the real systems represented in the minds of individuals and similarities among these. (1980:120)

There may be a causal explanation of an individual's knowledge of a certain language or grammar in terms of the community: those who grow up among English speakers, for example, come to speak English because of the social environment:

> In the case of human language, there evidently is a shaping effect; people speak different languages which reflect differences in their verbal environment. (1980:33)

But the answer to the conceptual question, 'What constitutes knowledge of grammar?', will make no essential

reference to other speakers. Thus the primary form of language, for Chomsky, is an idiolect, the language of an individual. A dialect is a collection of similar idiolects; a "language" in the sense in which English, Japanese, etc., are languages, is at best a larger collection of idiolects, but may not be a (scientifically) coherent concept at all, according to Chomsky:

> The term "language" as used in ordinary discourse involves obscure sociopolitcal and normative factors. it is doubtful that we can give a coherent account of how the term is actually used' (1988:37)

Chomsky gives as an example of the sociopolitical factors in the ordinary concept of "language" the fact that languages spoken in various parts of China are reckoned dialects of one language, although they are as diverse as various European languages; and the classification of dialects of German as such, although some are more like some dialects of Dutch than other German dialects. The proper object of linguistic study is therefore, in principle, the language of an individual, in Chomsky's view. He assumes that the concept "language" can be given a scientifically respectable definition. If, by contrast, it is thought of as a social (or sociological) concept, the assumption that it can be defined on the model of concepts in the natural or mathematical sciences might be abandoned. Chomsky's objection to the ordinary concept of language, that it's no good for a scientific linguistics, only motivates his redefinition of "language" if it is assumed that

there can or must be a 'scientific', ie. psychological/biological, linguistics, in his sense.

However, one aspect of "idealization" in linguistics is abstraction from differences among the idiolects ("languages") of members of a linguistic community, so that in practice Chomsky, and other linguists, refer to "English", "Spanish", etc., in more or less the normal way. This is presumed to be comparable, for example, to the abstraction involved in a physical theory which disregards the differences between the objects in its domain.

Subjectivism: Chomsky denies that a grammar has any reality apart from the individual's knowledge of it. Grammar is not, for example, a property of utterances and inscriptions of sentences, nor of the totality of these, nor of the set of possible utterances. Grammar (as theory) does not describe the use of language, in any way:

> A generative grammar is not a set of statements about externalized objects constructed in some manner. Rather, it purports to depict exactly what one knows when one knows a language: that is, what has been learned, as supplemented by innate principles. (Chomsky 1986:25)

And

> ..it seems that when we speak of a person as knowing a language, we do not mean that he or she knows an infinite set of sentences, or sound-meaning pairs taken in extension, or a set of acts or behaviors; rather, what we mean is that the person knows what makes sound and meaning relate to one another in a specific way, what makes them "hang together", a particular characterisation of a function, perhaps. The person has a notion of structure, and knows an I-language as characterized

> by the linguist's grammar. When we say that it is a rule of English that objects follow verbs, as distinct from the rule of Japanese that verbs follow objects, we are not saying that this is a rule of some set of sentences or behaviours, but rather that it is a rule of a system of rules, English, an I-language.' (Chomsky 1986:27)

This distinguishes Chomsky's theory from its predecessors in American linguistics, such as the structuralism of Bloomfield. He rejects, too, the Platonic-realist idea of language, 'what we might call "P-languages" (P-English, P-Japanese, etc.), existing in a Platonic heaven alongside of arithmetic and (perhaps) set theory' (C.1986:33). According to such externalist, or objective, views, knowledge of a language, or of grammar, is a relation between an individual and something independent of the individual. By contrast, Chomsky understands "language" to mean 'some element of the mind of the person who knows the language, acquired by the learner, and used by the speaker-hearer.' (Chomsky 1986:22) In Chomsky's view, if knowledge of a language (grammar) is any kind of relation, it must be an internal relation of the individual (as, for example, I might be said to stand in a relation to my leg, or to my thoughts or memories or ambitions).

> The question of truth, conformity to an external reality, does not arise in the way it does in connection with our knowledge of the properties of objects. (1980:27)

It follows from this view that an individual "knows" his or her "language" perfectly, which is not to say that they always

use it in accord with the rules they "know":

> There is no objective, external standard against which to check the system of rules and principles relating sound and meaning - the grammar - constructed by the mind. By definition, a person knows his language (or several dialects and languages) perfectly, though we can ask how the system created by one speaker matches that of another. (1971:21-2)

Mentalism: Grammar, and "language" in Chomsky's usage, is a psychological phenomenon, an aspect of the mental life of the individual who knows the grammar (language). Mentalism distinguishes Chomsky's position from that of behaviourism, which might also be construed as a kind of subjectivism. What language an individual "knows", what grammar, is not explained in terms of behaviour, linguistic or otherwise.

"Mentalism", then, is the thesis that the object of linguistic study, what is described by a linguist's theory (a grammar in the first sense), is a state of the speaker's mind underlying the use of language, a state of "knowledge". It follows that knowledge of grammar (language) is to be explained just in terms of properties of the minds of speakers.

Chomsky's position thus has the seemingly paradoxical consequence that knowledge of grammar has itself for its object. A grammar is, for Chomsky, the speaker's state of competence, knowledge of grammar. This is the reason for the ambiguity which Chomsky acknowledges in his use of "grammar", to refer to both the linguist's theory and the state of

knowledge of the individual which the linguist's theory seeks to describe, and perhaps also for his hesitation in speaking of "knowledge". (Hence a problem for Chomsky's theory: Is it coherent to attribute knowledge in the absence of something (distinct) that is known? Can there be knowledge if there is no possibility of being wrong?)

The ambiguity in Chomsky's usage of the term "grammar", arises from the theses of competence and mentalism together. "Grammar" can mean a linguist's theory of a particular language; it can also mean the "knowledge" in the mind of the speaker of which grammar (in the first sense) is a theory:

> We must be careful to distinguish the grammar, regarded as a structure postulated in the mind, from the linguist's grammar, which is an explicit articulated theory that attempts to express precisely the rules and principles of the grammar in the mind of the ideal speaker-hearer. The linguist's grammar is a scientific theory, correct insofar as it corresponds to the internally represented grammar.....it is common to use the term "grammar" with systematic ambiguity, letting the context determine whether it refers to the internalized grammar or to the linguist's theory. (1980:220)

Something like this ambiguity is present in ordinary usage. "Grammar" can refer to the theory of a language, in the sense in which two linguists may propose different grammars of one language; close to this is the sense in which a book is described as a "Latin grammar", for example. On the other hand, "grammar" can refer to a property of a language, in the sense in which we can say that the grammar of Spanish is similar to that of Italian. Chomsky's use of "grammar" to

refer to the speaker's hypothesised knowledge is in effect the latter ordinary usage, modified in accord with Chomsky's view of the proper object of linguistic study - of what a language is, or what "language" refers to. His subjectivism, specifically his mentalism, determines this version of the ambiguity of the term "grammar".

Intellectualism: Intellectualism is the attempt to explain intelligent (intentional, deliberate, rational) behaviour and abilities by reference to underlying rational processes. In particular, Chomsky proposes to explain intelligent (rational) linguistic behaviour - speech, writing, comprehension, and so on - by the hypothesis of underlying, unconscious reasoning:

> Learning is a matter of fixing the system within the permissible range; in the language case, by setting the parameters of UG [Universal Grammar, the innate system of principles which enables us to learn language] and adding a periphery of marked exceptions. What we know is then determined by the functioning of the mature system, sometimes involving moderately complex inference-like computations. (Chomsky 1986:264)

"Competence" is the theory that a speaker has knowledge of abstract rules and principles, expressible in theoretical terms, and that use of language (including understanding), is explained by something like inferential reasoning from the abstract rules to particular consequences or applications. Derivations from the rules are hypothesised, in a sense analogous to that in which, for example, a student of logic may derive consequences from a given proposition in accordance

with general rules of logical deduction.

Or again, the process hypothesised in the mind of the speaker-hearer is analogous to that carried out explicitly by a code-breaker working from knowledge of the principles of the code. 'It does not seem problematic to entertain the hypothesis that the mechanisms of mind permit something akin to deduction as part of their computational character' (Chomsky 1986:270). Intellectualism is one characteristic of traditional rationalism; the thesis of "innate" ideas or knowledge being the other principle doctrine. This thesis is characteristic not only of Chomsky's theory of knowledge of language, but of the whole approach to problems of psychology and philosophy of mind in 'cognitive psychology'. As Dennett remarks, 'the reigning ideology of cognitive science sets itself so defiantly against Ryle [who attacked the 'intellectualist myth', in The Concept of Mind, 1949] that it might with some justice be called intellectualist science.' (Dennett 1987:214) Intellectualism is articulated in the thesis of representationalism:

Representationalism: Chomsky holds that to know a grammar, or to interpret a sentence (to understand it by deriving its underlying grammatical structure), is to have "mental representations" of the (rules of) the grammar, or of the grammatical structure of the particular sentence, in one's "mind/brain":

> To know a language, I am assuming, is to be in a certain mental state, which persists as a relatively steady component of transitory mental states. What kind of mental state? I assume further that to be in such a mental state is to have a certain mental structure consisting of a system of rules and principles that generate and relate mental representations of various types. (1980:48)
>
> I see no reasonable alternative to the position that grammars are internally represented in the mind....(1980:86-7)
>
> I am assuming grammatical competence to be a system of rules that generate and relate certain mental representations, including in particular representations of form and meaning....(1980:90)

Interpreting a sentence (or, conversely, producing one) is then conceived as a mental process of producing a set of representations, structural descriptions of the sentence at all levels - phonetic, syntactic, perhaps semantic - by using one's knowledge of grammar, perhaps consisting of representations of rules. The intuitive way to think of this is by analogy with the explicit, written activity of a linguist, working on a chalkboard or paper or a word processor. Fodor (1975) has drawn the conclusion that such theories assume a <u>language</u> employed in the mind/brain. It is difficult to imagine what "mental representations" could be if not something <u>like</u> expressions in a language.

It has been suggested (Stabler 1983) that such a theory need not be committed to the representation of the rules of grammar, but only to the representations which constitute interpretations of sentences in accord with the rules

attributed to the person. Stabler says that there are two types of computer program, those which employ a representation of the (rules of) the program as part of their operation, and others which have no such representation, but merely compute in accord with the rules which specify them. In both cases the program can be described in terms of the rules, as following such-and-such. Dennett remarks that a pocket calculator has nothing in it equivalent to a representation of the mathematical functions it carries out, ie. nothing that could be translated as the rules of addition, etc.; but these do describe and explain its operations (Dennett 1987:221). Similarly it might be thought that the human mind/brain has no representation of the rules of grammar as such, but that knowing them just consists in deriving interpretations in accord with them.

Chomsky seems committed to the stronger view, that the grammar is represented as such: 'We do of course assume that the rules are somehow represented in physical mechanisms' (1986:257).[12] This is perhaps implied by his view that knowledge of grammar could exist without being used at all - if "use" includes application in mental operations. It might also be inferred along the lines of the basic inference from use of language to knowledge of grammar: the ability to derive interpretations of novel sentences is best explained by knowledge of the rules of derivation, which is best explained

---

[12] See also Fodor (1975) p74 fn.

as representation of those rules. (Here begins a regress, which I will discuss further in the next chapter.) In any case, the idea that analyses are "mentally represented" is common to these various positions, whether or not representation of the rules is thought necessary. I will discuss both possibilities.

Computationalism: Chomsky describes the process of applying knowledge of grammar in the derivation of structural descriptions of sentences as "computation", suggesting that the mind/brain should be thought of as some kind of computer, or as like computers in some respect:

> certain aspects of the mind/brain can be usefully thought of on the model of computational systems of rules that form and modify representations and that are put to use in interpretation and action. (1986b:1)

> Linguistics is the abstract study of certain mechanisms, their growth and maturation. We may impute existence to the postulated structures at the initial, intermediate, and steady states in just the same sense as we impute existence to a program that we believe to be somehow represented in a computer...(1980:188)

In this view, the person interpreting a sentence carries out "computations over representations". The "language" in which grammar and grammatical analyses are represented is then thought of as analogous to the machine-language, or a higher-level "language" of a computer. The metaphor of "mental computations" pervades Chomsky's recent writings, as it does much of the literature in linguistics and cognitive science. A "grammar" is, in this analogy, a program run on the brain-

computer.

Physicalism: The thesis of mentalism implies that linguistics is a branch of psychology, a conclusion which Chomsky explicitly accepts. Chomsky also holds that psychology is a branch of biology:

> Linguistics, conceived as the study of I-language and S0[13], becomes part of psychology, ultimately biology. Linguistics will be incorporated within the natural sciences insofar as mechanisms are discovered that have the properties revealed in these more abstract studies (C. 1986:27).

Chomsky holds that talk of the mind is talk of the brain at some level of abstraction, and in particular that grammar, in the sense of the linguist's theory, is an abstract description of 'structures in the brain':

> When I use such terms as "mind", "mental representation", "mental computation", and the like, I am keeping to the level of abstract characterisation of the properties of certain physical mechanisms, as yet almost entirely unknown. ...we may think of the study of mental faculties as actually being a study of the body - specifically the brain - conducted at a certain level of abstraction. (C.1980b:2)

Grammar, in the sense of the speaker's knowledge or competence, is thus a property of the brain, or a structure or function of the brain abstractly described. To know a grammar, then, is for one's brain to be in a certain state. Just as linguistics is a part of psychology, so also

---

[13] "S0" is the "initial state" of a person's "language faculty", constituting the innate knowledge of principles of Universal Grammar, which in Chomsky's view a child must possess in order to be able to learn a language.

psychology is a part of biology, in Chomsky's view:

> The statements of grammar are statements of the theory of mind about the I-language; hence they are indirectly statements about structures of the brain presented at a certain level of abstraction from mechanisms. (C. 1986b:8)
>
> Linguistics, so conceived, becomes part of psychology, ultimately biology. (C.1986b:10)

This position suggests that in seeking an answer to the question 'What does knowledge of grammar consist in?', we might look to a physical (physicalist) theory. For example, some version of "functionalism", which attempts to explain how states or changes in the brain could constitute intentional mental states of a person. Chomsky expresses this view of the identity of mind and brain by using the hybrid term "mind/brain".[14]

---

[14] Also 1980: 5,31,202; 1986:23

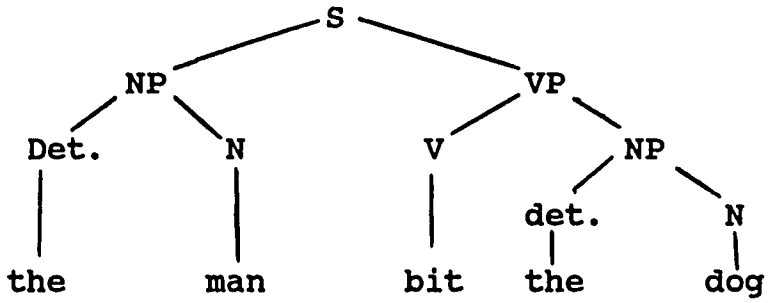
CHAPTER 2

UNDERSTANDING A SENTENCE IN CHOMSKY'S THEORY

2.1 INTRODUCTION: HOW TO UNDERSTAND A SENTENCE

According to Chomsky, understanding a sentence is an act of interpretation. When a person hears or reads a sentence in a language she knows, she derives a grammatical analysis, by applying her knowledge of grammatical rules and principles. (Perhaps also 'semantic' and 'pragmatic' knowledge or understanding.) In this way she produces a structural description (or a set of structural descriptions) of the sentence. For the sentence 'The man bit the dog', the analysis might be something like this[15]:

[X]



This analysis is carried out, in Chomsky's view, at an unconscious level of the interpreter's mind. The structural description occurs as a 'mental representation' in the mind,

---

[15] Other forms of syntactic analysis or representation have been proposed, eg. Chomsky's X-bar theory. Nothing turns on the choice among such theories, for my discussion. Something of this sort occurs in the speaker-hearer's head, according to Chomsky.

or brain, of the interpreter. The rules she knows may also be 'mentally represented'. Her knowledge of grammar consists of either representations of rules and principles, or (the capacity to generate) representations of structural descriptions, or both. Whether rules are represented is controversial, since the hypothesis that they are threatens to generate a regress. (See section 2.4 below for discussion.)

Chomsky would describe such an analysis as part of the theory of <u>competence</u> - the speaker-hearer's knowledge of grammar - not <u>performance</u> - how a speaker-hearer actually applies this knowledge to interpret a sentence. But a theory of performance would 'incorporate' a theory of competence (Chomsky 1965:9), which means that something like this analysis would actually be derived, somehow, in comprehending a sentence. As I argued in Chapter 1, that 'competence' (knowledge of a grammar) is <u>used</u> in 'performance' is assumed in the argument that we <u>have</u> such knowledge.

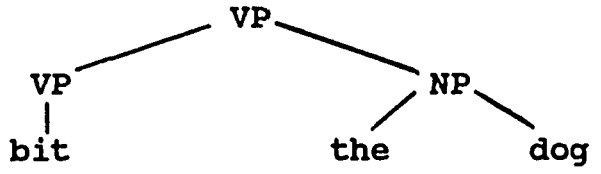
By deriving such an analysis of the (underlying) grammatical structure of a sentence, the hearer grasps the grammar of the sentence - comes to know the grammatical relations that obtain between the words and phrases that make up the sentence. This partly explains her understanding of the sentence. 'Partly', because she must also use her ('semantic') knowledge of the meanings of the

words and expressions in the sentence, and perhaps also relevant general and 'pragmatic' knowledge, eg. about human and canine behaviour, or about the context, such as which dog and which man are referred to, and so on.

Knowledge of grammar is not proposed as a complete explanation of language understanding. In the narrower sense, in which "grammar" means "syntax", it does not include even knowledge of the meanings of words. But it should explain a particular aspect of the understanding of language, namely, the ability to grasp the grammar (syntax) of a novel sentence. Understanding the sentence 'The man bit the dog', for example, involves understanding the grammatical relations among its constituent words: that "the man" is the agent of the verb "bit", that "bit" is a transitive verb which has "the dog" as its object, and so on. There is a totality of grammatical (syntactic) relations among the five words, all of which must be grasped in understanding the sentence. This is what must be explained by the hypothesised knowledge of grammar, in Chomsky's theory.

A structural description is a kind of translation of a sentence. The sentence is translated into a grammatically explicit form: one which represents grammatical relations among the words which are not explicitly represented in the original. That aspect of understanding the sentence which consists in knowing what these relations are, according to

Chomsky's theory, is explained by the derivation and representation of the analysis. Thus to know that "the dog" is the direct object of the verb "bit" in the sentence 'The man bit the dog', is to derive the representation [X], in particular to have a 'mental representation' part of which is equivalent to:



This (partial) representation describes or represents the words "bit" and "the dog" as standing in the relation of verb and direct object, within the sentence. Similarly for all the other grammatical properties and (internal) relations of the sentence. Knowledge of this grammatical property of the sentence appears to be explained by the occurrence of such a mental representation.

The argument for tacit mental representation of the underlying grammatical structures of sentences, can be summarised:

1. Speakers understand sentences.
2. Understanding a sentence involves grasping its grammar - knowing what grammatical relations obtain between the words and expressions of which it is composed.
3. Mental representation of structural descriptions would explain grasp of grammar.
4. It is therefore reasonable to hypothesise that speakers

derive structural descriptions, which they represent mentally, when they understand sentences. (This is an empirical hypothesis, to be judged by the normal standards of empirical science, in Chomsky's view.[16])

(1) is the explicandum. (2) is surely correct in some sense. (4) opens the whole debate about 'mental representation'; it looks as though it's a question for empirical science, not philosophy (alone) to answer. Is (3) true? In this chapter, I will argue that conscious or explicit representation of structural descriptions would not explain grasp of grammar. This puts a new burden of proof on the theory of unconscious mental representation. It must be shown that there is a relevant difference, such that it would explain understanding.

---

[16] eg. 'Focusing on the I-language, however, the problem is a rather different one: to find the mental representations that underlie the production and perception of speech and the rules that relate these representations to the physical events of speech. The problem is to find the best theory to account for a wide variety of facts, and we do not expect that analytic procedures exist to accomplish this task, just as there are no such procedures in other fields.(Chomsky 1986:41). And 'theories of grammar and UG are empirical theories, not part of mathematics. .. we propose certain principles, parameters, representations, modes of computation, and so forth, and we seek to explain the facts in these terms, taking this account tentatively to express the truth about the language faculty.' (ibid. 248-9)

2.2 EXPLICIT REPRESENTATION

The idea of mental representation of grammatical rules and analyses of sentences relies on an analogy with written rules and analyses produced by linguists: these are the explicit model for the tacit theory. The idea is that something like the linguist's analysis of a sentence - as found in a textbook of linguistics - occurs in an unconscious region of the mind (brain). Would explicit knowledge of rules, and application of this knowledge to derive structural descriptions, explain how we understand a (novel) sentence?

Imagine an individual who has conscious knowledge of grammatical rules, in the sense that she can recite them on demand (or at least recognise correct formulations of them), and applies this knowledge by writing out structural descriptions of (novel) sentences she hears and reads. (Her knowledge of the rules might be imagined in a similar way, as possession of a rule-book in which they are listed, which she refers to as necessary. See 2.4 below.) I will call this the "external model". In effect, it assumes Chomsky's theses of intellectualism and representationalism, but not mentalism (as discussed in Chapter 1).

Suppose the speaker-hearer in the external model hears or reads the sentence 'The man bit the dog'. She recalls the appropriate rules, or finds them in the rule-book, applies them to the sentence, and writes out the structural

description [X]. Does she grasp the grammatical structure of the sentence? Does she know what grammatical relations obtain between the words of which the sentence is composed, such as the verb-direct object relation of "bit" and "the dog"? This relation is explicitly represented in the analysis she has derived in the model. So if representing is knowing, she should now know that this is the relation between "bit" and "the dog" in the sentence analysed.

But she won't know that [X] represents this relation, hence she won't know that "ate" and "the bone" stand in this relation in the sentence, unless she understands [X]. Just as a French translation of a Spanish sentence won't enable a person to understand the Spanish sentence (to know what it says) unless he understands the French sentence. [X] represents this grammatical relation between these words; it says that they stand in this relation. So being in possession of the representation [X], she will know that the words are related in this way, <u>if she understands [X]</u>. What explains this understanding?

The representation [X] is a complex sign analogous to a sentence of English. It is composed in a technical notation derived from ordinary English ("VP" is defined by means of the English terms "verb" and "phrase" etc.). It is composed of signs in a certain spatial arrangement. Its meaning is a function of the meanings of the component signs and the syntax of the notational system. At least, if this is true

of ordinary sentences, as Chomsky assumes, it is also true of this structural description (eg. 1980:220ff). Fodor and Lepore explain the motivation:

> the point of assuming compositionality is primarily to explain the productivity of linguistic (/cognitive) capacities. The productivity of English answers the question of how it is possible for an English speaker to grasp the sense of new expressions on the basis of a finite acquaintance with his language. (1992:241)

Obviously this assumption is essential to the explanatory aim of Chomsky's psychological theory.

