

The Physical Analysis of Mental
States and Events

by

Brandon Taylor

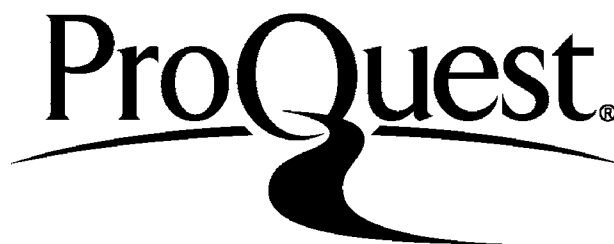
ProQuest Number: 10107307

All rights reserved

INFORMATION TO ALL USERS

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

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



ProQuest 10107307

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

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

Abstract

The purpose of the essay is to arrive at a clear assessment of what the doctrine of physicalism means, as it concerns mental events and mental states; and to exhibit and adjudicate the factors which bear on its truth or falsehood. The main themes of the essay are the reduction of the mental to the physical, and the identity theory.

Chapter I explains why mental language is so apparently indispensable to various theoretical enterprises, and shows why it seems impossible to dispense with mental language in favour of behavioural language. The reductive programmes of Carnap are discussed, and alternatives to them are introduced.

In Chapter II two theories with a physicalistic conclusion are examined, and found wanting. The first is Putnam's theory that mental states are functional states of the organism, and that a Turing Machine table can represent the relation between mental states without implying what physical realisations they have. The second is Davidson's "proof" that mental events are identical with physical events; its main weakness lies in the premiss that the mental and the physical interact causally.

Chapter III is addressed to the ontological question of what mental phenomena there are. The main conclusion is that the evidence suggests that, in a strict sense, there are no mental events. This entails that physicalism is to be best understood in terms of the truth-relations between mental and physical sentences (or equivalently, in terms of the identity either of mental properties with physical properties, or of mental facts with physical facts).

Chapter IV argues that physicalism can only be coherently stated in terms of the nomological equivalences between mental and physical sentences. The arguments which obstruct the truth of this doctrine for the case in

which the physical sentences have a behavioural content cannot be applied to the case in which the physical sentences have a cerebral content. One important general difference between the truth-grounds of mental sentences and the truth-grounds of physical sentences (explained in terms of a consciousness condition) provides an explanation for why a reduction of the mental to the cerebral is the only possibility open for physicalism.

Acknowledgements

I should like to record my gratitude to the following philosophers for the help and criticisms they offered me while this essay was being written: Glen Crowther, D. C. Dennett, Peter Dill, Gareth Evans, Myfanwy Evans, Stanley Godlovitch, Anthony Savile, David Wiggins and Andrew Woodfield. I am also grateful to the officials of the Department of Education and Science and the Senatus Postgraduate Studies Committee of the University of Edinburgh, for their financial assistance.

Contents

	page
<u>Chapter I: Physicalism Introduced</u>	
1. Preamble	1
2. The need for mental terms	3
3. Behavioural reduction	13
4. Alternatives to a behavioural reduction	29
5. Programmatic remarks	34
<u>Chapter II: Two Theories Examined</u>	40
<u>IIA: The "Functional" Theory of Mental States</u>	
1. Introduction	41
2. Exposition of the Turing Machine Theory	47
3. Criticism of the Turing Machine Theory	56
4. Relevance of the theory for physicalism	71
<u>IIB: A Theory of Mental Events</u>	
1. Exposition of Davidson's theory	75
2. Causal interaction between the mental and the physical: reasons as causes	78
3. Causal interaction between the mental and the physical: mental states as causes	90
<u>Chapter III: Grammatical and Logical Complexities</u>	99
1. Description of the problem	100
2. Events in general: the evidence from logic	104
3. The ontology of the mental	120
4. Actions and causes again: some remarks on their grammar	146

	page
<u>Chapter IV: Prospects for an Identity Theory</u>	161
1. Introduction	162
2. The Truth-Value of Physicalism	166
3. The Mental and the Physical	183
<u>Appendix: Identification and Explication</u>	203
<u>Bibliography</u>	220

1. PREAMBLE

One of the central purposes of this essay is to examine the intelligibility and the meaning of physicalistic theories of mental phenomena. Another is to try and reach some conclusions about the truth or falsehood of these theories. But rather than give a cursory treatment to every single complexity which attends this difficult subject, I have chosen to restrict the scope of the essay to a small number of themes which, I believe, are of central importance. I have also tried to discuss them in the detail which they deserve.