Hence, on the assumptions of the basic argument for tacit knowledge of grammar discussed above, understanding this representation is achieved by use of knowledge of the rules and principles of its grammar. This argument therefore leads to an explanatory regress. Understanding (the grammar of) the original sentence is not explained by derivation of an external representation of its syntactic structure, unless understanding of (the grammar of) the representation is assumed: and that requires exactly the same kind of understanding as the representation was supposed to explain.

In particular, the comprehension of <u>novel</u> sentences - sentences the hearer has never encountered before - is explained only if comprehension of a novel structural description is assumed. For if the sentence is novel, so is its structural description.

A grammatical analysis, such as [X], is a kind of

explanation of the sentence analysed. It seems to explain what must be known about the syntax of the sentence in order to understand it. So it's plausible that to be in possession of this explanation is to be part way to understanding the sentence. But this is only so if the grammar of the explanation is understood.

The point is not that more analysis of the sentence is necessary to reveal its underlying structure. We can assume that the analysis reached is correct and complete (although the concepts of correctness and completeness may be problematic in view of Chomsky's subjectivism). But 'an interpretation is something that is given in signs' (Wittgenstein 1974:47). The problem is that an analysis, however complete, is still a representation, ie. a complex, sentence-like sign, the meaning of which is - if this is generally true of sentences - a function of its component symbols and their arrangement (the syntax of the representation). Therefore, to understand it requires knowledge of its syntactic structure, which (according to Chomsky's argument) requires knowledge of the relevant syntactic rules. Explicit representation of a structural description would only explain the grammatical aspect of understanding a sentence, if the grammatical aspect of understanding a structural description is assumed.

2.3 MENTAL REPRESENTATION

(a) Conscious representation. We must therefore ask whether understanding a representation is explained by its internalisation: the thesis of mentalism, that knowledge of rules, and derivation of representations of syntactic structures, takes place in the mind (or brain) of the speaker-hearer, rather than externally, on paper or in some other public medium.

As a first construal of "internalising" the rules and representations of sentences, suppose the procedure carried out on paper, in the external model, is conducted in conscious thought. Suppose that the structural description of the sentence is "seen" in imagination, in the mind's eye; or that the interpreter says it to herself (as one can go through a calculation in silent speech). Representations in this sense might be thought of as Lockean "ideas". This internalisation will not resolve the problem of the external model: an imagined or visualised or mentally "heard" sentence must be understood in order to give the interpreter any knowledge of what it represents.

If a monolingual English speaker "hears" or "sees", in this sense, a Spanish translation of a French sentence, he won't be any nearer to understanding the French sentence or knowing what it means. If a person who has had no training in linguistics has a vision or mental image of the structural description {X}, it will tell him nothing about the syntax of the sentence [S]. What is needed, of course,

is some understanding of (X), and as before, this seems to assume that its syntax is understood, which by Chomsky's reasoning involves knowledge of relevant rules of grammar, and so on.

In general, a sentence, or a sentence-like complex sign, visualised or heard in imagination, just like a written or spoken one, is the kind of thing that, in Chomsky's theory, is understood by the application of knowledge of grammar. Again, this hypothesis generates an explanatory regress; whether the inner representation is assumed to be composed in the speaker-hearer's public language or in some other language or sign system. (This is the kind of picture of understanding criticised by Wittgenstein in the 'rule-following' considerations of the Philosophical Investigations, especially #86ff.)

(b) Tacit representation. Chomsky maintains that knowledge of rules, and the derivation of representations this knowledge makes possible, is "tacit" or unconscious. Now we need to ask whether this aspect of the hypothesis resolves the difficulty in explaining sentence-comprehension. Chomsky has suggested that tacit knowledge (or "cognizing") differs from ordinary knowledge only in being unconscious, or inaccessible to consciousness.[17] If the same is to be said of deriving and

[17] eg. Chomsky 1986:269.

representing structural descriptions, by applying this knowledge, then the difficulty remains. That is to say, if we think of tacit representation of rules and structural descriptions as essentially just like conscious representation, there is no reason to think understanding is explained by representation.

The characterisation of knowledge of grammar, and derivation of structural descriptions, as mental and unconscious, may be the kind of conceptual manoeuvre that Wittgenstein described as the 'decisive move in the conjuring trick' (Philosophical Investigations). We hypothesise processes whose nature we do not understand, and suppose that an explanation will be found sometime. It's assumed that some coherent explanation can be given; that the hypothesis is therefore explanatory.

If being unconscious, inaccessible to the conscious mind, is the only difference between tacit knowledge and interpretation and ordinary mental events, then mental representations of this kind will be as ineffective as written or conscious representations. Theories that imply a more radical difference of kind, especially the computational theory of mind, will be considered in the next chapter.

2.4 REPRESENTING RULES

According to Chomsky, the ability to use and understand

sentences in a language is explained, in part, by knowledge of grammar, consisting (in part) of knowledge of syntactic rules and principles. This knowledge is used in the derivation of structural descriptions of sentences. The rules and principles, or the structural descriptions they are used to derive, or both, are present in the speaker's mind (brain) as 'mental representations'. This representation constitutes the speaker-hearer's knowledge.

Knowledge of grammatical rules and principles explains the ability to derive structural representations of sentences, in Chomsky's theory. Hearing or reading a sentence, you apply the rules you know to the sentence (or its phonetic representation) to derive the structural descriptions. Knowing the rules, you follow them to produce the analysis. Would <u>conscious</u> knowledge of rules explain this ability? There are two ways to think of knowledge of rules, in Chomsky's model. The rules are represented, or they are not.

(a) Suppose they are represented. This view was endorsed by Fodor (1975), explaining why the behavior of 'organisms' is to be explained by unconscious rule-following, while eg. the movements of planets is not explained by saying that they 'follow' Kepler's Laws:

> What distinguishes what organisms do from what planets do is that a <u>representation of the rules they follow constitutes one of the causal determinants of their behavior</u>. (Fodor 1975:74 fn)

Suppose the representation of rules to be external, in the following sense. The interpreter (the speaker-hearer exercising her ability to understand, ie. in Chomsky's theory interpret, novel sentences) has a rule-book, in which all the rules and principles of the language are represented: this constitutes her "knowledge" of the grammar. When she hears a sentence, she makes a phonetic transcript of it. Then she looks up the relevant rules in the rule-book, applies them to the phonetic representation of the sentence, and in this way derives, ie. writes out, a set of structural descriptions of the sentence - something like [X]; or judges that the sentence is or is not grammatical, or the like.

The (representations of) rules in the rule-book enable her to derive descriptions, etc., by guiding her interpretation or judgement. This explains how she is able to produce an analysis, interpret and hence make sense of a novel sentence.

But this explanation makes sense only if she understands the rules she looks up. A rule in the rule-book is a complex sign, a kind of sentence. It consists of words and symbols in a spatial arrangement, in accordance with a syntactic system. To understand it means (ex hypothesi) to know the syntactic system in which the rule is formulated; which apparently means to have knowledge of the relevant

rules. Otherwise, the rule is just a complex mark on the paper, which would not guide her analysis of the sentence; it would not mean anything to her. For example:

[S] The books is on the shelf.

How does the reader know this is ungrammatical (assuming that it is ungrammatical, in his or her idiolect)? Following Chomsky, we should say: by applying knowledge of the rule of English (the hearer's idiolect) that a verb must agree in number with its subject. "Knowing" this, in Chomsky's theory, means having a representation of it (in mind):

[R] A verb must agree in number with its subject.[18]

So the interpreter, in the external model, looks up the rule R, applies it to the sentence [S], and discovers the grammatical error, ie. sees that the sentence is not in accord with the rule.

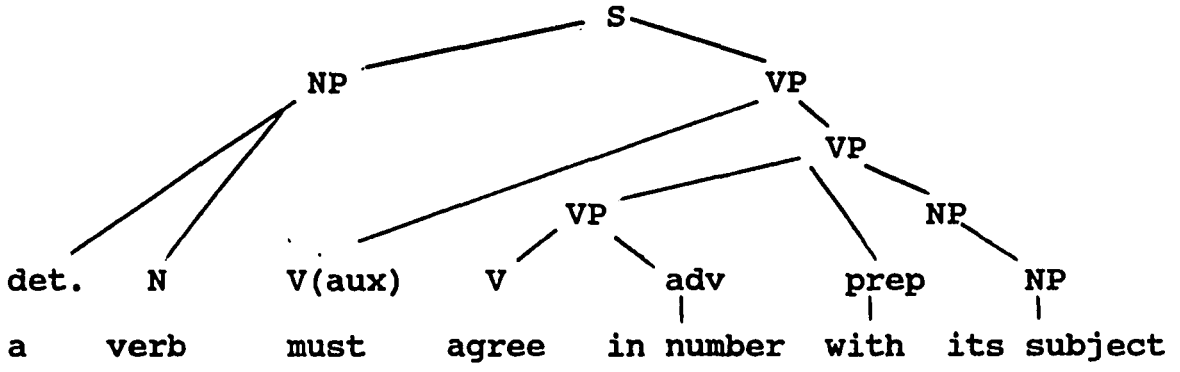
The problem is this: Applying the rule to the sentence (or comparing the sentence with the rule, etc.) presupposes understanding of the rule. A representation of [R] would be no use in assessing the grammaticality of [S], or deriving a structural description of the sentence, if this rule was not understood. But [R], the representation of the rule, is a sentence, and by (Chomsky's) hypothesis, understanding a sentence means, in part, deriving its structural

---

[18] This is not to suggest that a serious generative grammar will actually include <u>this</u> rule.

description, by applying relevant grammatical rules; which in turn must be understood. And so on.

It would not help to suppose that the rule is formulated in a syntactically explicit form, such as the labelled-node tree form, as it were pre-analysed, something like:



This would still be a complex sign consisting of words and symbols in a spatial arrangement, so that its meaning is a function of the syntactic system in which it is composed (again, assuming 'compositionality'). The problem is not that the rule is inadequately analysed, but rather that no matter how fully analysed, it is still a complex sign which must be understood in accordance with its grammar.

Suppose the knowledge of rules is conscious, in an ordinary sense, rather than consisting in representations in a rule-book. The speaker can think of the rule, say it to herself (or out loud), and write it out if necessary, like the engineer's knowledge of principles of mechanics, discussed in Chapter 1. When confronted by a novel

sentence, she recalls the relevant rules, writes them down, and applies them to the sentence to derive structural descriptions. But this supposes that she understands the rules she recalls, says to herself, or writes out. Again, if the rule is thought of as any sort of representation - in the mind's eye or voice, or on paper - the problem arises of understanding a complex, sentence-like sign. Without this, it can provide no guidance in analysing a sentence.

Perhaps knowing a rule involves understanding it. Fisher (1974) points out that there is one sense of 'knowing a rule' in which, for example, a child who can recite a rule is said to 'know' it, even if he can't follow it, and another sense in which a child who can (and usually does) follow a rule but not recite it is said to know it. But in Chomsky's theory, if knowledge of rules, hence the ability to follow them, is said to consist in their (mental) representation, the assumption that the rules can be followed would be question-begging. Having a representation of a rule explains the ability to act in accord with it if it is understood. But understanding linguistic representations - such as sentences, including rule-formulations - is the explicandum. And on the other hand, if knowledge of rules consists in their representation, but they are not understood, how could this begin to explain the ability to follow them? This cannot be simply assumed.

(b) Suppose, then, that rules are not explicitly represented. Knowledge of rules is, rather, like the second kind Fisher distinguishes. A speaker-hearer who can derive structural descriptions is said to 'know' the relevant rules - the rules that govern her derivations - though she cannot cite them and does not represent them. But then this ability to derive structural descriptions needs explaining. It is, by hypothesis, the ability to interpret <u>novel</u> sentences, deriving novel sets of structural descriptions. This is the kind of ability knowledge of rules of grammar was invoked to explain. Suppose we ask for an explanation of this ability. The answer should be that the speaker-hearer knows rules and principles that enable her to analyse a novel sentence. What does this knowledge consist in? If it consists in representations, there is a problem about how these are understood. If it does not consist in such representations, but consists rather in the ability to derive the (novel) analyses, then we seem to have <u>no</u> explanation, but merely the hypothesis of an unexplained ability. Furthermore, for the reasons I gave in the previous section, it would be a mysterious ability that leads to a further mystery - how she understands the resultant novel representations. (Chomsky, in any case, argues against appeal to ability as a final explanation of use and understanding of language eg. Chomsky 1986:9-13.)

The external model thus presents a dilemma, resolution

of which is a condition of adequacy of an internal model. Either rules are represented, and the representations guide interpretation of sentences, in which case understanding the (representations of) rules needs explaining, and a regress threatens; or rules are not represented, and the ability to analyse or interpret sentences stands in need of explanation, in the same way as and for the same reason that the ability to produce and understand sentences needed explanation in the first place.

2.5 PHYSICALISM

Does the thesis of physicalism help? Representations on the outside of the body would not help explain understanding. If structural descriptions appeared on the back of my hand when I heard a sentence, this would not enable me to understand the sentence, for reasons already discussed. It would be no more use than analyses on paper. Nor would it help if they appeared on the inside of my body - in my heart, for example. In fact, this would present an additional problem of access. It is thought to make a difference, however, if something of the kind happens in the thinking part of the body - the brain. But why?

Suppose that structural descriptions of a sentence appear in my brain every time I hear a sentence. I hear 'The man bit the dog', for example, and [X] forms in my language-processing lobe. On the face of it, this is useless unless (a) It is perceived by (some other part of) my brain, and (b) it is understood by whatever part of me perceives it. Both of these conditions threaten to generate a regress. (a) What is perception of one part of the brain by another? (b) How is a representation understood by the perceiving part? These problems look quite analogous to that facing Hume's model of the mind, in its need for a self to perceive and unite the 'bundle of impressions'. They are also analogous to the initial Chomskyan problem of

understanding language.

This kind of objection is considered quite naive and unscientific. It is supposed to be answered - or dismissed - by a scientific model of the brain-as-mind, in which view the brain is a sort of computer.

2.6 THE COMPUTER MODEL OF MIND

I have argued that there are two conceptual problems in Chomsky's theory of language understanding and speech. One is to explain how mental representations are understood. The other is to explain how rules can be followed. Both threaten to generate explanatory regress.

There is an influential idea of the nature of mind which proposes an answer to these problems. The essential idea of this model is that the brain is a kind of computer. Representations are physical states of the system in causal connection with other states, and with perception and behaviour - 'input' and 'output'. Mental activity is (like) data processing. Block explains:

> The basic idea is that the mind is the program of the brain and that the mechanisms of mind involve the same sorts of computations over representations that occur in computers. (Block 1990:247)

And in a similar vein, Fodor writes that

> contemporary cognitive theory takes it for granted that the paradigmatic psychological process is a sequence of transformations of mental representations and that the paradigmatic cognitive system is one which effects such transformations....[ie.] computational systems of

one sort or another. (Fodor 1983:29)

In this view, a theory of linguistic performance - a theory of how an individual derives structural descriptions, etc. - is a specification of a program run on the brain/computer. It no doubt operates in conjunction with other programs - those which process visual information, store and retrieve memories of previous experience, and so on.

This model of mind seems to answer the problems of understanding and rule-following: A mental representation is a state or property of the brain which has a causal role in the operations of the brain/computer. It is connected with other representations, and ultimately with perceptions and motor-functions ('behaviour'), in a way that matches its meaning. For example, a representation meaning 'All men are mortal' might cause another meaning 'I'm going to die!' - which might eventually cause the making of a will. A representation meaning 'NP1-VP(passive)-prep.-NP2' might cause another meaning 'NP2-VP(active)-NP1' - representing the transformation of a passive sentence to an active.

In answer to the problem of understanding, it may be said either (a) Understanding a representation consists in its having an appropriate causal role - the right sorts of causes and effects, actual or potential, ie. causal connection is understanding, or (b) Mental representations don't have to be understood, because they have causal roles:

a sentence heard is understood by being processed in the brain, while a mental representation is processed rather than understood.

The idea of (a) is that understanding consists in the ability to use a language, or system of representation - where "use" refers to internal as well as external capacities. Mental representations are understood if they have a suitable causal role in the individual's (physically realised) mental processes. Schiffer makes this point:

> Now it is natural to equate a conceptual-role theory for the language of thought with a theory of language-of-thought understanding. For to understand a language is to know how to use it, but to "use" a system of mental representation is just for its formulae to have conceptual roles. (Schiffer 1987:186).

(b) emphasises the disanalogy between external representation - ordinary language - and 'representation' in a computer. The latter does not involve an agent perceiving, interpreting or 'using' words and sentences. It is part of a mechanism, not a kind of behavior. Fodor explains:

> ... we don't need a combinatorial semantics for mentalese because using a productive language as a medium of computation doesn't require access to its semantics; by definition, computational processes are exhaustively syntactic, and it's not in dispute that mentalese has a combinatorial syntax. Making the syntactic character of computation clear was Turing's foundational contribution to the philosophy of mind. Turing's way of getting mental processes to be symbolic without having to postulate a regression of understanders is what the idea of computation buys you in the philosophy of mind. (Fodor 1990:188)

The significant idea here is that because mental representations have causal roles in the mental mechanism - as in an electronic or mechanical computer - they don't have to be understood. (The point of the distinction of syntax and semantics in this passage is that the causal roles of 'representations' in a computer depend entirely on their physical properties and relations, not on what they 'mean', if anything.)

Either way, the problem seems to be solved - understanding mental representations is explained or explained away. This is also essentially the reply of several critics to Searle's famous 'Chinese Room' argument (Searle, Bridgeman, Dennett, Fodor, Lycan, Pylyshyn: 1980).

Dennett states the problem:

> It also seems (to many) that understanding a heard sentence must be somehow translating it into some internal message, but how will this message in turn be understood; by translating it into something else? ....On the one hand, how could any theory of psychology make sense of representations that understand themselves, and on the other, how could any theory of psychology avoid regress or circularity if it posits at least one representation-understander in addition to the representations? (Dennett 1978:122-3)

Dennett explains that theories in artificial intelligence are full of 'homunculus talk', which is to say that discussion proceeds as though there are understanders in computers - likewise in brains - but that this way of talking is in fact - in principle - eliminable.

One discharges fancy homunculi from one's scheme

> by organizing armies of idiots to do the work.....One never quite gets completely self-understanding representations (unless one stands back and views all representations in the system from a global vantage point), but all homunculi are ultimately discharged. (Dennett 1978:124)

In other words, 'homunculi' - understanders - are eliminated by replacing them at the ground level by causal connections; and understanding is a function or property of the whole mechanistic system - computer or brain (robot or human).

Rule-following is explained in the same way. Causal connection IS rule-following, in so far as the representations connected are suitably related in meaning. There is no need for a rule to be understood, in the final analysis, because representations are connected causally.

Computer programs are specified in terms of 'instructions'. 'A typical instruction might say - "Add the number stored in position 6809 to that in 4302 and put the result back into the latter storage position."' (Turing 1950) It might look as though my arguments against the possibility of rule-following would be just as applicable - or inapplicable - in this case. Someone might wonder how a computer can understand an instruction, and complain that this explanation of how computation is possible - by reference to instructions executed - generates a regress. But this would be a misguided objection. Of course there is no such problem about a computer. It doesn't have to 'understand' instructions in any problematic sense. It's just constructed to do things this way, so that for example,

the signal '6809430217' (in binary form) causes the appropriate changes in states of the machine - which Turing described as adding numbers, and so forth. Ultimately the instructions and 'representations' in a computer can be translated into a purely descriptive, physicalistic language. And that, according to the computational theory of mind, is how it is in the human brain. Ultimately, a theory of linguistic performance could be spelled out in terms of states and causal connections of neurons - in principle. (Fodor gave this explanation in response to Ryle's charge that 'mentalism' leads to a regress: Ryle 1949; Fodor 1968).

For example - no doubt simplifying drastically - the brain follows the 'add rule' if 'mental numerals' "2" and "3" CAUSE the mental numeral "5". When a cognitive psychologist proposes a theory of some mental ability, such as the ability to estimate distance, or to understand a novel sentence, she may explain it in terms of rules the subject knows, follows or applies. In principle, however, every such explanation could be spelled out in simple steps that are realised by causal connections in the central nervous system. Mathematical abilities, it might be, are finally reducible to binary calculations, which are physically realised by causal systems of 0s and 1s (states that can be described as 0s and 1s) in the brain. Block again:

> Intelligent capacities are understood via decomposition into a network of less intelligent capacities, ultimately grounded in totally mechanical capacities executed by primitive processors. (ibid 256)

The basic problem for this model of the mind, it is generally supposed, is to explain what it means to say that brain states - patterns of neural activity, levels of activation of individual neurons or groups of neurons, 'distributed' electrochemical states, or whatever it may be - are "symbols" or "representations"; and what makes them particular representations - what determines their meaning or 'content'. The computer analogy is inadequate in this respect, because the meanings of inputs and outputs of machines, and of their internal states, are determined by the humans who design and use them. (I will argue later that the analogy fails also at the level of 'uninterpreted' states and processes: brains cannot even be described as having a 'syntax' or a 'machine language'; see chapter 4.)

This way of dealing with the problem of understanding - explaining it or explaining it away - makes sense only if sense can be made of the kind of theory of mind it assumes. In the following chapters I will examine such theories.

2.7 CONCLUSION

Ordinary knowledge of grammatical rules (conscious knowledge, like that of the linguist) would not explain how a sentence is understood. In particular, it would not

explain the understanding a speaker-hearer has of the grammatical structure of a sentence (knowledge of the grammatical relations of the words and phrases in the sentence), except on the assumption of the same kind of understanding of the grammatical structure of the structural description derived. If we had conscious knowledge of rules, this would not explain what tacit knowledge of rules, in Chomsky's theory, is supposed to explain. Explicit knowledge and use of grammatical rules, in the ordinary sense, is a use of language. So it can't be a general explanation of the ability to use language. There is a disanalogy here with the case of the engineer, discussed in chapter 1. Her knowledge and use of principles of mechanics depends on her ability to use the language of mathematics; but there is no circularity, because it is not this ability which the attribution of the knowledge was intended to explain.

Furthermore, if knowledge of rules is thought of as representation of the rules, in the sense in which a written rule is a representation, the ability to analyse a sentence would not be explained by the theory that the interpreter is guided by the rules, if Chomsky's model of understanding a sentence is assumed. This is because a rule must be understood if it is to guide behavior, and a representation is a kind of sentence, and understanding a sentence (in Chomsky's theory) means knowing its grammatical structure.

Therefore knowledge of syntactic rules is necessary to understand, hence to be guided by, (a representation of) a rule. On the other hand, if the rules are not represented, the ability to derive analyses of sentences seems to be unexplained.

The computer model of the mind purports to deal with both the problem of understanding and the problem of rule-following. This is the contemporary form of physicalism in philosophy of mind, and is peculiarly suited to representational theories. The problem, then, is whether the analogy of the human brain with a computer, mind with computer program, and neural states with computer data or symbols can be justified.

CHAPTER 3

FUNCTIONALISM AND THE COMPUTER MODEL OF THE MIND

3.1 INTRODUCTION

Chomsky's idea that understanding and speaking a language involves the tacit mental knowledge and use of a generative grammar, creates a problem about understanding which threatens to generate a regress. If understanding a sentence involves the unconscious derivation of structural descriptions by application of grammatical rules, these 'mental representations' must in turn be understood. And if knowledge of rules is to explain the ability to derive such representations, it should, it seems, involve understanding of the rules[19].

The solution to this Rylean challenge in recent philosophy of mind, compatible with Chomsky's assumptions (as explained in Chapter 1), is some version of <u>functionalism</u>, or the <u>computational theory of mind</u>. In this view, understanding rules and representations is just a matter of their having an appropriate <u>use</u>, which is to say <u>causal role</u> in the brain[20]. Cognitive activity is, in the final analysis, a matter of causal processes. A representation is understood in so far as it has a role in a causal or functional system, a 'mechanism' or an organic analogue; as a computer may be said to

---

[19] See Chapter 2.

[20] eg. Fodor 1968, 1975.

'understand' the symbols or formulae it employs in so far as they have an appropriate causal role in its operations.

The computational theory of mind is the idea that the brain is a kind of computer, and that the mind is the brain's program. Intelligent capacities, such as the ability to reason, to act intelligently, to speak and understand a language, are possessed by virtue of the brain's program or programs; and their exercise consists in computational processes and the control of behaviour by such processes. As Block puts it:

> The basic idea is that the mind is the program of the brain, and that the mechanisms of mind involve the sorts of computations over representations that occur in computers. (Block 1990:247)[21]

So the explanation of mental capacities, and of behaviour, consists in the discovery of the 'programs' and computational procedures we employ. Intelligent behaviour is explained as the output of computational processes in the brain, which in turn are explained, in part, by reference to the input to this organ, the cerebral computer, and in part in terms of its output, bodily movement. The discovery of these programs and procedures is the research project of computational psychology.

I will argue that this idea is based on a confusion about the concepts of 'behaviour' and 'perception'. There is an equivocation between two concepts of behaviour, one often

---

[21] Also eg. Boden 1990:2; Jackendoff 1987:15-16; Pylyshyn 1984:xiii.

referred to as 'output', covering physical events, especially bodily movements, caused by processes in the body, and the other being the notion of behaviour in the sense of human actions, the things we do as a result of thought, intention, and so on. Similarly there is an equivocation in the use of 'input', between the ordinary notion of perception, exemplified by a person's seeing something, and the notion of 'input' appropriate to a computer. But 'output' and behaviour, 'input' and perception, are not the same. Lewis provides an example of this equivocation in these two passages from his 'Psychophysical and Theoretical Identification' (discussed below):

> Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. (p212)
>
> Ascriptions to me of various particular beliefs and desires, say, cannot be true if there are no such states as belief and desire; cannot be true, that is, unless the causal roles definitive of belief and desire are occupied. But these roles can only be occupied by states causally related in the proper lawful way to behavior. (p213)

Lewis's idea, as I shall explain, is that the 'platitudes' referred to in the first passage collectively define the 'causal roles' referred to in the second passage. But then for the 'mental states' so defined to be eg. 'belief and desire', 'motor responses' must be equivalent to 'behavior'. This, I will argue, cannot be right, and so the causal roles defined in terms of 'motor responses' (and 'sensory stimulations') cannot be such mental states as belief and

desire, defined in terms of behavior and perception.

A Note on 'Mental States'

I will use the expression 'mental states' (or 'events'), in the familiar contemporary fashion, to refer to beliefs, intentions, wishes, expectations, fears, and so on. But I intend no commitment to any philosophical view this usage suggests, and in fact I think the expression is problematic and potentially misleading. At the very least, this use, ubiquitous in the philosophical literature, is a misuse of the English expression. 'Mental states', in non-philosophical English, are such conditions as anxiety, excitement, hopefulness, confusion; not 'propositional attitudes', such as the belief that it's raining or the intention to carry an umbrella, but states of mind, general mental conditions. 'Mental state' in the philosophical usage, is thus a term of art, but one which inclines us to particular views of the subject matter, certain kinds of philosophical theory. I take the risk, in the interests of brevity and fashion.

3.2 HORNSBY'S CRITICISM OF FUNCTIONALISM

The argument I will develop is based on Hornsby's criticism of functionalism, in her essay 'Physicalist Thinking and Behaviour' (Hornsby 1986). Hornsby diagnoses the problem:

> I suggest that some of the allure of functionalism has resulted from failure to keep track of the use of the simple term 'behaviour'. The elements of common sense that give rise to the idea of a psychological theory seem correct when 'behaviour' is understood in Ryle's way, as including all the many things an agent does. The idea of a functional theory realized in neurophysiological states seems correct when 'behaviour' is understood in (say) the physiologist's way, as an agent's moving her body in all kinds of complex fashions. These two notions of behaviour overlap, and when 'bodily movements' is used to catch them both, they are made to appear to coincide. But the two notions do not coincide. And if one wants to preserve both common sense and the idea about functional theories, then one can only conclude that there is a complexity in propositional-attitude psychology that does not derive from any complexity in people's bodily movements conceived in ways available to common-sense psychology. [Footnote:] I think there are things to be said about perception which lead in the same direction as the things I have said about action. (Hornsby 1986:106)

Functionalism, and also I shall argue the computational theory of mind, appeal to a particular conception of behaviour: it is essentially the idea of bodily movement. This conception of behaviour is the one that fits a scientific view of the human body, or the central nervous system or brain. Bodily movements are the results of physiological processes, hence movement, and physical change generally, is explained by a theory of such processes. But human behaviour, as the object of explanation by reference to reasons, beliefs, desires, and so on, is not reducible to, or explicable in terms of, bodily movement. So a functional theory of the internal processes of a human body cannot explain behaviour in this sense. Behaviour in the physical sense may be explained

by such a theory, but as Hornsby says:

> If mental states are to be thought of as dispositions of any sort (or, if you prefer, as states that are parts of systems that exhibit an overall structure), then, to the extent that they are dispositions to behave (or states connected systematically with ways of behaving), the relevant notion of behaviour is the broad one that the philosopher behaviourists [eg. Ryle] used and the functionalists left behind. If we do employ the ordinary and richer conception of behaviour in specifying the upshots of mental states, we cannot hope to circumscribe mental states in anything like the way that the functionalist envisages. (ibid. 107)

We talk of the 'behaviour' of physical objects, meaning their movements, or movements of their parts. For example, the 'behaviour' of molecules in a gas consists of their movements, elastic collisions, and so on. The 'behaviour' of a machine is similarly conceived. Behaviour in this sense can be described in a physical vocabulary - the vocabulary of physical science, or of common sense physical description. We could talk of the 'behaviour' of humans in the same sense: describe their bodily movements and changes in a physical vocabulary, scientific or non-scientific. It is possible - but not necessarily easy - to describe the <u>movements</u> involved in the performance of an ordinary action, for example using the telephone. This would be like describing the actions of a mime without saying what she is <u>doing</u>. But this is not the normal sense in which we talk of human behaviour. Human behaviour in <u>this</u> sense consists of actions, things people do, such as walking, cooking, writing a letter, asking a question,

paying a bill, looking for a book or a job.

Human behaviour usually involves some bodily movement or change, and for some kinds of behaviour there are more or less characteristic movements, but there is, in general, no analysis of types of human behaviour into types of bodily movement. The categories of human behaviour and of bodily movement divide the world in ways that are incommensurable. Identical bodily movements may constitute different actions; different bodily movements the same action.

Some actions are named by reference to particular bodily movements. For example, raising one's right hand may be an action, what one intends to do. 'Raising my hand' would then be the description under which the agent intended to act. This is what Hornsby means, I think, by saying that the 'two concepts of behavior overlap'. But this is not <u>generally</u> so. Functionalism supposes a <u>general</u> coincidence of the two concepts; of the categories of human behaviour and bodily movement; or at least some systematic correspondence between them. There is none, so the functionalist thesis is false. Analogous points concern perception and sensory stimulation, as I will explain.

If an action - a piece of behaviour - consists in making a bodily movement, in a certain context, or is a (causal) result of such a movement, then the physiological causes of the movement are causes of the action, in that context, and so in <u>some</u> sense explain the action. As Fodor has remarked,

'nothing shows that a causal explanation may not be provided for each behavior [ie. each particular action] under some description.' (Fodor 1968a:47) My argument will therefore be, in effect, that this is not explanation in the relevant sense, so as to be equivalent to, or explicate, "belief-desire explanations" and their conceptual kin. This is because the non-coincidence of physical movement-types, on the one hand, and behaviour or action-types on the other, means that no explanatory generalisations of one sort are equivalent to explanatory generalisations of the other sort. But functionalism, and the computational theory, assume the coincidence or equivalence, or at least some systematic correspondence of, the two sorts of generalisation.

3.3 THE IDEA OF FUNCTIONALISM

The basic idea of functionalism is that the identity or type of a thing - what kind of thing it is - is determined by its causal role in a system or mechanism: its place in a system of causes and effects. A heart is an organ that pumps blood around the body of an animal. Shape, substance, evolutionary history, are irrelevant to the identification of an organ as a heart. If an organ has the function of pumping blood, it's a heart. A transformer is a device that changes the voltage of an electric supply; given an input voltage, it delivers an output voltage, in a certain ratio to the input. That's what it is to be a transformer. The material of which

it is made doesn't matter, nor does its design - as long as it works, performs its function. (Of course, a transformer can fail to function correctly: its defining function is the causal role it's supposed to have.)

The functional theory of mind says that the same is true of such mental phenomena as belief, intention, wishes, etc. A belief is a 'mental state' with a certain (kind of) functional role in a person's mental economy, ie. a functional role defined in terms of other mental states and the person's interactions with the world. This functional role is implied by the ordinary explanations we give of each other's mental lives and behaviour, or by generalisations from which we infer such explanations.

It is also claimed by functionalists[22] that this explains the 'content' of particular beliefs, and other kinds of mental state, such as the belief that it is raining, the desire for a drink, etc. The belief that it is raining, for example, would be a mental state with a particular functional role definable in terms of other mental states (beliefs, desires, etc.) and experiences and behavior. Someone believes that it is raining if he has such a 'state', ie. a state - an internal item - with such a functional role. Crudely

---

[22] eg. Loar 1981, whose view I will consider in detail below. The 'conceptual role' theory is a variation of the functionalist explanation of the 'contents' of mental ('propositional attitude') states, eg. Block 1986. Fodor 1987, 1990, holds that kinds of mental state, such as believing, intending, wishing, hoping, etc., are distinguished by their functional roles, but that their contents are not to be explained this way.

speaking, a belief that it is raining is an internal state, property or process of a person that is caused by perception of rain, reports of rain, and the like; and causes the desire to stay indoors, the act of taking an umbrella when obliged to go out; and so on. (The last clause is, of course, a source of difficulties.)

Functionalism derives its plausibility from thoughts such as this: Beliefs often result from perception, and have characteristic effects on behaviour. Beliefs can be thought of as dispositions to behave in certain ways, or as the effects of perceptual experience, or some combination of these. The belief that it will rain, for example, might be explained in terms of its characteristic expressions: the disposition to wear a raincoat or take an umbrella, to say certain things, and so on. This would be a behaviorist analysis. Functionalism, it's often pointed out, is a development of behaviorism, adding an internal dimension to behaviourism's stimulus-response analysis. If a disposition is thought of as an underlying state (a brain state, for example), the belief might be thought of as the state that causes these kinds of behaviour. But as critics of behaviourism pointed out, what behaviour results from a particular belief depends on what one wants, and what other beliefs one has. Similarly, what belief, if any, results from a certain sensory experience may depend on other beliefs. So if the possession of a particular belief is to be identified

with a state that tends to cause certain kinds of behaviour - and tends to be caused by certain kinds of perception - the explanation of these tendencies will have to be given in tandem with explanations of other states, kinds of behaviour and perception. In short, such explanations as 'The belief that it is raining is the state that tends to be caused by such-and-such perceptions, and tends to cause such-and-such other states and kinds of behaviour', will have to be combined into an overall explanation of all the mental states a person might have - since this belief might interact with any other beliefs, desires, etc. - and all the kinds of perception and kinds of behaviour with which these could be causally related. Mental state-types must be defined all together. This is sometimes called the thesis of 'holism' of the mental. The overall explanation implied is referred to as a 'psychological theory'. It will consist of, or entail, counterfactual-supporting generalisations, or laws, of the relations between types of mental state, perception and behaviour, such as 'If A believes that it will rain, and intends to go out, and wants to stay dry, and believes that umbrellas keep the rain off, ... , then she will take an umbrella'.

3.4 FUNCTIONALISM AND THEORIES

A theory, according to functionalism, determines the functional roles of the theoretical entities it posits. A functional role is defined by actual and possible relations of

a theoretical entity with other entities similarly posited by the theory, and with the non-theoretical phenomena the theory mentions. These relations are specified by the theory. Mendel's theory of heredity fits the picture.

Mendel developed a theory of the causes of observable characteristics of plants and animals, which introduced and implicitly defined the concept of a gene. A gene, for Mendel, was a property of gametes that mediated the characteristics of parents and offspring, in accordance with a certain calculus (interactions of dominant and recessive characteristics, etc.). The physical stuff that fills this functional role, the biochemistry of genes, was not discovered until the 1950s, a hundred years after Mendel's hypothesis. He knew something must have this function, but he had no means to discover what it was. (Organisms of another world might have genes made of different stuff; the functional role defined by Mendel might be filled in their case by XYZ, rather than DNA.)

According to functionalism, in giving everyday explanations of human thought and behaviour, such as 'K believed he was being followed, because he saw a man waiting outside his house', and 'J carried an umbrella because she expected it to rain', we employ or imply a psychological theory, in the same way that Mendel's explanations of pea flowers employed a theory. These theories make use of 'theoretical terms'. In Mendel's case, the significant theoretical term in explanations of the physical

characteristics of individual plants and animals was 'gene' ('Erbfaktor', in German). In the 'psychological' case, according to functionalism, the theoretical terms are 'believes', 'wants', 'intends', and similar expressions - or their nominal forms, 'belief', 'desire', etc. The theories implicitly define the theoretical terms, by relating them to non-theoretical expressions or descriptions: descriptions of flower colours, or more generally of the characteristics of plants and animals, in one case; of human behaviour and perception in the other.[23]

This idea can be put in a material, rather than linguistic mode: The theories describe the relations of theoretical entities - genes, in one case; beliefs, wants, intentions and so forth, in the other - to other (non-theoretical) phenomena - flower colours and other characteristics of plants and animals in the first case; human behaviour and perception in the second. Genes are those entities which are related to the characteristics of plants and animals as Mendel's theory claims. Beliefs, intentions, and other mental phenomena are those entities related to the phenomena of perceptual and behavioural events as the theory of 'folk psychology' claims. The theories explain the real natures of the sorts of entity in question - genes, in one case, beliefs and desires, etc., in the other.

---

[23] This is articulated by Lewis 1970, 1972, developing the view of Ramsey 1929.

Functionalism interprets psychological (eg. 'belief-desire') explanations of behaviour as hypotheses about the internal mechanisms that produce the behaviour. Explaining human behaviour, from this point of view, is a kind of 'black box' problem: we know the external effects, and have to infer the internal causes. 'Behaviour', in this model, is external; mental states and events are internal; and since we only see the external behaviour, and don't normally look inside, the mental states are theoretical posits to explain what we observe.[24] It is therefore important to consider how the non-theoretical phenomena would be characterised in such a theory.

3.5 WHAT PSYCHOLOGICAL THEORY?

Functionalism claims that explanations of behaviour and of 'mental states' such as belief, are analogous to Mendel's explanations of the phenotypes of plants and animals. As Mendel's explanations implicitly defined the functional roles of genes, so the explanations of beliefs and other 'mental states', and of behaviour, provided by common sense or 'folk psychology', implicitly define the functional roles of beliefs, intentions, and wishes. As Mendel's explanations

---

[24] Lewis (1972) distinguishes 'theoretical terms' ('T-terms') and 'O-terms' (for 'other' terms - not observational). This allows 'O-terms' to include eg. mathematical vocabulary which may not be straightforwardly 'observational'. The crucial point is that the 'O-terms' must be logically independent of the 'theoretical', in this case 'psychological', concepts.

comprised, or derived from a theory, so common sense psychological' explanations comprise, or are expressions of, a theory of human behaviour and psychology. Loar explains:

> The hypothesis is that our mastery of attitude-ascriptions involves a certain theory that consists at least in: (a) input generalisations relating perceptual circumstances to beliefs, (b) internal constraints of rationality on beliefs, and (c) output generalisations relating beliefs and desires to actions. (Loar 1981:9)

In explanations of functionalism, the idea of a psychological theory is often sketched in terms of 'mental states' - beliefs, intentions, desires, hopes, fears, expectations, and so on; and their relations with 'sensory inputs' or 'stimuli', on the one hand, and 'behavioural outputs' or 'motor responses', on the other. Particular explanations of mental states and of behaviour are assumed to be given in such terms, and so also the generalisations or laws that constitute the 'theory'. Lewis (1972), in an influential formulation of the functionalist thesis, gives this sketch of the psychological 'theory':

> Think of common sense psychology as a term-introducing scientific theory, though one invented long before there was any such institution as professional science. Collect all the platitudes you can think of regarding the causal relations of mental states, sensory stimuli, and motor responses. Perhaps we can think of them as having the form: When someone is in so-and-so combination of mental states and receives sensory stimuli of so-and-so kind, he tends with so-and-so probability to be caused thereby to go into so-and-so mental states and produce so-and-so motor responses. Add also all the platitudes to the effect that one

> mental state falls under another - 'toothache is a kind of pain', and the like. Perhaps there are platitudes of other forms as well. Include only platitudes which are common knowledge among us - everyone knows them, everyone knows that everyone else knows them, and so on. For the meanings of our words are common knowledge, and I am going to claim that names of mental states derive their meanings from these platitudes. (Lewis 1972:256)

Putnam, in a similar vein, refers to 'sensory inputs' and 'motor outputs' (Putnam 1975a:433). But ordinary explanations of behaviour, belief, and so on, are not given in terms of 'stimuli' and 'motor responses', but rather in terms of perception and action. To say that someone sees something, in explanation of a belief she acquires or something she does, is to say something about her mental life. 'Leigh stood up because she saw the bus' says that she noticed it, recognised it as a bus, took in the fact that a bus was there. This attributes to her a cognitive response to the presence of a bus; it says that some mental state or event occurred - recognition of the bus, or belief that there was a bus, or something of the kind. This is not to say that all seeing is 'seeing that'; but the reference to perception in explanation of mental states such as belief is 'epistemic' or 'cognitive'.

Likewise, ordinary belief-desire explanations explain behaviour, not 'motor responses'. 'Leigh took the bus because she wanted to go across town' explains an intentional action, taking a bus, not - or not directly - any particular bodily movements she made in doing so. I will discuss perception and behaviour in turn, and then contrast them with 'input' and

'output'.

3.6 FUNCTIONALISM AND PHYSICALISM

Functionalism is sometimes said to be in principle 'topic neutral' (eg. Armstrong; Putnam): The functional roles defined by the relevant theory could be filled by entities of any kind, physical or not; if physical, then mechanical, hydraulic, biochemical or electronic. In this sense, Mendel's genetic theory was also in principle 'topic neutral': genes could have turned out, perhaps, to be non-physical; or if physical, biochemical, mechanical, hydraulic or electronic.

Another way of making the same point is to say that, in other creatures, perhaps the fauna and flora of another planet, genes as defined by Mendel's theory might be made of a different stuff. Or we can imagine artificial plants and animals, capable of reproduction, but made of synthetic materials, in which the genetic material consists of some silicon-based variant of DNA, or some structurally dissimilar compound, or electromagnetic states of semiconductors.

Similarly, the functional explication of mental states might apply to other intelligent beings, whose constitution might be quite different from ours; even, it's been suggested, non-physical beings.[25] But in our case, we know the stuff is

---

[25] This assumes that the same, or at least corresponding, notions of perception and behaviour - or 'input' and 'output' - would be applicable to these beings as to us, which is not obviously coherent, especially in the case of non-physical beings.

biochemical, complex hydrocarbons and such like. So 'mental states' will turn out to be physically realised, by some properties of the organic material that we're made of.

Which is not, of course, to assume 'type-physicalism'. Some functionalists (eg. Smart, Armstrong) held that the doctrine was at least compatible with the thesis of materialism, type-identity of mental and physical phenomena, ie. that types of mental state or event might be (contingently) identical with types of physical, presumably neural, state or event. Believing that it's raining, for example, might always consist in having a certain configuration of neurons activated. Others (eg. Putnam, Fodor) argued that functionalism shows that type-identity is false: mental states and events are functional states or events of individuals, which may be, and probably are, realised by different physical states or events in different individuals (human or not). But these functionalists too assume that functional states are physically realised, at least in humans. This implies 'token-identity': mental states and events are identical with, or realised by, physical states or events, although the physical states that realise a type of mental state may vary from species to species, individual to individual, or from time to time in one individual. In this sense of 'physicalism', then, functionalism is a physicalist doctrine. The point of functionalism is to explain how physical creatures can have mental states, without resorting

to Cartesian dualism.

I think, in fact, that functionalism is the most plausible version of physicalism, for the following reason. Assume physicalism to mean at least this: Each particular 'mental state' or 'propositional attitude' had by an individual consists in a particular physical state or event - the thesis of 'token identity', which seems to be the highest common factor of physicalist doctrines. So when a person believes that it is raining, there is some state or event in her brain that constitutes that belief: a group of neurons in a particular state of excitation, for example, or a certain pattern of electrical activity in some region of the brain, or a distributed electrochemical property of some kind.

Now it may be asked what makes this state or event or property the belief that it is raining: what is it for such a state or event to be the belief that it's raining? The answer can't be merely that this event occurs when the person believes that it will rain, nor even that it <u>invariably</u> occurs when she holds this belief, for such co-variance might hold because of a causal connection between the belief and the event, so that the event is a symptom or expression of the belief. For example, A might have a distinctive twitch that always occurs when he believes that it is going to rain. But the twitch surely does not <u>constitute</u> the belief. What's needed is evidently that the right sort of dependence should hold between the state that constitutes the belief and its

causes, or its expressions, or both. The state should underlie, in some relevant way, such events as the avowal of the belief, or its manifestation in non-verbal behaviour. But this is just the first step in the argument for functionalism. The next is to recognise that the expressions of a belief are dependent on the belief <u>together with</u> other beliefs and preferences, and so on.

The psychological theory implicit in explanations of behaviour, according to functionalism, defines the <u>functional roles</u> of mental states and events. Beliefs are those mental states related to other mental states and to perception and behaviour in the way described by this theory. The belief that it will rain is that mental state related to other mental states and to perception and behaviour in the particular ways described by the theory - or by whatever psychological theory is true of believers. These counterfactual relations constitute the functional roles of the mental state-types in question.

Genes turned out to be realised by - to consist in - DNA molecules. DNA molecules fill the functional role of genes; they have the causal relations specified by genetic theory. Similarly, mental states and events are supposed to be realised by physical - neural - states and events, in virtue of the counterfactual <u>causal</u> relations these have with other neural states and with 'inputs' and 'outputs'. As Block explains:

> Metaphysical functionalists characterize mental states in terms of their causal roles, particularly, in terms of their causal relations to sensory stimulations, behavioral outputs, and other mental states. (Block 1980:172)

So the functional roles of mental states and events are the causal roles of neural states and events; thus neural states and events are mental states and events. The internal and external dependencies of mental states, described by the psychological theory, are realised, in us humans at any rate, by the causal connections of our neural processes. The connections among mental states, such as one belief resulting from another by inference or association, or an intention arising in the course of practical reasoning, are realised in the causal connections by which one neural impulse excites another, or one wave of excitation causes another; or the connections of some other kind of brain process to be discovered by neurophysiologists. 'Inputs' and 'outputs' will likewise consist in causal connections between neural processes and some physical objects or events or properties.

Functional and causal roles consist of counterfactual relations, ie. the causal relations physical states or events would have with other physical states or events. This counterfactual character of the role-defining relations means, as Loar explains, that functional roles are filled by physical state-types, in an individual at a certain time.[26] The

---

[26] Loar 1981:62; also Schiffer 1987:20-22.

causal roles of (types of) neural states consist in their counterfactual connections with other (types of) neural states, and also with (types of) events - or objects or states of affairs - corresponding to behaviour and perception. A functional role may be filled by more than one physical state-type (hence the possibility of 'multiple realisation'); and a physical state-type may fill a functional role in an individual at one time, but not in another individual, or in the same individual at another time. 'Token physicalism', together with functionalism, thus implies a kind of limited 'type-physicalism'. But the principle advantages advertised for functionalism over simple physicalism, in this regard, are preserved: 'multiple realisability', hence 'non-chauvinism'.

Functionalism can be thought of as claiming that two theories are isomorphic. One is the psychological theory, specifying relations among mental states, perception and behaviour; the other is a causal-functional theory of the physical system consisting of the central nervous system and its 'inputs' and 'outputs'. To put it another way, the network of mental states, perception and behaviour, is isomorphic or congruent with the causal network of the CNS and its inputs and outputs, under some functional description of the latter. The congruence, according to functionalism, is guaranteed by the identity of the two systems. A psychological theory, in this view just is a functional theory of the causal system of the CNS etc., expressed in a

mentalistic vocabulary.

To make sense of the idea of mental states being physically realised, behavioural output must be conceived as consisting in causal relations between neural states or events, and physical events that are their effects (Davidson's picture of reasons as causes of action); and input must be conceived as consisting in causal relations between physical objects or events - 'inputs, or distal objects and their properties' (Schiffer 1987:20) - and the neural states they cause. As with the internal relations among neural states, these connections must be counterfactually defined: the 'causal role' of a (type of) neural state or event consists in its actual and possible causes and effects. So the causal relations in question are relations between (types of) neural state or event, and (types of) objects or events which are their causes and effects.

3.7 BEHAVIOUR

We often explain things people do by reference to their beliefs, desires and other mental states. The explananda here are not normally 'motor responses', but actions. For example, I go to the store because I want to buy milk (and believe that I can buy milk at the store, etc.); I carry my umbrella because I believe it is going to rain. Going to the store, or carrying an umbrella, involves bodily movements, but it does not follow that these movements are explained by the ordinary

reference to my belief and desire. Typically, but not invariably, the explanandum is an action that involves things in the person's environment:

(1) L bought the newspaper because she wanted to find out what happened in the world.

'Buying a newspaper' describes an event that involves another person, a newspaper, money to be transferred, as well as the person who performs the action.

(2) D flipped the switch because he wanted to turn on the light

D's flipping of the switch involves at least a switch, as well as the agent D.

Sometimes the action may involve no person or thing other than the agent:

(3) She clapped her hands because she wanted to express her appreciation of the performance.

But most action is world-involving: the description under which it is explained involves reference to things or other people involved in the action. Such explanation gives the agent's reasons for performing the action. The explanation is

relative to a particular description of the action, because, as Davidson points out, 'reasons may rationalize what someone does when it is described in one way and not when it is described in another' (Davidson 1963/1980:5). The truth of (3) does not imply the truth of

(4) She caused her neighbour to have a heart attack because she wanted to express her appreciation of the performance,

even if her clapping was the cause of the heart attack. Nor, and for the same reason, does it imply:

(5) She moved her hands at such-and-such velocity with such-and-such trajectory because she wanted to express her appreciation of the performance.

Davidson's 'belief-desire' analysis of explanations of behaviour brings out why this is so. Davidson suggests that such explanations imply the existence of a primary reason why an agent did something, which consists of a pair of a 'pro attitude' and a 'related belief', which he claims is the cause of the action. So the primary reason corresponding to (3) would be:

(6) She clapped her hands because she wanted to express her appreciation of the performance, and believed that she would

do so by clapping her hands.

The explanation would be shown to fail if it turned out that she did not believe that clapping would express her appreciation, or something similar. But even if her clapping and her causing the heart attack are identified as the same action, in some sense, it does not follow that

(7) She gave her neighbour a heart attack because she wanted to express her appreciation of the performance, and believed that she would do so by giving her neighbour a heart attack;

because, it may be supposed, she had no such belief. Likewise, and for an analogous reason, it does not follow that:

(8) She moved her hands with velocity x and trajectory y because she wanted to express her appreciation of the performance, and believed that she would do so by moving her hands with velocity x and trajectory y.

In general, alternative descriptions of an action cannot be substituted in 'belief-desire' explanations. This point is significant, because functionalism, as I shall explain below, is constrained to a narrow conception of 'behaviour', roughly that of bodily movement. Even if an action can be identified

with the bodily movements involved, these considerations show that the agent's reason's for performing the action do not thereby explain his making those bodily movements. Buying a newspaper involves making various movements, and uttering various noises. It can be said, perhaps, that these movements and noises, in the context, <u>are</u> buying the newspaper. But

(9) L made such and such movements and noises because she wanted to find out what happened in the world

is not an explanation of the relevant kind (perhaps not of <u>any</u> kind), unless she believes that by making such and such movements and noises she can find out what happened in the world. Perhaps this explanation would have to be further elaborated by the ascription of the belief that these movements and noises constitute, or would result in, buying the newspaper. Functionalism is committed to the claim that <u>all</u> explanations of action are essentially enthymemes of this sort: they all have bodily movements as their primary <u>explananda</u>, which are believed by the agent to be means to the performance of actions in the usual sense. (eg. Loar 1981:88; discussion below.)

We don't normally think about the bodily movements we make in performing actions. We don't, typically, have either beliefs or intentions, by normal standards of ascription, concerning the bodily movements we make in acting. (Ordinary

behaviour may differ in this respect from dance or mime.) The movements necessary for the performance of ordinary actions are often quite complex and specific, and have to be made with a high degree of precision. For example, I go to the store by making a series of specific movements of my legs (to simplify). The exact nature of these movements depends on the number and size of the stairs I have to descend, the distance along the street to the store, the people and other obstacles I have to avoid, and so on. A 'belief-desire' explanation of my making these movements would have to refer not only (if at all) to my desire to buy milk and my belief that they sell it at the store, but also to beliefs about the location of the store, the heights of the steps, and so on - a great deal of very specific information about the physical environment. The 'belief-desire explanation' required by functionalism would be something like:

(10) I make such-and-such movements because I want to buy milk and believe that I can do so at the store, and believe that by making these bodily movements I will get to the store.

Loar calls such beliefs 'instrumental beliefs', and claims that other kinds of belief only influence action by the meditation of these (ibid. 87). But ordinary explanation of behaviour - 'folk psychology' - would not countenance this explanation, because I do not, in any ordinary sense <u>believe</u>

that by making these movements I will get to the store. I have no ordinary beliefs about the movements necessary to get to the store. These movements depend on facts about the environment - the size of the stairs, etc. - concerning which I have no beliefs of the necessary specificity, in any ordinary sense. If asked, for example, why I raised my leg by just so much in climbing from one step to the next, I could not say that I did so because I believed (or knew) that the step was such-and-such a height.

A cognitive or computational psychologist may want to claim that I do have some belief about, or representation of, the height of the step; that I must use this representation in computing the movement I make. But common sense, manifested in ordinary psychological explanation, does not ascribe such beliefs. Even if there's some sense in which it's true, it's no part of the explanation of my going to the store by reference to my desire for milk. That explanation says nothing about the bodily movements I must make. Nor does any ordinary explanation. There are no 'platitudes' concerning detailed beliefs about the relations of bodily movements and the physical environment. If a psychological model hypotheses such beliefs, it stands in need of a theory - such as functionalism, perhaps - to explain and justify the hypothesis.

The irrelevance of motor responses or bodily movements to most ordinary psychological explanations is clear when we

consider the possibility of generalisations, anything that could answer to Lewis's request for 'platitudes'. If there are any commonly known generalisations concerning the relations of mental states and the things people do, these would not, in most cases, say anything definite about bodily movements. Suppose one such generalisation is 'If a person wants to buy milk, and believes that he can do so at the store, ... , then he will go to the store.' But there are innumerable physical movements, with no common description (other than 'going to the store') that could constitute going to the store. No description of movements could complete the generalisation just suggested.

Sometimes a particular bodily movement <u>is</u> the intentional action explained by reference to beliefs and desires. For example, the conductor makes a particular gesture with her left hand because she wants the violins to play quietly. This may be an example of the 'overlap' of the two notions of behaviour mentioned by Hornsby. (Even so, the relevant bodily movement-description may not be specific enough to satisfy a 'motor response' model of behaviour.) But this is not generally the case. If 'folk psychology' rests on or implies generalisations (or laws) relating mental states to things people do, they will not usually have motor responses or bodily movements as <u>explananda</u>.

Particular belief-desire explanations normally explain actions without reference to the specific bodily movements

that are involved in the actions; and generalisations, if there are any, would relate types of mental state to types of action the instances of which may have no common physical character. For example, when a person wants to take a taxi, and sees one approaching, she will signal to the driver. There are innumerable ways of doing this, more or less typical. There's no reason to think there's any bodily movement-description, other than an open-ended disjunction, that applies to all these particular acts; but all fall under the single description 'Signalling to the taxi driver', and it is as such that the belief-desire explanation applies to them. This is so even in a case for which there are characteristic gestures (so that it would be possible to recognise a mime hailing a taxi). Many kinds of action are achieved by far more diverse bodily movements.

The 'psychological theory' functionalism appeals to as implicitly defining the 'psychological' vocabulary - or as determining the functional roles of psychological states - must deal in <u>generalisations</u>, explicitly or by implication. If they are to be derived from common sense explanations - 'folk psychology' - they will concern the typical causes of types of <u>action</u>, not types of bodily movement, except where bodily movements are the actions we intend to perform.

An action, as explained by an ordinary 'belief-desire' explanation, is something done intentionally, deliberately. The intention with which such an action is performed is an

essential aspect of the action: without it, the action does not occur, or a different action occurs. This is obvious in such cases as lying, promising, and requesting. If there's no intention to deceive, it's not a lie, for example; and a wholly different explanation of the behaviour would be called for.

Dretske (1988) argues that an action is a complex of the intentional cause of a bodily movement, the movement, and (some of) its effects. The intentional cause is included to distinguish eg. a rat's moving its paw from its paw moving (by some other cause, such as an external agency or a reflex). I don't want to endorse the causal-chain picture in Dretske's analysis; but I want to use the distinction between a movement which is an action, and the same movement not performed intentionally or with a different intention, hence not an action or a different action. The explanation

(11) L raised her hand because she wanted to ask a question,

explains L's deliberate act of raising her hand, not merely her hand going up. The explananda of ordinary belief-desire explanations, I claim, are always intentional actions, not mere bodily movements.

Actions, as the explananda of ordinary, folk-psychological, belief-desire explanations, are WORLD-INVOLVING and INTENTIONAL. They are typically described in abstraction

from the bodily movements by means of which they are performed, and the explanations apply to them under such non-bodily descriptions. Bodily movements are not normally the objects of beliefs and intentions; there are typically no folk-psychological explanations of bodily movements, described as such. If there are generalisations assumed or implied by, or derivable from folk-psychological explanations, they will not, typically, concern bodily movements, but action-types, which may be realised by physical movements having no common description (except as instances of that action-type).

3.8 OUTPUT

What would the 'output' generalisations of the psychological theory be like? Presumably, they would be conditionals with descriptions of mental states in the antecedent, and of behavioural effects in the consequent. The intuitive idea would be something like this:

If D wants to turn on the light (and believes that flipping the switch will turn on the light), then he will flip the switch.

That is, we would <u>generalise</u> from the particular belief-desire explanations we give of behaviour. (Wanting to turn on the light is that state which, inter alia, causes switch-flipping.)

But functionalism cannot use a notion of 'behaviour' that retains the intentional aspect of the ordinary conception, since the intentional element or aspect is to be explained by functionalism in terms of 'input', internal connections, and 'behavioral output'. If 'behavioral output' is understood as intentional action, a functional theory would assume part of what it is supposed to explain. So 'output' cannot be described as, for example, flipping the switch, in the normal sense in which this description occurs in 'He flipped the switch because he wanted to turn on the light'. Nor can it be described as, for example, moving one's fingers, as in 'He moved his fingers because he wanted to flip the switch'. Output would have to be described instead as something like the movement of his fingers, with no implication that he moved them. The description must be non-intentional. (The idea, of course, will be to explain intentional actions as those with the right kinds of causes. But 'output' is what is caused, and must therefore be distinguishable from its - perhaps intentional - causes, for the purpose of functionalism.)

'Output' in a functionalist theory must be understood as no more than bodily movements - not behaviour or action in the 'wide' or world-involving sense. Thus an 'output' might be described as moving one's hands in a certain way, but not as grinding coffee. This is because grinding coffee involves, in addition to the hand movements, environmental conditions which are independent of the agent's mental states: the arrangement

and properties of coffee beans, grinder, and so on. This restricted conception of behaviour, common to contemporary 'causal' theories of action, is sketched by Davidson:

> Our primitive actions ... mere movements of the body - these are all the actions there are. We never do more than move our bodies; the rest is up to nature. (1980:59)

'He flipped the switch' describes an action in terms of events in the world beyond the agent's body, the <u>effects</u> of a bodily movement. But the action, in this sense, that results from any combination of 'mental states' depends in part on the way the world is. Flipping the switch won't result from the belief and desire, or even the <u>attempt</u>, if the switch is stuck fast, for example. In general, the world has to cooperate; and in particular, the 'instrumental belief' (eg. that the switch will flip if it's pressed) has to be true, for the intended action to result. Functionalism is not entitled to *assume* these conditions, for the following reason.

'Outputs' are supposed to be the effects of neural processes. The causal role of such a neural item - hence what mental state or event it realises - consists, in part, in its actual and counterfactual relations with its 'behavioural' effects. A functional theory of the agent would specify (or entail) the effects a neural state would have in combination with any other possible neural states with which it might interact. (This is an aspect of the 'holism' of functional explanation.) If it's a desire for water, for example, it would have one effect if another state of the agent was a

belief that there's a glass of potable water in front of him; but a different one if he believes (ie. has a neural state that is the belief) that the glass contains vodka.

But the functional theory would entail the possible effects of a neural event on the world at large - the local environment, say - only if it also took account of all the possible states of that environment. A neural state that realises a driver's desire to stop quickly, together with one that realises her belief that she can do so by pressing her foot on the brake pedal, will have one effect if the car's braking mechanism is in working order, but another if the cable has broken.

In general, for a functional theory to entail the world-involving effects of the agent's mental - ie. neural - states, it would have to be a theory of the world, or at least all possible local environments, as well as the agent. it would have to specify the effects of neural states <u>conditionally</u>, by reference to (all) possible environmental conditions. This, I take it, is an absurd requirement for a <u>psychological</u> theory. So the causal roles of neural states and events will be defined by the functional theory only in so far as they involve those effects which are independent of environmental variables.

The output-conditionals of functionalism must therefore relate mental states to things that can be done - or that will happen - whether the world cooperates or not; whether the

instrumental beliefs are true or false. Loar calls such effects 'basic actions':

> Certain actions are the primary explananda for belief- and desire-ascription - bodily movements, and intentional mental acts, which I shall not discuss. Consider a non-basic action like flying to Abu Dhabi. Although the intention to fly to Abu Dhabi occurs in its explanation, given enough information such an explanation can be factored into two components: (1) actions like boarding the plane and staying aboard, whose explanation contains the intention to fly to Abu Dhabi, certain beliefs and intentions to do those specific actions, and (2) independent facts, like the fact that the plane flew to Abu Dhabi. This division of explanatory labor ends with bodily movements, basic actions that are not thus explained by further actions and independent facts. Since the independent facts are not themselves explained by those intentions, save per accidens, the primary explananda of the belief-desire theory are bodily movements..... The intentions that explain them ..... have those bodily movements as their objects. They are "willings". (Loar 1981:88)

'Basic actions' are things that a person is simply able to do, if he 'wills' to do so, and there are no 'external impediments'. This category, according to Loar, 'is virtually restricted to bodily movement types' (90).

Counterfactual conditionals relating internal physical states and bodily movements would thus form the output-aspect of the description of causal roles of the internal states. Suppose the functionalist thought to include also the further effects, such as the switch being flipped, in this description. This would be in keeping with Harman's idea of 'wide functionalism':

> I claim that psychological explanations are typically wide functional explanations. That is, I claim that such explanations typically appeal to an

> actual or possible environmental situation of the creature whose activity is being explained. A narrow functional explanation appeals only to internal states of the creature and says nothing about how the creature functions in relation to an actual or possible environment. (Harman 1988:11)[27]

I agree with Harman that 'psychological explanations are typically wide ... explanations'. This was the point of my 'world-involving' condition on the descriptions of behaviour in ordinary belief-desire explanation. The question is whether functionalism can respect this condition.

The 'wide' effects of a neural state-type - that which the theory seeks to identify with the 'primary reason' for flipping the switch, for example - are dependent on conditions independent of the agent's mental, or neural, states - the condition of the switch, and so on. (The environmental facts might be explained in turn by psychological ones, if they are a result of human design; but not the same psychological facts. The properties of the switch may be the result of intentional states of its designer, but they are not a result of the mental events responsible for the particular act of switch-flipping.) In some circumstances the effect would not be switch-flipping.

A causal role account would describe the effects of neural states - narrow or wide - in physical terms. If they sometimes cause events which are, from the intentional point of view (ie. relative to the agent's intentional states)

---

[27] Also Kitcher 1985.

unsuccessful, these are, from the causal point of view, just effects like any other. The causal role of a neural state includes its 'unsuccessful' effects. So the deviant cases cannot be distinguished from the 'successful' cases, for purposes of a causal-role description, without reference to the content of the primary reason. But that reference would be question-begging, because the causal role theory is supposed to explain the mental state-identities of neural states, eg. the content of 'primary reasons', in the functionalist account.

From the point of view of a strictly causal account, the 'unsuccessful' cases are indistinguishable from the 'successful'. The causal role description would have to include counterfactual conditionals relating the internal states to their effects, including the unsuccessful ones. But what is common to the successful and unsuccessful attempts, to all effects of a particular internal state (or states), is just the bodily movements. The driver's desire to stop the car - and related belief - result in her foot hitting the brake pedal, for example, whatever happens next. (When the car fails to stop, she will acquire new beliefs and desires, and further action will result; but that's another story.)

By contrast, a factory robot, of the kind that performs one function on a production line repeatedly - bolting the left rear door on cars, for example - does have a regular correlation between the causes of its 'behaviour' and their

effects. The difference is explained by the uniformity of the robot's environment, compared with the endless variability of human environments. 'Wide functionalism' may be true of factory robots.

On the other hand, there won't be generalisations about desires and intentions to turn on lights, for example, and bodily movement-types, because there are no bodily movement types common to attempts to turn on the light. So functionalism must suppose that all actions are ultimately performed by the mediation of beliefs, desires or intentions - or as Loar calls them "willings" - whose objects are bodily movements. These are things the agent can do without threat of the world's non-cooperation. They are, in Loar's theory, the objects of basic belief-desire explanations. The further effects of bodily movements - such as grinding coffee - are the objects of beliefs, desires and intentions which influence behaviour only by the mediation of 'basic' beliefs and desires and their objects, bodily movements. In this way, 'J grinds the beans because he wants to make coffee', is a sort of enthymeme. The full explanation would be something like, 'J makes such and such bodily movements, because he wants to make coffee, and believes that by making these movements he will grind the beans, and believes that grinding the beans is a means to make coffee'.

I think the hypothesis of ubiquitous beliefs, intentions or 'willings' with bodily movements as their objects is a

dubious piece of psychology. It's not folk-psychology. But in any case, it leaves all the beliefs, desires and intentions whose objects are not mere bodily movements - most of mine, for example - in need of explanation by some other part of the functional theory. 'Behaviour', in the restricted sense, fails to bring any worldly things into the picture.

Hornsby questions whether even 'bodily movements' should be the end of the analysis of behaviour in the functionalist's theory. Her thought is that the bodily movements that result from brain states and events depend on such facts as the state of the person's arms and legs, which might seem to be 'independent' of the psychological facts - beliefs, intentions - or the physical (presumably brain) states that realise them. The belief-desire pair that is, according to Davidson, the cause of flipping the switch, or the 'willing' posited by Loar, will only result in a switch-flipping sort of bodily movement if the physiological state of the agent's hand cooperates. The reasons to retreat from ordinary actions to bodily movements in functionalism's output-conditionals, seem to be reasons equally to retreat from bodily movements to their physiological causes, perhaps efferent nerve impulses. But it's even less plausible that agents have beliefs and intentions concerning these. This raises questions about the distinction of the psychological and the non-psychological within the individual, which I shall not try to resolve here. Suffice it to say that bodily movements are the most the

functionalist can claim as 'behaviour' or 'output' for his theory.

So the notion of 'behavioral output' available for functionalism will be one shorn of (a) those aspects of actions that involve anything beyond bodily movement, especially objects and other people, and (b) intentional, or generally mental aspects of actions understood in the normal way. As Loar acknowledges, 'output conditions' on this basis will not serve to individuate beliefs, to explain for example the difference between a belief (or an intention) concerning coffee and one concerning newspapers. Coffee and newspapers are not mentioned in functionalism's reduced notion of 'behaviour'. Nothing is mentioned, except bodily movements, so 'output-conditions' won't do much of the explanatory work in the functional explanation of mental states.

Loar reaches this conclusion concerning the role of 'output' in the functional scheme:

> Finally there is the output condition of the belief-desire theory, which I suggest is simply this: If z wills at t to do A now, z is able to do A at t, and no external condition prevents z from doing A at t, then z does A at t. (Loar 91)

Substitutions for "A", as we have seen, will be descriptions of (at most) bodily behaviour. If beliefs, and the rest, concerning things in the world can be given a functionalist explanation, then, most of the explanatory work will have to be done by 'input conditions'.

3.9 BEHAVIOUR AND OUTPUT COMPARED

Behaviour is not the same as bodily movement. Action-types are not reducible to bodily movement-types. The functionalist notion of 'output' is bodily movement, but behaviour is not the same as bodily movement; the individuation of kinds of behaviour, as actions, is incongruent with any individuation of types of bodily movement - except perhaps by a wholly ad hoc scheme.

Actions typically involve bodily movement. (Perhaps not all behaviour involves bodily movement or change; I leave this as an open question.) An action such as grinding coffee, on a particular occasion, is performed by means of certain bodily movements. But the classification of bodily movements into types of behaviour, or actions, is not made on the basis of the kind of physical movement in question. Classification of movements into actions is quite unsystematic with respect to the physical character of the movements. Different movements are classified as the same (type of) behaviour; and movements that are the same are classified as different behaviours, or actions.

Some kinds of action involve characteristic bodily movements. Washing one's hands, opening a door, slicing bread - these actions typically involve movements that are very similar from occasion to occasion. They can be mimed, and it would be easy to recognise what the mime is doing. But there is no general correspondence of actions and movements, even in

these cases. A door can be opened with very different movements - with the foot instead of the hand, or by backing into it. Other types of action are performed with quite diverse movements: paying a bill, watering a plant, grinding coffee.

I can grind coffee in an electric coffee grinder, or in an old-fashioned manual mill. There's nothing in common to the physical movements I make in grinding coffee these two ways. In fact, the bodily movements I make on two occasions of grinding coffee in the electric mill may be quite different: I hold the mill differently, apply pressure to the relevant part with a different movement of my hand, and so on. (I could hold it between my knees and press it with my head.)

A mime does not open a door, wash his hands or slice bread: he <u>mimics</u> these actions. Performing them requires a door, a knife and a loaf of bread, or soap and water. (Lady Macbeth does not wash her hands in the sleepwalking scene.) Bodily movements are not, in general, actions, because actions are world-involving. For this reason, identical movements can constitute different actions: taking an apple or an orange, hitting a baseball or chopping a tree, grinding coffee or mixing paint.

A type of action, then, is not reducible to, or explicable in terms of, a type of bodily movement, or any range of movement-types. Although many actions involve characteristic movements, there seems to be no prospect of

specifying the range of movements that could realise the action. Actions aren't, in general, individuated according to the movements that realise them. A particular movement could realise any one of a heterogeneous range of actions. There is no general congruence of action-types and movement-types.

Even for those actions which are individuated by reference to a kind of bodily-movement, such as the action of raising one's hand, the action is not the same as the movement. A's hand going up is not the same as A raising his hand. His hand might go up because B raises it. The physicalist, following Davidson, would explain the difference as consisting in the cause of the bodily movement. Raising one's hand, in this view, is one's hand going up with the right cause, namely an intention, or 'primary reason', or some other 'mental state' whose content is related to the movement in the appropriate way. But the functionalist can't use this idea, because the nature of that internal cause is one of the things functionalism has to explain. Its effect, as part of the explanans must be independently specifiable: the hand going up.

3.10 ACTIONS AND BODILY MOVEMENTS ARE NOT 'TOKEN IDENTICAL'

It's obvious that there is not, in general, 'type-identity' of actions and bodily movements. But there are also reasons to doubt that actions can be identified with bodily

movements one by one.[28] Functionalism assumes a view of the relationship of actions and bodily movements that is part of Davidson's treatment of the explanation of actions, in 'Actions, Reasons and Causes'. The idea is that bodily movement-descriptions and action-descriptions are two ways of describing the same events. In this view, an action is (the same event as) a bodily movement, or a series of movements. This might be called a 'token-identity' thesis about actions and bodily movements. Actions and bodily movements are not 'type' identical. But each particular act of opening a door, for example, might seem to consist in a particular bodily movement. With this assumption, it is easy to accept Davidson's suggestion that the agent's reason for an action consists in a physical event, or events, which is the cause of the physical movement that constitutes the action.

The token-identity thesis of actions and movements is especially plausible when we think of cases in which the action is performed by relatively simple movements. Throwing a ball, for example, is done by swinging one's arm in a certain way, and relaxing one's fingers at the right moment. This involves, it seems, one or at most a few physical movements, and it's easy to think of these as being caused by one, or a few, neural events. In fact, of course, a great deal of neural activity underlies such an action: many

---

[28] These doubts are suggested by Stoutland (1985), especially in the last section of that paper, 'The ontology of intentional behavior' (pp54-59).

impulses in efferent nerves from the CNS to many muscles in different parts of the body. So the idea must be that these somehow all flow from one state or event in the controlling part of the brain, and this is the 'reason', or the intention or whatever. The simple picture immediately becomes more complex. We might imagine interpreting processes that mediate the 'reason' and the movements necessary to achieve the intended action. The brain must somehow figure out what movements are necessary, or sufficient, in the present context: the posture of the body, environmental circumstances.

Which movements are identical with the action of throwing the ball? Many bodily movements and adjustments take place in throwing a ball. Should compensating shifts of weight, involving muscle activity in the legs, back, neck - perhaps throughout the body - be included? When a baseball pitcher delivers the ball, his entire body seems to be involved, including the muscles that control his eyeballs. This suggests that the 'bodily movement' that is identical with the action of pitching the ball consists in all the movements of his whole body. (Starting and ending when?) Perhaps this should be thought of as composed of a series of movements, not just one. But pitching the ball is a single action, with, presumably, a single 'primary reason', in Davidson's sense. If all the bodily movements that occur when a person performs an action are to be identified with the action - whether described as one movement or several - cases in which a person

performs two or more actions simultaneously will appear problematic. The baseball pitcher might chew a wad of tobacco as he pitches, for example. Which bodily movements constitute which action, in this case? Some movements may be necessary to enable him to do both things. Which action are they part of?

The uncertainty in the identification of particular bodily movements with the performance of an action is all the greater in less simple or less obviously physical actions: putting out the cat, reading a letter (which involves bodily behaviour - but which exactly?), hiding, standing still.

If a description of bodily movements was given by an observer who had no notion of the actions being performed - perhaps a description of human movements by an intelligent alien - it would surely not individuate movements in a way that corresponds to our individuation of actions. It seems unlikely there would be any way of fitting one scheme onto the other - of identifying discreet sequences of movements, in one scheme, with actions in the other. Not only would a one-to-one correspondence be unlikely, but even a one-to-many. Some movements are parts of more than one action. Actions, as we individuate them, don't always have spatial and temporal limits of the kinds movements have.

I think these considerations cast some doubt on the 'token identity' picture of behaviour and bodily movements - that for every action or piece of behaviour, there is some

movement or series of movements that constitutes the action; or which together with the relevant mental states and events, constitutes the action. To deny this would not be to imply that behaviour consists in something non-physical. It may be that the individuation of actions is more closely related with the intentions of the agent, and with the extra-bodily things the agent's movements affect, than with the movements themselves. In identifying events as actions, we are not, typically, identifying any definite bodily movement or movements.

3.11 PERCEPTION

The actions people perform, and also the things they believe, expect, wish for, and so on, are often explained by reference to the things they see and hear, and sometimes the things they smell, feel and taste.[29] These explanations sometimes refer to an act of perception, as in:

(1) A believed it would rain because he saw black clouds,

and sometimes to the object or state of affairs perceived, as in:

---

[29] I will consider mainly vision, as is common in philosophical discussions of perception. It is also common to remark in footnotes to such discussions that what is said should also apply to the other modes of perception.

(2) B stopped because the light was red.

I want to note two characteristics of such explanations. One is the epistemic or cognitive character of the perception attributed to the person (or implied, in the second example); the other is the reference they make to an object or state of affairs perceived - what I will call the world-involving character of perception. These aspects of perceptual explanations are problematic for functionalism.

Explanations like (1) refer to cognitive states, acts or events. To see black clouds, in the sense of (1), is to notice or recognise black clouds, to see that there are black clouds. Seeing in this sense implies believing or a related, cognitive or intentional state: perceiving or noticing that there are black clouds, or seeing something as black clouds. These are 'mental events' or 'propositional attitudes'. (They are 'intentional' by the criterion of non-substitutivity: 'A sees/ notices/ recognises/ perceives that ...' creates an opaque context in which alternative descriptions of the same thing cannot be substituted salva veritate.)

If A does not believe there are black clouds, he did not notice them or see them in the sense of (1). If he does not believe there are black clouds, (1) would not explain what it purports to explain; it would be false. (1a) is

contradictory[30]; its second conjunct undermines the explanation offered in the first.

(1a) A believed it would rain because he saw black clouds, but he did not believe there were black clouds,

This is sometimes described as 'epistemic perception'. Whether or not there is such a thing as 'non-epistemic perception', explanations like (1) explain by reference to an epistemic variety, perception that involves a cognitive response to what is seen: belief or a closely related epistemic state. Explanations like (2), although they do not explicitly mention an act of perception in the same way, imply that epistemic perception occurred. If B did not see the red light - see that the light was red - then (2) is false. Similarly:

(1b) A believed it would rain because there were black clouds in the sky,

implies that he (epistemically) saw the black clouds, or knew of them in some other way - from another's report, for example.

Common sense, 'folk psychological' explanations of mental

---

[30]Or if not contradictory, it has something like the inconsistency of Moore's paradox: 'I believe p, but not p'.

states and behaviour by reference to perception are typically, if not invariably, like (1) or (2). If there is a 'theory' implicit in such explanation, then, we should expect its generalisations or laws to relate types of perception, in the sense explained, to types of mental state or behaviour. 'If x sees heavy black clouds in the sky, and ....., then he believes that it will rain', perhaps. The lacuna might be filled by reference to beliefs about black clouds as natural signs of certain weather conditions, and such like. I don't mean to imply that the conditional could be completed in any useful way. But if there's any possibility of deriving such 'laws' from common sense psychology, the antecedents should refer to perception in the same sense as in ordinary singular explanations.

To say that someone sees something is also, in the normal case, to say something about the world. 'She saw the bus' implies that there was a bus she saw. (1) implies that there were black clouds, which A saw. (2) implies that there was a red light. When perception is referred to in explanation of what a person believes or expects or does, it is normally perception of something in this sense, which implies the existence of what is perceived. Perception is a relation between a person and an object or state of affairs. 'A saw x' implies both a cognitive or epistemic state of A, and the occurrence or existence of x. As Harman says:

> Ordinary psychological explanations are not confined to reports of inner states and processes.

> They often refer to what people perceive of the world and what changes they make to the world. (Harman 1988:15)

Explanations of belief and behaviour can also be given by reference to what is thought to be seen:

(3) Leigh raised her hand because she thought she saw a bus approaching.

I think that explanations of this form are conceptually dependent on explanations that imply the existence of the perceived state of affairs. But in any case, it is clear that functionalism could not rely on this kind of case as the model of 'input', because thinking (or imagining) one sees something is a paradigmatic intentional, mental state - a 'propositional attitude' - and hence the kind of thing functionalism is to explain, not assume.

The history of philosophy includes many attempts to abstract from perception a class of objects or events independent of the things perceived. In so far as these have been conceived as 'mental' phenomena, such as Cartesian thoughts, Humean impressions and ideas, and 'sense data', they could not be cast by functionalism in the role of 'inputs', without circularity, since the idea of functionalism is just to explain mental phenomena in terms of non-mental. In general, these attempts have not, in any case, met with great success.

Heil (1983) argues that perception is always 'epistemic'. The distinction of 'epistemic' and 'non-epistemic' is not, in his view, a difference between two kinds of perception, but rather between two ways of saying what is perceived, of describing the object of perception: If Lois Lane sees Clark Kent, it is also true that she sees Superman, but if she is aware that she is seeing Clark Kent and not that she is seeing Superman, then 'Lois Lane sees Superman' is a non-epistemic description of her perception, whereas 'LL sees Clark Kent' is an epistemic description. To say that perception is always 'epistemic' is to say that it always involves some belief about what one sees; an object is always perceived 'under a description'. There are always two (or more) possible ways of characterising a perception: 'epistemically', in terms of the perceiver's cognition - roughly, what she believes the thing to be; and by some description of the object of perception which is not the content of the perceiver's epistemic state. These are not descriptions of different kinds of perception, but different kinds of description of perception.

This seems to me the right way of thinking about perception, but I do not assume it. My claim is rather that in ordinary - 'folk psychological' - explanations of belief and behaviour, it is always an epistemic description of perception that is explanatory. 'Lois Lane believed the world would be saved because she saw Superman on the scene' would be false, would not explain her belief, if she believed she was

seeing Clark Kent, but not that she was seeing Superman, even if it is true that in seeing Clark Kent she sees Superman. Similarly, 'Leigh believed it would rain because she saw black clouds' explains her belief by reference to a 'mental event', her recognition of black clouds through the use of her eyes.

I conclude that there are two essential conditions of the appeal to perception in ordinary, 'folk psychological' explanations of beliefs and other mental states, and of behaviour. The perception is COGNITIVE, or epistemic; and it is WORLD-INVOLVING, or in Harman's sense, WIDE.

3.12 INPUT

Functionalism cannot use the ordinary, 'epistemic' notion of perception, as in 'She stopped because she saw the red light' - or '...saw that the light was red' - in the role of 'input'. Such explanations ascribe a propositional attitude or epistemic state - recognition, or belief - to the perceiver, and these intentional mental phenomena are to be explained, according to functionalism, by reference to 'inputs', 'outputs' and internal relations of mental states.

Functionalism must therefore analyse perception into a non-cognitive aspect, some notion of 'input' that does not assume any propositional attitude on the part of the perceiver, and a cognitive aspect which can be explained in terms of 'input', 'output' and internal functional role. The

idea will be to identify an 'input' that can be thought of as the cause of the relevant intentional state. This means - assuming a physicalist construal of functionalism - some cause of neural states or processes which constitute or realise the relevant mental states.

There are, broadly speaking, two sorts of thing that might be thought of as the input-causes of neural states and events: (a) physical events at the 'surface' of the body, especially the receptors of sense organs, such as the retina and the inner ear; and (b) the objects, events or properties of things that cause these sensory events - the objects from which light is reflected to the eye, for example, or their reflective properties. There are causal chains, as it were, leading to processes in the central nervous system, and the 'causes' of these processes might be identified with objects or events at various points on those chains, more or less distant from the place where the mental events are thought to occur - the brain. The distinction is sometimes marked by talk of 'proximal' and 'distal stimuli' (eg. Fodor 1983). Schiffer notes the choice in this passage:

> A functional role is simply any second-level property of first-level state-types possession of which entails that the state-type possessing it is causally or counterfactually related in a certain way to other state-types, to outputs, to inputs, or to distal objects and their properties. (Schiffer 1987: 21)

'Inputs' presumably means the events at the sense organs, and 'distal objects and their properties', the causes of these,

the things one perceives. This distinction seems to coincide with that made by Harman (1988) between 'narrow' and 'wide functionalism' (assuming that he would include states of the retina among 'internal states'):

> I claim that psychological explanations are typically wide functional explanations. That is, I claim that such explanations typically appeal to an actual or possible environmental situation of the creature whose activity is being explained. A narrow functional explanation appeals only to internal states of the creature and says nothing about how the creature functions in relation to an actual or possible environment. (Harman 1988:11)

(a) Narrow Inputs If 'inputs' are treated in the first, 'narrow' way, as 'proximal stimuli', or 'sensory stimulations', then the causal roles of neural states and events will consist of their - actual and counterfactual - causal relations with other neural states and events, and with states or events at the retina, inner ear and other sense-receptors. A functional theory or description of the physical system that realises mental states would then consist, in part, of 'input conditions' in the form of generalisations concerning causal regularities among eg. patterns of retinal stimulation and patterns of neural activity in the brain. The 'input' side of the theory would thus mirror the 'output' side, in that both would stop at the boundaries of the physical individual, the human body. (These boundaries may not be very well-defined, but the idea is, I suppose, roughly intelligible.) The functional theory would then be a theory of the individual, not of the individual and the world, which

has sometimes been considered a methodological desideratum of a psychological theory (eg. Fodor 1980).

This view of 'input', as events at the sense organs, seems to be generally accepted in cognitive science. The problem for the theory of vision is taken to be that of explaining how a person, or her brain, infers - or computes - representations of objects in the world from <u>two-dimensional retinal arrays</u>. This assumption is expressed by Hurlbert and Poggio:

> What does vision do? The plain answer is that vision transforms light signals into internal representations of the things that transmit them. Human vision starts with a two-dimensional pattern of light (an image) on each retina and ends with a description of three-dimensional objects in terms of their shape, color, texture, size, distance, and movement. (Hurlbert and Poggio 218[31])

And Pylyshyn, in his book <u>Computation and Cognition</u>, says that,

> recognizing or identifying a stimulus as an instance of some category does involve inference; that's what we mean when we speak of seeing a stimulus <u>as</u> something. And, of course, it is what we see things as that determines their effect on our behavior. (Pylyshyn 1984:15)

The 'inference' is from the pattern of stimulation at the eye or ear, as it may be, to an idea or representation of the object responsible for it. Fodor gives a similar picture:

> the computational relations that input systems mediate - roughly, the relations between transducer outputs and percepts - are quite remote. For

---

[31] In ed. Graubard: The Artificial Intelligence Debate, MIT 1988. Marr's work makes the same assumption (Marr 1982).

> example, on all current theories, it requires elaborate processing to get you from the representation of a proximal stimulus that the retina provides to a representation of the distal stimuli as an array of objects in space. (Fodor 1983:53)

The 'proximal stimulus', the retinal array or some pattern of nerve impulses that it generates, is taken as the 'input' to the brain, from which facts about objects in the world - the 'distal stimuli' - must be inferred by some computational process.

(b) Wide Inputs Suppose inputs are treated 'broadly'. Then the causal roles of neural states and events will consist in part of their actual and counterfactual causal relations with 'distal' objects, properties and events, namely those things that cause the 'proximal stimulations'. (I assume the relevant causal connections are via the eyes, ears and other sense organs. I will concentrate on vision; similar things could be said about the other senses.) The relevant functional story will concern causal regularities between such 'distal' phenomena and neural processes. The essential idea of functionalism, in this respect, is that perception can be explained in terms of a causal relation between the thing perceived and neural processes in the person who perceives it. In this way, the causal roles of neural states would be world-involving, to match the world-involving character of the ordinary concept of perception (see above).

Loar's proposal is of this sort. He claims that part of

the psychological theory employed by each of us in ascribing 'propositional attitudes' to each other consists of 'input generalisations relating perceptual circumstances to beliefs' (Loar 1981:9). These 'input conditions' refer to objects or events in the individual's environment:

> The recognitional abilities we can ascribe to virtually any adult member of our society are numerous, involving types of artifacts, features of geography, kinds of flora and fauna, classifications of sounds as language, music, barking, laughing and engine noise, not to speak of numbers, spatial and temporal relations, relative size, and shape. What generalizations, corresponding to these recognitional abilities, might serve as input conditions of an appropriate functional theory? ....Their general form is: If z is in circumstances C, z believes that p. (67)

The 'circumstances C' include, eg. for the belief that there is a horse present, the presence of the horse:

> p must vary with C, which is secured by letting the truth of p be part of C. (Observational beliefs are not guaranteed to be true; they are those states whose functional roles are in part determined by p's truth being, with other conditions, counterfactually sufficient for them.) (ibid. 68)

The problem is to identify the causal connection between an object or state of affairs, and a person - or more precisely, the person's brain processes - which realises the perception of the object or state of affairs, without saying that the person sees the object, or the state of affairs. The specification of this causal connection must be free of ineliminable references to mental states or events - beliefs, for example. 'Inputs' are aspects of the causal roles of neural states and events by virtue of which they realise

certain mental state- and event-types, according to functionalism. Mental state-identities of neural states (or events) are to be explained in terms of the causal roles of the neural states (or events). These causal roles must therefore be described, in principle, without appeal to the mental state-types any of them realise. In particular, 'inputs' must be described in non-mental, non-intentional terms. But they must be identified so as to coincide with acts of perception: the relevant causal connection between eg. a horse and a person's brain activity must be perception of the horse.

A horse may have an effect on a person's eyes, or other sense organs, and hence on her brain, without her seeing it in the epistemic sense relevant to belief-desire explanation. It might be partly obscured by a tree, and therefore unrecognisable; in poor light conditions it might be taken for a mule or a cow. So the explanation of perception in terms of a causal connection would have to specify the range of environmental conditions in some way. Loar suggests that 'we can spell out physical relations between z and objects and spatial regions whose features make p true, relations which are normally sufficient for z's being able to observe that p' (Loar 1981:68). He does not provide details, and it is hard to imagine how the relevant conditions could be spelled out.

The environmental circumstances in which a person is able

to see a horse, or those 'normally sufficient' for observing that a horse is present, are enormously varied, even for an individual. It would be necessary to delimit, in physical terms, the conditions in which light reflected from a horse to a person's eyes is sufficient for the person to see the horse: the distance between the observer and the horse, the angle of sight, the degree to which the line of sight must be unobstructed (but allowing, for example, that some tall grass might come in between them), the level and quality of the light, how clean the observer's glasses must be, and any other environmental factors that affect the perception of horses. Furthermore, the relevant conditions would be different for horses, spiders, telephones, buildings, mountains, planets and stars; and generally for every kind of thing a person can see. So the theory can't hope to get away with one specification of the conditions in which a causal connection between an object or state of affairs, and a person, is sufficient for seeing the thing. The qualifications necessary to the causal specification of 'input', if it is to explicate perception, will be as many and varied as there are kinds of thing we are capable of perceiving. The conditions would vary, also, from one person to another: an experienced naturalist can spot an animal in conditions in which a city-dweller sees nothing but trees, for example. If both are looking in the same direction, we may suppose that the light reflected from the horse to their eyes is the same for both, but one sees what

the other fails to see. These considerations are not an a priori argument that the project Loar suggests is impossible, but they do, I think, make it seem quite implausible.

That a person perceives an object or state of affairs - or acquires an 'observational belief' about it - depends not only on the right sort of external relation obtaining between the thing and the person, but also on the person being, so to speak, suitably receptive. If the external conditions just discussed - physical relations between observer and object, and environmental conditions - could be specified, this would define circumstances in which a reliable image of the thing is produced on the observer's retina, or similar conditions for the other senses. But this is not a sufficient condition for seeing something, in the relevant epistemic sense. At a minimum, the observer must be conscious, and not distracted. As Loar says, other qualifications are necessary, having to do with the state of the person z: That 'z is attending to those aspects of z's observational field that .... are normally sufficient for z's perceiving that p', and that 'countervailing beliefs do not forestall believing that p'. (Loar 68-69) These 'internal' and 'external' conditions, together, would then be sufficient to complete generalisations of the form 'If z is in circumstances C - related to the state of affairs p in such-and-such ways - and is in state S, then z perceives, or believes that p'.

Schiffer (1987) has given reasons to doubt that such

generalisations can be produced. One kind of doubt concerns the possibility of completing the antecedent, by specifying the conditions on the state of the person. Schiffer remarks on the difficulty of completing a putative 'law' such as:

> If there is a red block directly in front of x and...., then x will believe that there is a red block in front of x. (Schiffer 1987:31)

The problem Schiffer raises is that the antecedent needs completing with conditions concerning the state of the subject - eyes open, not colour blind, not drugged, etc. But common sense, folk psychology, suggests no completion of this. There are certainly no 'platitudes', in Lewis's sense, answering to this demand.

It also seems doubtful that the condition of being attentive to the relevant aspects of the situation could be spelled out without saying something like, 'x notices that there is a horse', or 'recognises the thing as a horse'. To what aspects of a situation does one have to be 'attentive' to recognise something as a horse? I can stare right at something and not recognise it. My coming to believe that it is a horse seems to require my <u>seeing it as a horse</u>. Any condition to the effect that I recognise the horse as a horse, as falling under the concept 'horse', would render the explanation circular. Yet without such a clause, it does not seem that any conditions would be <u>sufficient</u> for having the belief that there is a horse.

The presence of a large, red patch in front of me,

blocking all other possible objects of vision, does not explain even my believing that there is something red in front of me, unless I notice it, see it as red. No completion of Loar's generalisations would suffice, I think, that did not include question-begging conditions, such as that the subject has the concept red and sees the object as falling under it.

An object or state of affairs in z's environment may cause a neural process in z's brain, via reflected light stimulating her retinas, for example, although she does not see the object, in the epistemic sense, or acquire the appropriate 'observational belief'. There seems no way to specify conditions such that the causal connection would be sufficient for the perception or belief, without rendering the account circular. (I think this is because perception is not a passive relation of a person and an object or state of affairs. It's not merely the object having an effect on the observer, or her brain, but an act, in which the observer responds to the object of perception.)

What is perceived, in the epistemic sense, or what 'observational belief' is acquired, depends not only on what is there to be perceived, and on the physical relations of the observer and the object or state of affairs, but also on the observer's 'mental states', including her possession of concepts, other knowledge, beliefs and expectations she has, and so on. Where I see only an insect, an entomologist may

see a Blatta orientalis. I may see a horse if I am expecting or looking for one, and not otherwise. Loar includes some such factors in his sketch of 'input conditions' by mentioning the conditions of 'attentiveness' to relevant circumstances, and being 'open' to certain beliefs (not having 'countervailing beliefs') (Loar 1981:69). Mental states of the observer cannot be irreducible in a functionalist theory, on pain of circularity. Loar notes that states of being 'attentive' and 'open' must be in turn functionally specifiable. This means that they must be explicable in terms of internal causal relations, input and output - 'causal role'. But this means that 'inputs' must be specifiable, in principle, without reference to such other 'internal' mental states as are conditions of perception, in the epistemic sense. When the complete theory is spelled out, 'inputs' will be referred to under descriptions independent of internal states and their causal connections, other than the immediate effect of the external factor. An 'input' must be nothing but an environmental impact on the central nervous system. It follows directly that 'inputs', as such, cannot be identified with perceiving, in the epistemic sense, since this does depend on other mental states.

It is not clear that the reference to such objects as horses and buildings in the 'input conditions' of a functional theory is legitimate, in view of the assumption that mental states are physically realised. The functional roles of

'mental states' are assumed to be filled by physical - neural - states or events. So the relations posited by the theory, internal and external, are to be realised by causal relations with physical objects, properties or events. If believing that there is a building in front of one is explained as having in one that neural state normally caused by present buildings (and with certain counterfactual connections with other mental processes and behaviour), then, the implication is that there is some neural state that is normally caused by buildings in one's environment. But buildings seem to be quite the wrong kind of thing to be involved in causal regularities with neural state-types. If there are regular causal connections between environmental factors and types of neural event, so that an appropriate science might generalise about them, these generalisations would surely not concern buildings.

Some states of sense organs are no doubt caused by buildings, on those occasions when buildings are seen, for example. But there's not likely to be any regular correlation of states of the retina and buildings. Buildings are just too diverse; they don't form a physical kind in any relevant sense. Buildings come in a great variety of sizes, shapes, colours, and textures (not to mention distances, and even rates of movement). Houses (and their images) constitute thoroughly heterogeneous categories from the point of view of any physical or biological science. And the same must be said

of innumerable mundane concepts: cloud, umbrella, newspaper, person, bus, sandwich, horse, large, authoritative, mild - and most of the vocabulary in the dictionary. It's absurd to suppose that a scientific theory of the causal connections between the brain and the environment would employ the concept of a sandwich. The kind 'sandwich' does not occur and is not definable in the vocabulary of any science, basic or 'special'. The science of optics might refer to the properties of reflective surfaces, but sandwiches do not form an interesting class from this point of view.

Stones break windows because of their density, hardness and velocity, not because of their colour or beauty or value. Buildings and books cause retinal stimulation because of their shapes and colours and reflective properties, not because they are books and buildings. Bookhood and buildinghood, so to speak, would be irrelevant properties for a theory of input to the central nervous system. And there is no prospect of translating these properties into the relevant kind: there are neither necessary nor sufficient conditions, in physical terms, for being either a building or a book.

I conclude that a 'wide functional' story would not refer to the kinds of things we normally perceive, and about which we form beliefs and have intentions: buildings, people, books, buses and traffic lights. At most, it would refer to the physical properties of these that are responsible for the pattern of light that reaches the eye: shape, colour,

movement, etc.

3.13 INPUT AND PERCEPTION COMPARED

Perception is not 'input' in any sense available to functionalism. I have argued that this sense is either that of stimulations, events at the sense organs, or that of 'distal stimuli', extra-personal causes of stimulations. Perception, as it figures in folk psychological explanations, is intentional - or epistemic - and world-involving. The first point is that to see something, for example, is to recognise it, or see it as a thing of a certain kind[32]; the second point is that perceptions are individuated by reference to their extra-personal objects, what is seen. The causal notions of input available to functionalism cannot meet these conditions. The kinds of causal circumstance definable within the constraints of functionalism will not generally coincide with acts of perception (perceptual events), individuated epistemically and with reference to their objects.

The interpretation of 'input' as sensory stimulation, events at the eye, ear, etc., is inadequate as a substitute for perception because it makes no mention of objects or states of affairs perceived, an essential aspect of perception in the ordinary sense, in which it occurs in explanation of beliefs, other 'mental states', and behaviour.

---

[32]That is, perception of something as a thing of a certain kind, explains belief or behaviour; 'seeing as' is explanatory, not mere, non-epistemic seeing.

Whether 'input' is understood in the 'narrow' way, as events at the sense organs (eg. retinal stimulations), or the 'wide' way, as objects or events in the world that cause sensory stimulations, there is no reason to think that these are sufficient for perceiving something. Loar recognises this, and adds conditions, such as the observer being 'attentive' to relevant circumstances. But as I argued, there's no reason to think these conditions could be made sufficient, without including something that would render the explanation circular, ie. something equivalent to 'the observer sees (that) S'. If this is right, there might be an external or sensory cause of some neural process, in the specified conditions, without the relevant perception taking place, or belief occurring. The functional explication of perception or 'observational belief' would be inadequate. So types of causal 'input' will not coincide with types of perceptions. 'A sees that p' cannot be explained as a non-intentional, causal relation between A and p.

Perceptions are individuated by reference to their objects: seeing a bus, or a building. But a causal explanation of retinal stimulation, by reference to features of the environment, would not be in terms of buses and buildings, but rather in terms of shapes, colours, and angles of surfaces, direction and character of light sources, and so on. Common objects of perception, such as buses and buildings, are not definable in such terms. There's no common

physical character of buildings, for example. So the types of 'cause' involved in a causal account of vision would not coincide with the types of object of ordinary perception. As in the comparison of bodily movements and behaviour, there are two incongruent schemes of individuation at issue.

3.14 REFUTATION OF FUNCTIONALISM

The functionalist idea that neural states fill the functional roles of mental states, as defined by the theory of folk psychology, depends on the identification of perception and input, and behaviour and output. The causal roles of neural states can be identified with the functional roles of mental states, if the causal relations that constitute the former can be identified with the connections that constitute the latter. For example, if seeing a horse can be explicated in terms of a causal connection between horses and neural events.

I have argued that behaviour (or action) and perception, as these are referred to in ordinary folk psychological, belief-desire explanations, are world-involving, or 'wide' in Harman's sense. As Harman puts it:

> Ordinary psychological explanations are not confined to reports of inner states and processes. They often refer to what people perceive of the world and what changes they make to the world. (Harman 1988:15)

I have also argued that behaviour and perception, as referred to in ordinary, folk psychological explanations, are

intentional or cognitive. Perception, in these contexts, normally involves a cognitive aspect, belief or recognition. That is what explains the acquisition of beliefs, for example. And behaviour is normally intentional action: that is what is explained by 'primary reasons'.

But I have also argued that the conceptions of input and output available to functionalism, assuming the thesis of physicalism, lack these features. The failure of functionalism follows.

Behaviour is not analysable in terms of bodily movement. Perception is not analysable in terms of the causes of neural processes. There is no systematic match of the categories, behaviour and perception, with the categories 'input' and 'output' - physical causes and effects of neural processes. But then there are no 'causal roles' of neural states or events to match the 'functional roles' of mental states and events. A functional description of the central nervous system would be strictly incomparable with a functional description of a person's mental life, because the systems have no common points of reference. It's definitive of mental states, according to functionalism, that they mediate perception and behaviour; this idea is what gives functionalism its plausibility. Neural processes mediate 'inputs' and 'outputs'. These are different in kind from perception and behaviour. There's no correlation of kinds of (physical, causal) input and output with kinds of perception

and behaviour. So neural processes do not 'realise' mental states and events, in the way functionalism claims.

3.15 FUNCTIONALISM AND THE COMPUTATIONAL THEORY OF MIND

The computational theory of mind may be thought of as a version of functionalism. Block, for example, claims that functionalism 'is the orthodox account of intentionality for the computer model of the mind' (Block 1990:270).[33] Computational psychology claims that all sorts of mental processes occur, which have the character of computations, the kinds of inferential processes modelled in computers. Vision, the comprehension and use of language, rational decision-making, and many other intelligent capacities and activities, are explained as involving or consisting in complex computations. Since we are not conscious of these processes, unable to confirm their occurence by introspection, they are explained as programs of the brain. A theorist may attempt to write a program which would enable a computer to parse a sentence, for example, or infer the nature of 'distal stimuli' from 'proximal stimuli' - and it is suggested that these tasks are actually, if unconsciously, performed by human beings, in their brains. Computational psychology is thus based on the idea that intelligent capacities are realised by programs of the brain, in the same sense in which computer programs realise a machine's computational capacities. Computational

---

[33] Similar suggestions are made by Boden (1990:2).

psychology aims to characterise - in theories of competence - or specify - in theories of performance - the programs that realise particular intelligent capacities of human beings.

A program describes a (possible) functional organisation of a machine. To say that a machine has, or runs, a particular program, is just to say that it has a certain functional organisation, ie. that a certain functional description is true of it. A program is instantiated in a machine - or perhaps in some other complex physical object, such as a brain - in virtue of the causal relations of its internal states and processes, and its inputs and outputs. To explain the computational capacities of a machine, how it reaches certain results from certain problems, for example, by describing its computational processes, is to explain its outputs in terms of its inputs and its internal functional organisation. There is nothing else to the operations of a computer, but its functional organisation and its inputs and outputs. (It's no coincidence that functionalism fits computers in this way, of course, since the idea was inspired by thinking about how computing devices work, starting with Turing's discussion of 'Turing machines'.[34])

A computational theory of a human mental capacity, which claims that the capacity is realised by a program of a particular kind, is therefore a theory of the functional

---

[34] Turing 1936, 1950; also Putnam 1960, 1967a, 1967b, 1975a, (1975b 291-303).

organisation of the brain. The claim that such-and-such a computational process is involved in, say, vision or decision making, is a claim about the functional organisation of the brain-processes underlying and mediating vision, choice and behaviour. The computational theory of mind is, from this point of view, a version of functionalism. It aims to give the details of the functional processes that are alleged to mediate 'input' and 'output'. In so far as this is so, then, the arguments against functionalism just rehearsed are arguments against the computer model of the mind. The events mediated by the processes in a person's brain are physical 'inputs' and 'outputs', not perception and behaviour. Describing these internal processes as 'computational', or providing a description of them in the form of a program, does not change this basic error in functionalism as a theory of mind.

The inputs and outputs of a computer are physical events. The processes that take place inside it are sensitive only to certain physical properties of these events. What happens inside depends only on the relevant physical properties of the 'inputs', and the internal processes determine outputs only according to their relevant physical properties. In a typical computer, these are electronic 'signals', patterns of electric impulses from the input terminal (keyboard) to the central processor, and from the processor to the output terminal (video display or printer). A program describes

functional relations between (physical) input-types, internal states, and (physical) output-types. That is, a computer instantiates, or is running, a certain program, if the pattern of relations between inputs, internal states, and outputs, matches the functional description specified by the program.

A machine, running a particular program, computes something - mathematical problems, or weather patterns, or share values - in so far as its inputs and outputs are 'interpreted'. Interpretation means assigning some value, or type of object or state of affairs to be represented, to each input- and output-type. This involves a systematic correlation of physical input-types (eg. electronic signals) to which the internal processes of the machine are sensitive, and values or represented objects, or whatever kinds of thing the computations are supposed to be about; similarly for outputs. Only things that are systematically correlated with input- and output-types can be described as computed by the machine. For if there is no such correlation, there will be no systematic, regular, predictable or rational relation with the internal, computational processes.

This point is illustrated by a simple calculator. What goes on inside is calculation of the results on the display, from the values entered on the keyboard, because there is a systematic correlation of numerals (or numbers), and signals to and from the circuit. This correlation is usually established for us by the manufacturers, who print numerals on

the buttons, and arrange for the display to light up in the appropriate way.

So if the brain is thought of as a computer, its 'inputs' and 'outputs' will be the physical events that are causes and effects of neural processes in it. That is to say, the same notions of input and output appealed to in functionalism. But, as I have argued, there is no systematic correlation of 'inputs' and perception, or 'outputs' and behaviour, in the human case. If the brain's a computer, then, it doesn't compute behaviour from anything corresponding to perception. At most it computes bodily movements from proximal stimuli (the physical effects of the environment on sense receptors).

3.16 SCIENTIFIC PSYCHOLOGY

Computational psychology accepts the limitations of 'input' and 'output', and hypothesises inferential processes to connect these with more familiar kinds of thought. This suggests that instead of 'folk psychology', functionalism should appeal to cognitive science, the overall goal of which is to define the program or programs that realise intelligent capacities in the human brain. For example, psycholinguistics aims to explain how we infer the syntax, and perhaps also the semantics, of sentences, from their 'acoustic forms'. The computational theory of vision claims that inferential processes lead from two dimensional 'retinal arrays' to

beliefs about (or 'representations' of) three dimensional objects of familiar kinds. In the explanation of 'output', we can expect theories of how the brain computes the movements necessary, in given circumstances, to perform a desired action.

Scientific, computational psychology will either attribute mental states like belief, desire, intention, and their kin, or it won't. If it uses the ordinary mental vocabulary, these terms should represent the same concepts as we use in ordinary ascriptions and explanations - in 'folk psychology'. (If not, it's just equivocation.) But what could be the justification for the use of the ordinary concepts of belief, intention, and the rest, in the articulation of computational psychology? Why should we think that functional states of the brain are beliefs, and so on? The plausible functionalist answer was that they mediate perception and behaviour, in the way that ordinary explanations suggest. This answer is not available to the computational theory, because it is concerned with states and events that mediate <u>inputs</u> and <u>outputs</u>, which, as I have explained, are not the same as perception and behaviour. Abandoning the intuitively plausible basis of functionalism, computational theory has no justification for the use of ordinary mental concepts like belief.

If, on the other hand, computational theory does <u>not</u> use the ordinary mental vocabulary, then it has no claim to be

psychology, the study of mental processes, rather than physiology. The brain has, of course, some functional organisation; hence some 'program' describes it, and explains the relations of inputs and outputs. Perhaps this is reason to think of the brain, or the central nervous system, as a computer, or a computational system. But the theory of this system won't explain behaviour, only (at most) bodily movements; and it won't define any functional states corresponding to belief or desire - because these are mental states related to behaviour and perception in certain ways.

It is recognition of this dilemma, I think, that leads eliminativists to suggest that we should reject ordinary explanation of behaviour, with its mental vocabulary, and replace it with a scientific theory of inputs, neural processes and outputs.[35] They assume that mental states must be physical or functional states if they are anything at all, and since they are not physical or functional states, conclude that they are nothing. The eliminativist proposal - which is not likely to find general acceptance - would entail giving up any explanation of behaviour. If we were only interested in the bodily movements of humans, and not also in their actions, this might make sense; but it's hard to imagine what such a change in our interests would be like. Computational descriptions of brain processes, and their role in the causation of bodily movements, would provide no understanding

---

[35] eg. Churchland 1981; Stich 1983.

of human beings and the things they do: no human understanding of persons. This kind of understanding is expressed in novels, but a 'computational' novel is unimaginable. 'Eliminativism' is a reductio ad absurdum of whatever assumptions entail it. Lynne Rudder Baker comments on the idea of abandoning common-sense mentalism:

> If within our power at all, it would be to give up the point of view from which thinking about anything, meaning anything, or doing anything intentionally is possible. To abandon the common-sense conception of the mental would be to relinquish the point of view from which the idea of making sense makes sense. (Baker 1987:174)

A sufficient response to the eliminativist is the question, 'Do you believe what you say?'

Theorists of computational psychology and 'artificial intelligence' typically assume that difficulties concerning 'inputs' and 'outputs' can be resolved by developing theories of the kinds of program that the neural computing apparatus employs in vision and the other modes of perception, and in the control of 'motor responses', including the production of speech. In the next chapter I will argue that the basic assumption of such theories, that neural processes can be described as 'computational', rests on a conceptual mistake.

CHAPTER 4

MEANWHILE, INSIDE THE HEAD

4.1 COMPUTERS, SYMBOLS AND BRAIN

> The basic idea is that the mind is the program of the brain and that the mechanisms of mind involve the same sorts of computations over representations that occur in computers. (Block: 'The computer model of the mind', in eds. Osherson & Smith 1990:247)
>
> ..the mind is taken to stand to the brain as the software and data of the computer stand to the hardware. (Jackendoff 1987:16)
>
> I want to maintain that computation is a literal model of mental activity, not a <u>simulation</u> of behavior (Pylyshyn 1984:43)

What is a computer, that the human brain may be compared with such a machine, and mental processes with the implementation of computer programs? In the last chapter, I contrasted the notions of 'input' and 'output', as these apply to computers, with the concepts of perception and behaviour we use in explaining human thought and behaviour. I concluded that, if a computer is understood as a physical system that processes input and generates output, then the human brain - or body - may be thought of as a computer, but explanation on this basis will be incomparable with <u>psychological</u> explanation, which has to do with perception and behaviour, not physical input and output. Some psychologists and philosophers who advocate a computational

theory of mind, however, place greater emphasis on the nature of the processes that, they claim, take place within computers and human brains. In this view, it is the formal, symbol-processing character of computers that is the basis of the model for mental activity:

> Computational psychology treats mental processes as operations defined in terms of formal manipulations of formally described representations. (Boden 1988:232)

The standard explanation of this kind - what might be called the classical analysis in work on the foundations of computational psychology - is that a computer is, in Newell and Simon's expression, a physical symbol system : a system of symbols and expressions, on which operations are performed in accordance with certain rules. Computational processes are thus rule-governed symbol manipulations. Newell and Simon explain:

> A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus a symbol structure is composed of a number of instances (or tokens) of symbols related in some physical way (such as one being next to another). At any instant of time the system will contain a collection of these symbol structures. Besides these structures, the system also contains a collection of processes that operate on expressions to form other expressions: processes of creation, modification, reproduction and destruction. A physical-symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves. (Newell and Simon:117)

This abstract definition is intended to capture what is common to computers, human brains, and any other physical 'systems' capable of intelligent activity. Thus Newell and Simon propose:

> The physical symbol system hypothesis: a physical symbol system has the necessary and sufficient means for general intelligent action. (In Boden 1990:111)

Similar explanations can be found in Anderson (1983), Boden (1988, 1990), Haugeland (1985), Johnson-Laird (1983, 1988), Newell (1990), Pylyshyn (1984), and in Turing (1936, 1950). For example, Johnson-Laird explains:

> ..a computer ... has two symbolic abilities. It can manipulate symbols so as to transform them or to construct new symbols out of them. And its internal operations are controlled by symbols. All the computer ever does is to manipulate binary symbols... (1988:34)

And Boden explains the view (without explicitly committing herself to it):

> The formalist sense of computational psychology .... covers those theories which hold that mental processes are, and/or are to be explained in terms of, the sorts of formal computations that are studied in traditional computer science and symbolic logic. These disciplines define "computation" as the formal manipulation of abstract symbols, by the application of formal rules. That is, the criteria for effecting one symbol-manipulation rather than another, and also for distinguishing the various transformations that are possible, are purely syntactic. (Boden 1988:229)

The hypothesis that the brain is a computer - hardware - and the mind a program or programs - software - is then

expressed in the claim that the brain is, in the sense to be explicated, a 'physical symbol system'; and mental processes consist in 'symbol manipulation'. Thus Johnson-Laird:

> A major tenet of cognitive science ... is that the mind is a symbolic system. It can construct symbols and it can manipulate them in various cognitive processes. (1988:34)

The processes that take place inside a computer, and in the brain of a rational creature, in this view, are analogous to the use of a formal symbolic "language", of the sort developed in mathematical logic, in which expressions composed from a finite stock of signs are produced, transformed, and put in various relations with each other in accord with rules of formation and transition. Computer programs are written in such languages, and programs describe what happens in computers.

4.2 HAUGELAND ON FORMAL SYSTEMS

I will follow Haugeland's explanation of this thesis. In his book Artificial Intelligence: The Very Idea, Haugeland defines a computer as an 'interpreted automatic formal system.' He explains that a 'formal system is a game in which tokens are manipulated according to rules.' (Haugeland 1985:48)

(Haugeland uses "token" rather than "symbol" in this context, reserving "symbol" for interpreted tokens, ie.

tokens that have a semantic value - meaning or reference. "O"s and "X"s in tic-tac-toe[36] are mere "tokens", in this sense, whereas "red" and "dog" in English mean something, so in Haugeland's sense are symbols. "7" is a symbol according to mathematical realists, but only a token according to formalists. This use of "token" accords with the use of "symbol" in the quotations given above.)

Haugeland gives as examples of 'formal systems' games such as tic-tac-toe and chess, and also formalisms such as the 'languages' of symbolic logic. These 'games' illustrate an essential feature of computational systems. A 'formal system' is a system of rules, or an implementation of such a system - such as a game of chess. The rules define the elements of the system - the pieces or tokens, and permissible combinations of these in complex expressions - and legal 'moves' - what may or must be done from a given position. A formal system is thus defined by rules of formation and rules of transition. So the rules of chess state that the game is played with such-and-such pieces, starting in a certain arrangement; and define permissible moves in any state of the game. They determine the difference between legal and illegal moves; moves and non-moves. A move that does not conform with the rules is 'illegal', a mistake or an error, incorrect. (Of course, you have to be playing the game, or intending to, in order

---

[36] 'Noughts and crosses' in British English.

to make a mistake or an illegal move. A child who has not learned the game, arranging the pieces and moving them around on the board, is not making mistakes, nor making legal moves when he happens to move a piece in a way permitted by the rules.)

Games and logical symbolisms, of these kinds, are systems of manipulation of symbols or tokens, in accord with certain formal or 'syntactic' rules. Semantic properties of the symbols, such as truth and reference, sameness and difference of meaning, etc. - if they have any - are irrelevant to the system. The rules don't refer to the meanings - or referents - of the symbols; only to tokens, or types of token, like the X's and O's of tic-tac-toe; or the signs used in symbolic logic; or the beads of an abacus. So even if the symbols have meaning (or reference), it's irrelevant to their role in the rule-governed system. The rules don't apply to them in virtue of their semantic values[37]. The system treats them as <u>tokens</u> (in Haugeland's sense), even if they have meaning, or refer, and are therefore <u>symbols</u> (in Haugeland's sense). (This corresponds to what Fodor calls the 'formality condition' on explanations of mental processes: he claims that mental processes consist of computations over 'mental representations' which are 'sensitive solely to syntactic

---

[37] This is not to imply that meaning and reference are the same. <u>Neither</u> meaning <u>nor</u> reference are relevant to the role of symbols in a formal system.

properties' (1987:19) of the representations.)

On the other hand, the physical properties of the tokens are irrelevant to the rules. Chess can be played with plastic or wooden pieces, or humans or helicopters, or ink marks on paper. All that is essential is that there is a certain number of tokens of identifiable types, in a certain ratio (32 pieces of twelve types, for chess); and a certain system of relations among them (eg. positions on the standard board, or algebraic coordinates). The first point can be expressed by saying that the symbols are 'meaningless' or 'uninterpreted'; the second by saying that they are 'abstract'. Of course, any particular set of chess pieces will be concrete, but the rules refer to them abstractly. The rules define symbols and moves abstractly. A physical symbol system is therefore a collection of objects manipulated in accord with these abstract rules - like a particular chess set. A physical symbol is an object used in such a system. The rules are 'syntactic': they refer only to tokens and types, and relations of these. A formal system is a 'game' defined by syntactic rules: rules that govern moves according to the 'abstract and meaningless' definition of symbols.

This picture is expressed in the description of a computer - and therefore also the human brain, according to the computational theory of mind - as a 'syntactic engine' (Dennett 1987) or a 'syntax-driven machine' (Fodor 1987:20).

Computational - and therefore, in this view, mental - processes are 'syntactic operations over symbolic expressions' (Pylyshyn 1980:113).

4.3 WHAT "SYMBOL" MEANS

When we call something a 'symbol', this normally implies that it stands for or represents something else. We speak of a symbol <u>of</u> something - as an expensive car is a symbol of status or wealth, a cross is a symbol of christianity, and a well-known design is the symbol for radiation. In describing a computer as a symbol-processing system, however, an abstraction is made from such semantic relations. Haugeland indicates this distinction by using 'token' instead of 'symbol' to refer to computational states in abstraction from their referential relations; he reserves 'symbol' to refer to tokens that represent something. So when a computer - a 'computational system', perhaps a human brain - is described as a symbol-processing or token-manipulating system, this is not meant to imply anything about the semantic status of states or components of the system. A 'symbol', in this usage, is not (necessarily) a symbol of anything. This seems to be the case of a general-purpose computer fresh from the factory. It has a 'machine language', operates according to certain complex transition rules, but has not yet been assigned any 'interpretation'. If it processes symbols, it does not yet

represent anything. 'Semantic interpretation' has to be assigned by its programmers or users. Similarly, the pieces of tic-tac-toe or Chinese checkers have no 'semantics'.

So what could be meant by 'symbol' or 'token' - symbol- or token-processing - in the absence of semantic properties or relations? This way of describing a 'system' implies that formal rules govern the moves or processes that occur. That is, 'symbol' (or 'token') in this usage means *that which is manipulated in a rule-governed process*. So when we call the letters and other signs used in symbolic logic 'symbols', this refers to marks on paper - or some other medium - with the status of things used in a rule-governed 'game'. Mere marks wouldn't be 'symbols'. What makes them 'symbols' or 'tokens' is that they are used, or designated to be used, in a procedure governed by rules, an implementation of a formal system.

Talk of 'tokens' or 'symbols' makes sense only in so far as there is some sort of rule-governed activity. Any physical process consists of objects or events which could be classified into kinds - or types - in various ways, affecting each other through causal connections directly or indirectly. To describe the elements of a system as 'tokens' or 'symbols' is to imply either that they stand for, or represent, other things; or that they have - or are intended to have - roles specified by rules. Otherwise 'token' - or 'symbol' - is a category that includes

everything. Similarly, to describe carved wooden objects as 'pieces' - in the sense of 'chess pieces' - implies that they are used, or are intended to be used, in a game, ie. moved according to rules. Any objects could be so used, but they are not 'pieces' until they are used, or are designated or intended to be used, to play the game. 'Token', used in discussion of formal systems, is equivalent to 'piece' in reference to games.

(There is a usage of 'token' in contemporary philosophy in which it is correlative to 'type', in the sense that to call something a 'token' is just to say that it is an instance or example of a certain kind of thing.[38] Everything is a thing of some kind, even if it's the only thing of that kind, so that in this sense everything can be called a 'token', if this is given sense by adding that one means a token of some kind specified. In this sense, 'token' is synonymous with 'example' or 'instance'. It cannot be this sense of 'token' that is intended when a computer is defined as a 'token-processing system', because it would be vacuous to say that, in this sense, computers operate with tokens.)

The method of calculation with numerals - arabic or binary, for example - is a formal system. The method could

---

[38]This usage is not unchallenged, eg. by Richard Wollheim in *Art and its Objects*

be defined by syntactic rules: rules referring only to types of symbol and permissible or mandatory 'moves'. We could teach someone to use a system with three types of symbol - triangles, circles and squares, say - specifying only what must be written where, given starting positions of specified kinds. These rules would constitute algorithms for the use of the system: conditional instructions (ie. instructions of the form 'If such and such is the case, then do such and such') with an order of execution. This game could consist of the methods of performing the basic arithmetic functions, in the tertiary numerical system. The person who learns this game need not realise that it is equivalent to the decimal (or binary) system, nor that it could be used to represent eg. stocks and shares, or to calculate 'mathematical truths'. Similarly, first-order propositional logic can be defined - and taught - as an 'uninterpreted' formal system, a method of manipulating symbols, with (merely syntactic) rules of formation and transition.

4.4 SYMBOLS AND REPRESENTATION

Formal systems can be used to represent values, or states of affairs and processes in the world. For example, an abacus is used as a 'formal system'. The symbols or tokens of an abacus are its beads. Correct use of an abacus involves the manipulation of these tokens according to

certain rules or algorithms. The rules could be formulated without reference to anything the beads represent. In principle, someone could use the abacus, following the rules, without understanding that its use is interpretable as calculation, ie. as anything to do with numbers. But it can, of course, be used to represent mathematical relations between numbers, because the formal or syntactic operations carried out mirror those relations. If you start with settings that can be interpreted as representing the numbers twelve and eleven, and follow the multiplication-algorithm - the rule for manipulating the beads that corresponds to multiplication - the end result will be the representation of 132, under the same system of interpretation. An abacus may be used to calculate values or quantities of goods in a warehouse. But such interpretations are irrelevant to the rules for its use. A different algorithm for an abacus, a different set of rules for manipulating its beads, might not be usable to calculate arithmetic functions. It would nevertheless be a formal system, and might have some other computational application, representing perhaps some physical process. (Formalists claim that there is nothing to (pure) mathematics except the manipulation of symbols according to syntactic rules; mathematics is just a formal system, in this view.)

A formal system can be used to represent, and compute facts about, things in the world - weather, stocks and

shares, students and grades, etc. This can be done by devising series' of moves in the symbolism such that, given a certain pairing of the initial sequence of symbols with a state of affairs in the world, and a certain process of change in that state of affairs, the result of the transformation of the symbols in the system will represent the result of the real-world changes, under the same method of pairing symbols and states of affairs. Transformations of symbols can then be used to predict what <u>will</u> happen or what <u>would</u> happen under certain - predictable - circumstances. This is what makes computers useful. Writing programs consists in devising sequences of 'moves' that will meet this pairing condition for given representational tasks.

4.5 FORMAL SYSTEMS AND THE POSSIBILITY OF ERROR

A person playing Chinese checkers, for example, or solving problems in symbolic logic, or calculating in binary notation, implements a 'formal system', in Haugeland's sense. These are cases of manipulating 'tokens' or 'symbols' in accordance with a system of syntactic rules: the activity is a case of 'rule-governed symbol-manipulation'. In such a case, there is a difference between correct moves and errors or mistakes. For example, if a person playing Chinese checkers ('solitaire') jumps over <u>two</u> pieces, or removes a piece without jumping over it,

he's made a mistake - or cheated. There's a difference between 'correct' moves and 'incorrect'.

Similarly, in solving problems in a system of symbolic logic - a formal language - some moves are permitted, some are not. Even if the person does everything correctly, some things he could do would be incorrect or 'illegal'. Describing the moves he makes as 'correct' implies this: some other possible moves would be incorrect. Not just anything that could be done with the pieces or symbols is permissible in the system. This is (part of) what it means to say that the activity is a 'formal system', such as a game of the relevant kind. If a person is merely moving things around, or making marks on paper, with more or less regularity, but nothing would count as incorrect or a mistake or an error, by some standard, then he's not playing a rule-governed game: he's not applying or following rules, implementing a formal system. This does not imply anything about what rule-following consists in - whether social or individual facts, conscious or unconscious states, relations to physical or abstract objects. Whatever constitutes rule-following (application, implementation), this condition must be met: some possible 'moves' would be wrong.

Sometimes the distinction of correct and incorrect moves - in particular, the identification of some as wrong - is made by reference to normal practice, or to explicitly formulated rules. That is, normal practice or official

rules are the standard of correctness. If the player jumps over two pieces, this is wrong because that's not how the game is played; this game is defined by the rule that only one piece may be jumped at a time, so to do otherwise is to make a mistake, violate the rule. The rules may be explicitly formulated, in a definitive version, as is the case with chess; or they may be common knowledge, in the case of less formally institutionalised games; or they may be merely implicit in the way the game is normally played, corrections made and accepted, and so on. In formal logic, a system is defined by explicit rules of formation and transition, written out in a formal presentation of the system. In these kinds of case, a standard independent of the individual player determines the difference between correct and incorrect moves.

A logician may invent a formal system, by formulating rules of formation and transition, and then operate the system - 'manipulate' (ie. read and write) symbols - accordingly. An individual may invent a game, or a variation of a game, and play it. For example, I may play a variant of Chinese checkers in which it is permitted to jump two pieces in certain circumstances - perhaps I allow myself to do this once or twice in a game if there's no other way to complete it. I would be playing according to a different rule (different, that is, from the normal rules). Or I may invent and play a wholly new game or 'formal system'. In

such a case, even if I have not explicitly formulated the variant rule, there must still be a distinction between correct and incorrect, moves permitted and violations. I must have the intention, at least, to conform to some rule or rules; I must be prepared to recognise some moves as mistakes.

For suppose there is no such standard of correctness. Suppose that a 'player' moves pieces on a board, or writes 'symbols' in various arrangements, but that there is neither an independent practice, nor formulated rules, nor any intention - conscious or unconscious - on the part of the player to conform to any particular rules, or to recognise a distinction of correct and incorrect moves. There is only the behaviour - the moves he makes. These may exhibit some regularity. But nothing he could do would be an error, for anything could be brought into conformity with his previous moves by some rule. If some possible move is to be an error - if there is to be a difference between correct and incorrect moves - there must be some standard other than the moves he makes, if only what he means or intends to do. If the rule he follows is identified with what he does, then anything he does must be in accord with the rule. But 'following (or implementing or obeying) a rule', as I have argued, implies a distinction of moves as correct and incorrect. If someone follows (or implements or obeys) a rule, then some different, conceivable action would be a

mistake, incorrect, an error.

So in a case in which there is no standard of correctness, other than the player's actual behaviour, he would not be following a rule. There must be something else - other than what he does - that determines the rule (what rule) he's said to follow or implement or obey. If there is only behaviour, there is only change in behaviour, not a mistake. But the rules that define formal systems define the difference between correct and incorrect moves. If a formal system is being 'played' (or implemented), there is a difference between correct and incorrect moves. So if there is no distinction of correct and incorrect moves, there is no 'formal system'; the activity is not 'rule-governed', only more or less regular.

4.6 LANGUAGE AS A 'FORMAL SYSTEM'

Chomsky's theory of competence manifests this assumption. Chomsky's treatment of grammar in effect proposes that use of language can be considered as a 'formal system', the rules of which constitute a generative grammar. These are syntactic rules that determine permissible combinations of words, considered only as syntactic types. (Assuming the independence of syntax and semantics.) They define the set of sentences, the grammatical strings of a language.[39] For an individual speaker, a definite

[39] See eg. Chomsky 1957.

(infinite) set of strings is grammatical; all others are ungrammatical (ignoring qualifications about 'degrees of grammaticality').

This distinction corresponds to the distinction of 'correct' and 'incorrect' - Chomsky's repudiation of 'prescriptive' linguistics notwithstanding. To say that one string is grammatical and another ungrammatical, for a speaker-hearer, is not just to say that she 'produces' one and not the other. In fact, it's not to say this at all, because (a) (infinitely) many strings grammatical-for-the-speaker are never produced, (b) the speaker may produce ungrammatical (for her) strings, (c) her judgments of grammaticality, although they constitute the linguist's primary evidence, are not wholly reliable: she may, in Chomsky's view, judge 'grammatical' a string which in fact is not grammatical in the sense of being determined by the rules of the grammar she 'knows' or 'cognizes'.

Rather, to say that a string is grammatical (or ungrammatical) for a speaker, is to say that it is in accord (or not) with the grammar (syntactic rules) she 'knows' or has 'internalized' or 'cognizes'. That is, the distinction of grammatical and ungrammatical strings, like the distinction of correct and incorrect moves, is in principle relative to a fact other than her behaviour; in this case, according to Chomsky, an individual, 'psychological' and 'biological' fact. Were there no such fact - individual or

social; mental, physical or abstract - there would be no difference between 'grammatical' and 'ungrammatical' strings; there would be only strings produced. Hence, 'competence' must be different from 'performance'.

4.7 AUTOMATIC FORMAL SYSTEMS

A formal system can be automated: a machine constructed to implement the system, follow the rules, execute the algorithms, that define the system. An automated formal system is just a formal system embodied in a mechanism in which causal connections replace human agency in the manipulation of the symbols or tokens. This is straightforward for a determinate formal system, such as an abacus. In such a system all rules are mandatory - from a particular starting position, only one move or series of moves is legal. (Non-determinate systems may raise problems, but I will ignore them. I assume the computational theory has adequate answers. Chess programs implement a formal system of non-mandatory rules, a game in which there is choice among legal moves. Of course, they do so by implementing determinate 'decision-making' algorithms. Random elements can also be built into computers. Whether human choice and decision can be explained by analogy with such methods is a question I will not attempt to answer. Ignoring this problem is a concession to the computational theory of mind, so it does not vitiate my criticism of that

theory.)

A computer, in Haugeland's terms, is an '<u>automatic</u> formal system'. This means that it is a formal system in which legal moves happen automatically, because of the construction of the machine. It is possible to construct a device in such a way that token-manipulations in accord with the rules are effected by causal connections among the parts of the device. The human actions that move chess pieces and abacus beads in accord with the rules are replaced by mechanical or electronic connections. Furthermore, the components - or states of components - of the device that constitute the symbols are the causally efficacious elements of the device, in a computer. The electronic states of components (eg. voltages) that are the machine's 'tokens' themselves <u>cause</u> the next 'moves', as required by the rules of the machine language (or at a higher level, a program). A series of 'moves' is a series of physical events each of which causes the next. This is what Johnson-Laird means when he says that a computer's 'internal operations are controlled by symbols' (1988:34). It is as though the beads - or their positions - caused the appropriate movements of other beads on an abacus. So the actions of humans in manipulating the symbols of a game like tic-tac-toe, or a logical symbolism, are replaced by causal connections of the 'symbols' themselves. A machine is so constructed, or programmed, that if the rules require, for example, that if

the symbolic state X occurs then Y should follow, the physical instantiation of X (the state of the machine that implements X) will cause the instantiation of Y to occur. In this way 'homunculi' can be 'discharged', as Dennett says (Dennett 1978): there is, in the final analysis, no need for agents to implement the rules.

An automated formal system, such as an old-fashioned mechanical calculator, or an electronic version, or a computer, is designed and constructed so that the token-manipulations that occur are the same as those a human would perform in following the rules correctly. As Alan Turing explained:

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human computer. The human computer is supposed to be following fixed rules; he has no authority to deviate from them in any detail. (Turing 1950:440)

A computer is a physical system, a complex of parts with causal connections, in which changes occur in ways explicable by laws of physics - electrical laws, for example, or at another level, laws of particle physics. These changes occur with a high degree of regularity, hence predictability (with knowledge of the relevant physical, or electronic, laws). What is it about such a machine that allows it to be described as a 'formal system', in which rule-governed symbol-manipulation occurs? The arrangement of its components, and their causal connections, is such

that a mapping is possible of a certain symbolic system onto the physical elements of the machine (or relevant properties of them, such as voltages), in such a way that the physical changes in the machine will respect the rules. This is what Pylyshyn calls an 'instantiation function', mapping the symbol-processing description of a machine onto a physical description:

> By mapping from physical to computational states, such a function provides a way of interpreting a sequence of nomologically governed physical state changes as a computation, and therefore of viewing a physical object as a computer. (Pylyshyn 1984:56)

A pocket calculator provides a simple example. The physical components of the device are arranged in such a way that processes that occur when buttons are pushed can be paired with "0"s and "1"s used in accordance with the rules of the binary system, for example in calculating : 110+101 = 1011. If certain states of elements of the device - voltages in micro-circuits - are interpreted as "0"s and "1"s, then the physical processes are such that the (syntactic) rules of manipulation of digits in binary calculation are 'obeyed' by the device. The 'instantiation function' is thus given by correlating "0"s and "1"s in the rules with certain types of physical events (properties of parts) in the device. Under this interpretation, the pocket calculator is an 'automatic formal system', a rule-governed symbol-processor. Under this interpretation, it does exactly what a 'human computer'

would do in performing calculations in the binary system, writing binary symbols in accord with the 'fixed rules' of this system - assuming, of course, that he does not 'deviate' from the rules, ie. that he makes no mistakes. (Machines, in fact, are more reliable - less likely to 'deviate' - than humans at this sort of task.)

A device can be constructed to implement a rule or system of rules. For example, the rule 'If either A or B, then
C' can be implemented by a simple electronic device called an OR-GATE. This has two inputs and one output, and the output is activated - generates a small potential difference - if either (or both) of the inputs is activated. Obviously, the device is construed as following (implementing, carrying out) this rule by identifying "A" with the activation of one of the inputs, "B" with the activation of the other, and "C" with the activation of the output. Under this 'instantiation function' (Pylyshyn's term), the device 'computes' the rule. A combination of OR-GATES, AND-GATES and NOT-GATES, suitably connected, can implement a <u>system</u> of rules, complex functions from inputs to outputs. Such devices are called 'logic circuits'. They can be built up into computers. In general, a machine can be construed as implementing a system of syntactic rules if it has the right functional organisation: such that the states and causal relations of the machine can be paired

with the symbols and rules of the formal system. Then the machine can be described as 'rule-governed' in the sense in which a person moving chess pieces, or writing "P"s, "v"s, etc., is engaged in a rule-governed activity. In the sense in which the person can be described as 'processing symbols', so can the machine. (A person and a machine could manipulate symbols according to the same system of rules.)

4.8 COMPUTERS AND NATURAL PHYSICAL SYSTEMS

The case of the computer has to be distinguished from that of physical systems such as the solar system, or the tidal movements of oceans. For these are also systems in which physical changes occur with a degree of regularity, and for which instantiation functions could therefore be defined for some set of 'syntactic rules'. That is, there is some projection of some formal system, rules for the manipulation of symbols, under which the movements of the planets or oceans would 'respect' the rules. But we would not say that the solar system or the oceans are rule-governed, or that they process symbols; even if we discovered some enormously complex pattern of movement of - say - all the stars in the universe, such that an IBM-type program could be projected onto the pattern - unless we assumed an intelligence behind the pattern, or using it.

In these cases, a change in the behaviour of the system - a change in the regularity - would not be incorrect, a

mistake or an error. There is no standard for such a distinction. What happens, happens. Whether there is a change, by some measure, or not, the laws of physics must explain what occurs (or else we have the laws wrong - that is, the laws are other than we thought). If activity is 'rule-governed', there is a difference between correct and incorrect moves. There is no such difference in these cases, so these are not rule-governed systems. This accords with our ordinary thinking: we are not inclined to call the solar-system or the oceans 'rule-governed', or to say that they involve 'symbols' or 'tokens'. These concepts are alien to the domains of planets and oceans.

4.9 COMPUTERS AND ERROR

A computing device - a calculator, or a standard digital computer - does fulfill the condition of 'normativity' - there being a difference between correct and incorrect 'moves'. The condition is met by designers' or users' intentions. If the device deviates from the rules that govern its computational processes, this is really a deviation from the rules, because those are the rules it's supposed to be implementing. A NOT-GATE, for example, or a complex of NOT-GATES, AND-GATES and OR-GATES in a logic circuit (the processing units of computers are just complex logic circuits), is designed to work a certain way, to have a certain input-output function, to operate according to

this functional description, not that.

In a computer, some physically possible events are computational errors, because they are deviations from the machine language, or some higher level program, which the machine is supposed to be implementing. The question, 'What formal system - rules - is it implementing?' is not to be answered merely in terms of the functional organisation it actually has. If that were the case, nothing could be an error, because every event would be part of its actual functional organisation. This is why the solar system is not 'rule-governed', does not implement an algorithm. In the case of the computer, there is another standard: the rules it's intended - by its designer or builder or programmer - to implement. This 'intentional stance' (to borrow Dennett's expression) makes the difference between correct and incorrect 'moves'. Hence a computer is a 'rule-governed' system; hence it makes sense to describe it as 'manipulating symbols'.

If an artifact is described as following or implementing rules - as running a program, for example - this implies the distinction of correct and incorrect moves or operations, just as it does when a person is described as playing a game or employing a formal system. As in the human case, this implies some standard of correctness, distinct from the activity itself. In the case of the computing device, the distinction of correct and incorrect

'moves' is made by comparison with the way the device is supposed to work: the rules it's intended to implement. Suppose a device is designed to implement a certain formal system. Then if it does something that is not in accord with these rules - under the intended 'instantiation function' - this is an error, incorrect by comparison with the way its designer or programmer meant it to work.

Consider a physical system with a certain functional organisation, ie. in which states of its parts (eg. kinetic, chemical, electrical or thermal states) are causally interdependent in a more or less complex, more or less regular way. Suppose that this organisation is such that it instantiates some formal system, under some 'instantiation function', ie. there is some way of pairing the symbols and rules with the states and causal relations of the system. That is to say, some formal system - rules of formation and transition - could, in principle, be found (or devised), with which the physical states and causal connections of the physical system could be paired. (cf. Stich 1983:149-151) Is such a physical system ipso facto implementing a 'formal system'? Is it 'rule-governed'? Is it therefore a computer? Any physical system can be so described: can be paired with some formal rule-system, under some method of projection (instantiation function). So if this is a sufficient condition of being a computer, all physical systems are computers, including those that implement the

rule - compute the function - 'Whatever happens, do nothing'. So a human brain, in particular, would be a computer, by this definition. But this would be a vacuous thesis.

So having a functional organisation that can be paired with some system of (syntactic) rules is not a sufficient condition of implementing a formal system, in the sense in which 'human calculators' use or implement formal systems - eg. in using an abacus or a logical symbolism. So this does not explain 'symbol-processing', or being a computational system, in the sense defined by Haugeland et al.

4.10 THE INTENTIONAL CHARACTER OF RULE-FOLLOWING

In each case of rule-following, playing a game, and using a formal system, that I have mentioned, the application of a normative standard to the person (or machine) described as doing these things, involves human intention or intentional behaviour. In some cases, the standards are set by human intention: rules of games are explicitly invented and formulated, or established by intentional behaviour, the practice of playing a game in a certain way. There is a correct way of playing a game only if people have played it that way intentionally, or have intended it to be played that way, whether or not the rules have been explicitly formulated.

For the standards of a game to apply to an individual,

so that what he does can be assessed as correct or incorrect, he must intend to play the game, follow the conventional rules (which is not to say that he must know all the rules). If he moves the pieces on a chess board without intending to play chess, or intending to play a variant with different rules, then he doesn't necessarily make a mistake when he moves a piece in a way not permitted by the standard rules. (This does not necessarily mean that there is some 'mental state', conscious or unconscious, that constitutes the intention. Perhaps the correct analysis would explain the intention as an aspect of the behaviour, rather than a distinct phenomenon. I mean to leave this open.)

A person writing logical symbols in various combinations makes a mistake when she deviates from a standard system, only if she intends to use them in the standard way, in accordance with the rules of a certain formal system. If she intends to use some of the signs in an idiosyncratic way, or intends to follow the rules of a non-standard system, then the assessment of her work must be in accordance with these intentions. Which standards apply, depends on her intentions in performing the activity.

This is so even if a platonic view of logic is assumed. Whether her deductions are in accord with such an absolute standard depends not only on the standard, the platonic objects or rules, but also on the way she intends to use the

signs she employs: whether "v" in her symbolism is the sign for conjunction or disjunction, for example. That behaviour of a certain kind is the implementation of an objective system, depends on the intention involved in the behaviour. If the stars in a certain part of the universe happen to form patterns that look like valid deductions in formal logic, they wouldn't for all that be deductions; and if they happen to form a pattern that looks like an invalid deduction, no mistake would have occurred, however platonic the rules of logic may be, because stars have no intentions, do nothing intentionally; nor as far as we know, are they controlled by the intentions or intentional behaviour of any intelligent being.

Most of the time, we simply use signs in accordance with their normal uses, just as we use chess pieces in the normal way, without having special intentions concerning them. But this supposes the general intention to conform to normal use: conforming is intentional behaviour, even if it's not often the subject of conscious thought. And it presupposes intentional behaviour in general: namely that which constitutes the normal use of the symbols.

This applies also to the assessment of computers. Any distinction of correct and incorrect in the functions of a computer, as implementing rules, or a program, is relative to the intentions of designers, programmers or users. There is no mistake except relative to such standards established

by intentions or intentional behaviour.

4.11 THE DIFFERENCE BETWEEN BRAINS AND COMPUTERS

Could the brain be described as 'manipulating symbols' - or 'tokens' - in a 'rule-governed' way? There is no intentional context for describing it as implementing or instantiating a certain system of rules. No one designed or programmed a human brain with the intention that it instantiate certain rules; nor does anyone use or interpret a brain in a relevant way.

Brain processes are not intentional, nor do any intentions apply to them. We have no conscious intentions concerning what happens in our skulls, and any hypothesis of unconscious intentions would require an explanation that would lead to a regress. No sense can be made of the idea that we intend to implement a certain formal system in our brains. Nor is there any external standard set by intentional design or interpretation, as there is for computers. If there is any sense in which we 'interpret' each other, this applies only to our visible and audible behaviour. Perhaps a brain <u>could</u> be used as a computer, in a way that involved rule-based assessment of its behaviour, but no such use occurs in the normal case.

So the standard that, in the case of a computer, determines what rule it implements - namely, that which it's <u>intended</u> to implement - does not obtain in the case of the brain. In the sense in which a calculator or computer can be said to implement a formal system, this cannot be said of

a brain. There is no standard, applicable to brains, of the kind involved in describing human behaviour and computer processes as 'rule-governed', or as implementations of 'formal systems'. The idea of symbol-processing therefore fails to apply to brain processes. If that is the essential idea of the computational theory of mind, then, it involves a conceptual mistake.

The conclusion I draw concerning computers and the computational theory of mind, is expressed succinctly by John Searle (though my argument is not his): 'syntax is not intrinsic to physics' (Searle 1992:208). This, he explains, is

> because ... syntax is essentially an observer-relative notion. The multiple realizability of computationally equivalent processes in different physical media is not just a sign that the processes are abstract, but that they are not intrinsic to the system at all. They depend on an interpretation from outside. (ibid. 209)

And so, concerning the analogy between computers and brains:

> In the case of the mechanical computer, the whole working system includes an outside homunculus [a programmer or interpreter], and with the homunculus the system is both causal and logical: logical because the homunculus gives an interpretation to the processes of the machine, and causal because the hardware of the machine causes it to go through the processes. But these conditions cannot be met by the brute, blind, nonconscious neurophysiological operations of the brain. (ibid. 220)

4.12 BIOLOGICAL STANDARDS

A biological system - a body or an organ, perhaps an ecosystem - has normal functions, and deviations from them.

We can recognise states or processes that are abnormal, pathological or degenerate, when an organism or organ - or ecosystem - 'malfunctions'. Why shouldn't this be the basis of distinguishing 'correct' rule-governed activity from 'mistakes' in the brain?

The brain consists of neurons which have normal functions. There is an objective difference between a healthy, properly functioning neuron and one that's 'broken', as there is between the functioning of a healthy heart and heart-failure or irregularity. But this is true not only of the brain, but also of any biological organ. So if this is a condition of describing a system as 'rule-governed', it's satisfied by, for example, the liver as well as the brain. But we wouldn't describe the liver as 'symbol-processing', just because its normal functions instantiate <u>some</u> rule-system under <u>some</u> method of projection.

Biological functions break down; organs malfunction; organic processes exhibit abnormalities and pathologies. But there's no <u>mistake</u> involved. The concept of a <u>correct</u> action, which characterises rule-following, has no application to organic processes. A mistake is a failure to draw the right conclusion, to do the correct thing, to follow a rule, to do what one intends or what serves to achieve an intended goal. We speak of heart-<u>failure</u>, but not of a <u>mistake</u> of the heart - not in this sense, anyway -

when for example it pumps irregularly or stops pumping. Similarly a neuron may 'misfire', fail, behave pathologically. but it does not thereby make a mistake.

We could identify pathological effects - breakdowns in normal neural functioning, for example. But this kind of distinction doesn't resolve the problem. Not all pathologies - neural malfunctions - would affect the presumed computational activity. A dangerous rise in the level of some chemical in the neural protoplasm, for example, would be a 'malfunction', something wrong with the brain, but might not affect the pattern of neural impulses - at least in the short term. Not all damage to a computer, even to its processing circuits, affects its computational operations. What counts as a computational error is independent of what counts as physical damage - though of course the latter may cause the former. <u>If</u> there's computational error, then this may be explained by physical damage or degeneration. But this order of explanation assumes that computational error can be identified independently of the physical damage that may be its cause. In the case of the brain, however, there is no independent criterion of error. We can identify error in a computer independently of its physical cause because there's an independent criterion of the rules that govern its symbol-processing: the machine-table it's <u>supposed</u> to implement. So we can't explain error in the brain as what happens when

there's a neurophysiological malfunction, because some such cases would not affect the computational processes, or would not affect them in a relevant way.

A computer is constructed or programmed to implement a certain machine-table or system of syntactic rules. If it deviates from this, an error occurs. This may be due to a mechanical malfunction, or it may be a failure of design or construction. That a mistake occurs is independent of these explanations.

4.13 CHOMSKY REDUX: IN WHICH THE CONCLUSION IS DRAWN

Chomsky's idea, that understanding and speaking a language is partly explained by a speaker's knowledge of a generative grammar - if this hypothesis is understood in the naturalistic way he suggests, as a theory about states or processes in the brain - is illusory. The suggestion that we interpret sentences by deriving their structural descriptions begs the question how we understand these descriptions; in particular, how we grasp their syntax.

The answer to this - suggested by Chomsky and developed by numerous philosophers and cognitive psychologists - is that the brain is a kind of computer, and that understanding the hypothesized structural descriptions - and 'mental representations' in general - consists in their having appropriate functional or causal or computational roles in the brain's operations. I have argued that the

computational model of mind is incoherent. It appeals to notions of 'input' and 'output' that do not correspond to the ordinary notions of perception and behaviour, contrary to the insight of functionalism that mental states such as belief are essentially related to perception and behaviour, in the ordinary sense. The alternative - or complementary - explanation of the computational model, in terms of internal syntactic processes of symbol-manipulation, commits a solecism in applying concepts - such as rule-following - in the absence of the necessary normative framework or background.

I developed these criticisms of the computational model as a theory of <u>belief</u> and other 'propositional attitudes'. But it might be thought that the model is useful as an explanation of knowledge of language, or of grammar, even if it does not provide a general theory of mind. This is suggested by the idea of a 'language module', a part of the brain responsible for processing sentences, deriving their syntactic structural descriptions, in relative independence of other mental processes (eg. Fodor 1983). It is possible to write a computer program to analyse the grammatical structure of sentences, or to distinguish grammatical from ungrammatical strings, so it might seem that this, at least, could be performed in us by a similar process.

The objection to this is that we are not merely sentence-processors. We can believe - or disbelieve - what

we hear; compare what we are told with what we see; formulate our thoughts and feelings in words; describe what we perceive. The hypothesis of a language-processing module is useless unless it can be integrated in a general theory of mind.

The hypothesis of a language-module merely raises the problem of Chapter 2. For suppose there is such a part of the brain, and that for every sentence heard or read it generates a set of structural descriptions. This would not begin to explain the understanding of sentences unless these descriptions were understood. But there are no bounds to the role in one's mental life of the comprehension of a sentence. The syntactic processing might, perhaps, take place in an insulated module, but the result of this processing must have a role in the mind at large. If the module is thought of as a specialised computer, or as a sub-program, its product would be the kind of representation or data-structure employed by computers. This might be a partial explanation of understanding language, if the mind as a whole were explicable in computational terms. In that case, the output of the language-module could be taken up by the general cognitive-computational processes of the brain. But this is just the model, as I have argued, that fails. In this context, the language-processing module would be idle. It would generate structural descriptions, in a computational medium, to which no one has access, and for

which the brain has no use. The hypothesis is idle. A little computer in the head, parsing sentences, or doing maths, might be useful if there were someone to use it, to feed it data and read its output. But our relationship to our brains is not like that.

The appeal to computers as a model of mental processes does not stop the Rylean regress without invoking undischarged homunculi, a point which Fodor in effect concedes:

> A lot is known about the transformations of representations which serve to get information into a form appropriate for central processing; practically nothing is known about what happens after the information gets there. The ghost has been chased further back into the machine, but it has not been exorcised. (Fodor 1983:127)

And therefore,

> If someone - a Dreyfus, for example - were to ask us why we should even suppose that the digital computer is a plausible mechanism for the simulation of global cognitive processes, the answering silence would be deafening. (ibid. 129)

But the idea of <u>peripheral</u> computational processes - mechanisms of 'input', for example - is not just an incomplete model of mind, in this case; it's <u>incoherent</u>: unless there is an interpreter, within or without, it makes no sense to speak of computational processes. If the ghost is exorcised, there will be no computation in the brain, only physiology.

If the problem of understanding (and using) the output of the 'language-processing module' is referred to the

central processor of the brain - a latter-day homunculus - then the problem of understanding recurs. How is understanding in the central processor to be explained? The functional or computational answer does not suffice, as I have argued. So the regress has been deferred but not avoided.

The idea that knowledge of a generative grammar partly explains the ability to speak and understand a language creates this puzzle: How do we understand the structural descriptions that application of this knowledge would generate? The answer implied by Chomsky's commitment to the computer model - and by the related functionalist theory of mind - is that undertanding consists in internal use, functional or computational role. But this answer collapses with the incoherence of the functionalist and computational theories of mind. Assuming Chomsky's conceptual framework, widely accepted in 'cognitive psychology' - representationalism, physicalism, the computational model - knowledge of a generative grammar would therefore not help to explain the ability to understand - or speak - a language. But in that case, there is no reason to think that we have such knowledge.

BIBLIOGRAPHY

Anderson, J R (1983): *The Architecture of Cognition* Harvard University Press, Cambridge MA

Baker, L R (1987): *Saving Belief* Princeton University Press, Princeton NJ

Block, N, ed.(1980): *Readings in the Philosophy of Psychology* 2 vol.s Harvard University Press, Cambridge MA

Block, N (1986): 'Advertisement for a Semantics for Psychology', in French, Uehling, and Wettstein, eds., *Midwest Studies in Philosophy 10: Studies in the Philosophy of Mind* University of Minnesota Press, Minneapolis (1986)

Block, N (1990): 'The Computer Model of Mind', in Osherson and Smith, eds. (1990)

Boden, M. (1988): *Computer Models of Mind* Cambridge University Press, Cambridge

Boden, M, ed. (1990): *The Philosophy of Artificial Intelligence* Oxford University Press, Oxford

Bresnan, J and Kaplan, R (1982): 'Introduction: Grammars as Mental Representations of Language', in Bresnan, J, ed. *The Mental Representation of Grammatical Relations* MIT Press, Cambridge MA

Brand, M and Harnish,R ed.s (1986): *The Representation of Knowledge and Belief* University of Arizona Press, Tucson AZ

Chomsky, N (1957): *Syntactic Structures* Mouton, Den Haag

Chomsky, N (1959): Review of B F Skinner's *Verbal Behavior*, Language 35:26-58, reprinted in Block (1980)

Chomsky, N (1965): *Aspects of the Theory of Syntax* MIT Press, Cambridge MA

Chomsky, N (1980): *Rules and Representations* Columbia University Press, New York

Chomsky, N (1980b): 'Rules and Representations' in *Behavioral and Brain Sciences* 3:1-61

Chomsky, N (1986): *Knowledge of Language* Praeger, New York

Chomsky, N (1986b): 'Knowledge of Language' in ed.s Brand and Harnish

Chomsky, N (1988): *Language and Problems of Knowledge: The Managua Lectures* MIT Press, Cambridge, MA

Chomsky, N and Katz, J (1974): 'What the Linguist is Talking About' *Journal of Philosophy* 71 (12):347-67

Churchland, P (1981): 'Eliminative Materialism and Propositional Attitudes' *Journal of Philosophy* 78:67-89

Davidson, D (1963): 'Actions, Reasons and Causes' in Davidson (1980)

Davidson, D (1970): 'Mental Events' in Davidson (1980)

Davidson, D (1980): *Essays on Actions and Events* Oxford University Press, Oxford

Dennett, D (1978): *Brainstorms* Bradford Books, Montgomery VT

Dennett, D (1986): 'The Logical Geography of Computational Approaches: A View From The East Pole' in Brand and Harnish, eds.

Dennett, D (1987): *The Intentional Stance* MIT Press, Cambridge MA

Dretske, F (1986): 'Aspects of Cognitive Representation' in Brand and Harnish ed.s

Dretske, F (1988): *Explaining Behavior* MIT Press, Cambridge MA

Fisher, J (1974): 'Knowledge of Rules', *Review of Metaphysics* vol. XXVIII (2) 237-260

Fodor, J (1968): *Psychological Explanation* Random House, New York

Fodor, J (1968b): 'The Appeal to Tacit Knowledge in Psychological Explanation', *Journal of Philosophy* 65 (20) reprinted in Fodor (1981)

Fodor, J (1975): *The Language of Thought* Harvard University Press, Cambridge MA

Fodor, J (1980): 'Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology' *Behavioral and Brain Sciences* 3, reprinted in Fodor (1981)

Fodor, J (1981): *RePresentations* MIT Press, Cambridge MA

Fodor, J (1983): *Modularity of Mind* MIT Press, Cambridge MA

Fodor, J (1985): 'Fodor's Guide to Mental Representation' *Mind* (Spring 1985), reprinted in Fodor (1990)

Fodor, J (1987): *Psychosemantics* MIT Press, Cambridge MA

Fodor, J (1990): *A Theory of Content* MIT Press, Cambridge MA

Fodor, J and Lepore, E (1992): *Holism: A Shopper's Guide* Basil Blackwell, Oxford

Foss, D and Hakes, D (1978): *Psycholinguistics: An Introduction to the Psychology of Language* Prentice-Hall, Englewood Cliffs NJ

Garfield, J. ed. (1990): *Foundations of Cognitive Science* Paragon House, New York

Garrett, M (1990): 'Sentence Processing' in Osherson and Lasnik ed.s

Graubard, ed. (1988): *The Artificial Intelligence Debate* MIT Press, Cambridge, MA

Graves, C et al. (1973): 'Tacit Knowledge', *Journal of Philosophy* 70 (11):318-30

Harman, G (1967): 'Psychological Aspects of the Theory of Syntax' *Journal of Philosophy* 62(2):75-87

Harman, G, ed. (1974): *On Noam Chomsky: Critical Essays* Anchor Press, Garden City NJ

Harman, G (1987): 'Conceptual Role Semantics', in ed. Lepore *New Directions in Semantic Theory*

Harman, G (1988): 'Wide Functionalism', in Schiffer, S. and Steele, S. ed.s *Cognition and Representation*, Boulder, Colorado: Westview

Haugeland, J (1985): *Artificial Intelligence: The Very Idea* MIT Press, Cambridge MA

Heil, J (1983): *Cognition and Perception* University of California Press, Berkeley CA

Higginbotham, J (1990): 'Philosophical Issues in the Study of Language' in Osherson & Lasnik, ed.s (1990)

Hornsby, J (1986): 'Physicalist Thinking and Behaviour' in Pettit and McDowell ed.s (1986)

Hurlbert and Poggio (1988): in ed. Graubard (1988)

Jackendoff, R (1987): *Consciousness and the Computational Mind* Cambridge MA: MIT Press

Johnson-Laird, P (1988): *The Computer and the Mind* Harvard University Press, Cambridge MA

Kripke, S (1982): *Wittgenstein on Rules and Private Language* Harvard University Press, Cambridge MA

Larson, R (1990): 'Semantics' in Osherson and Lasnik ed.s

LePore and McLaughlin, ed.s (1985): *Actions and Events* Basil Blackwell, Oxford

Lewis, D (1970): 'How to Define Theoretical Terms', *Journal of Philosophy* 67:427-446

Lewis, D (1972): 'Psychophysical and Theoretical Identifications' *Australasian Journal of Philosophy* 50:249-58, reprinted in Block (1980)

Loar, B (1981): *Mind and Meaning*, Cambridge University Press, Cambridge, England

Marr, D (1982): *Vision* MIT Press, Cambridge MA

McGinn, C (1984): *Wittgenstein on Meaning*, Basil Blackwell, Oxford

Nagel, T (1969): 'Linguistics and Epistemology' in Hook ed. *Language and Philosophy New York University Press*, reprinted in ed. Harman (1974)

Newell, A (1987): *Unified Theories of Cognition* Harvard University Press, Cambridge MA

Newell, A. and Simon, H. (1975): 'Computer Science as Empirical Inquiry' *Communications of the Association for Computing Machinery* 19:113-126 (1976), reprinted in Garfield (1990)

Osherson, D and Lasnick, H, eds. (1990): *An Invitation to Cognitive Science Vol. 1 Language* MIT Press, Cambridge MA

Pettit and McDowell eds. (1986): *Subject, Thought and Context* Oxford University Press, Oxford

Putnam, H (1960): 'Minds and Machines' in Hook, S. ed. *Dimensions of Mind* (1960), reprinted in Putnam (1975a)

Putnam, H (1967a): 'The Mental Life of Some Machines' in Castenada, H. ed. *Intentionality, Minds and Perception*, reprinted in Putnam (1975a)

Putnam, H (1967b): 'The Nature of Mental States' in Capitan and Merrill eds. *Art Mind and Religion*, reprinted in Putnam (1975a)

Putnam, H (1975a): *Mind, Language and Reality: Philosophical Papers Volume 2* Cambridge University Press, Cambridge, England

Putnam, H (1975b): 'Philosophy and our Mental Life' in Putnam (1975a)

Putnam, H (1983a): *Realism and Reason: Philosophical Papers Volume 3* Cambridge University Press, Cambridge, England

Putnam, H (1983b): 'Computational Psychology and Interpretation Theory' in Putnam (1983a)

Putnam, H (1988): *Representation and Reality* MIT Press, Cambridge MA

Pylyshyn, Z (1980): 'Computation and Cognition: Issues in the Foundations of Cognitive Science' *Behavioral and Brain Sciences* 3:154-169

Pylyshyn, Z (1984): *Computation and Cognition* MIT Press, Cambridge MA

Quine, WVO (1974): 'Methodological Reflections', in Harman (1974)

Ramsey, F (1929): 'Theories', in Mellor ed. *Foundations* Humanities Press, Atlantic Highlands NJ (1978)

Ryle, G (1949): *The Concept of Mind* Hutchinson's University Library, London

Schiffer, S (1987): *Remnants of Meaning* MIT Press, Cambridge MA

Searle, J (1972): 'Chomsky's Revolution in Linguistics' *New York Review of Books* 6.29.72

Searle, J (1980): 'Minds, Brains, and Programs' *Behavioral and Brain Sciences* 3:417-424

Searle, J (1990): 'Is the Brain a Digital Computer?' *Proceedings of the American Philosophical Association* Vol 64 (3) 21-37

Searle, J (1992): *The Rediscovery of the Mind* MIT Press, Cambridge MA

Stabler (1983): 'How Are Grammars Represented?' *Behavioral and Brain Sciences* 6:391-421

Stich, S (1971): 'What Every Speaker Knows' *Philosophical Review* 80, 4:476-496

Stich, S (1972): 'Grammar, Psychology and Indeterminacy'

Stich, S (1983): *From Folk Psychology to Cognitive Science* MIT Press, Cambridge MA

Stoutland, F (1985): 'Davidson on Intentional Behavior', in Lepore and McLaughlin ed.s

Turing, A (1950): 'Computing Machinery and Intelligence', *Mind* LIX 433-460

Wittgenstein, L (1953): *Philosophical Investigations* trans. Anscombe, Basil Blackwell, Oxford

Wittgenstein, L (1946-7): *Lectures on Philosophical Psychology*, ed. Geach, University of Chicago Press, Chicago Ill.