It would be otiose to stage a very elaborate introduction to a subject so familiar to modern philosophy as this one is, but for those for whom physicalism is not a permanent or living philosophical problem, a few brief and general opening remarks will help to set the scene for the complexities which are to follow.

There are several passages in both historical and contemporary philosophical writings which might be taken to suggest that physicalism, like so many other non-doctrinal "isms", is nothing more than a habit of thought, or some very general point of view concerning the structure of the world, and the sources of our knowledge about it. But philosophers and philosophically-minded scientists have taken care to devise more exact formulations. Although differing in detail, they all point in the same direction. They range from the consciously physics-oriented formulations of, say, J. J. C. Smart, to the more general hypotheses of those who believe that in some sense or other, everything is ultimately physical in nature. Physicalism, according to Smart, is

"the theory that there is nothing in the world over and above those entities which are postulated by physics and, of course, those entities which will be postulated by future and more adequate physical theories"¹

1. J. J. C. Smart Materialism; Journal of Philosophy 1963 p 651.

while others, accepting the principle that every phenomenon has either a mental or a physical nature, propose that mental phenomena can be "reduced" to physical phenomena, or that mental phenomena are really physical phenomena in disguise, or that in some sense a complete and adequate description of the world and its workings need incorporate no non-physical elements.

Now of course there is a certain amount of clarification to be done before we can regard any of these hypotheses as specific enough for the purposes of philosophical discussion. But the historical context of the theory of physicalism is not so difficult to describe. It will surprise no one who is acquainted with philosophy that one of the closest and most persistent allies which physicalism has had was (and still is) the positivistic movement in the natural sciences. The impressive tendency of the physical sciences gradually to provide explanations for more and more of what happens in the world clearly fortified the belief which lies at the centre of philosophical physicalism that everything is ultimately physical. But the real testing-ground for this belief was, and still is, the empirical study of man - although this is not of course to deny that physicalists of any persuasion encounter special and distinctive problems with colours, sounds, aesthetic qualities, and so on.

What physics could account for in the natural world, clearly, the better it seemed for the theory of physicalism. But there is a sense in which people are not yet part of the natural world, for it seems that their behaviour and mental life cannot, in some sense of cannot, be predicted and explained like other phenomena. It is not just a fact to wonder at, but a fact of the greatest importance, that the promised science of man has yet to make its appearance; indeed the world still awaits even the first small signs of its appearance. It is even fair to say, I think, that as far as the philosophical theory of physicalism is concerned, this fact has equally and oppositely compensated for the fortifying progress of the

natural sciences in other spheres. But of course neither of these opposing facts by itself has been sufficient to settle the issue. Supporters of physicalism will continue to sustain their beliefs on inductive grounds, arguing that the only rational thing to believe is that physical explanations can be found for everything;² while opponents of physicalism will continue to sustain their convictions largely on the grounds that theoretical reasons can be advanced to show why man is, and must be, exempt from science. Other less rational beliefs tend only to obstruct the issue.

Clearly, neither of these arguments by themselves is really sufficient to settle either the truth of physicalism or its falsehood. But my hope is that they give some credence to the idea that some of the most crucial points of discussion are methodological in character. In a way, I think, this is not surprising: for it is the very fact that we do at the moment lack a science of man that surely makes speculation about the possible truth of physicalism a matter worth worrying about in philosophy. In other words, it is the very existence of this gap which makes it inviting to speculate about what, if anything, might fill it.³

2. THE NEED FOR MENTAL TERMS

The interest which philosophers have developed in physicalism as a general metaphysic, and the ingenuity they have expended on refining this

-
2. This is the argument from the improbability of nomological "danglers". See again Smart, *op cit.* p. 661.
 3. If there is any truth in the "incubation" theory of philosophy, then we might view the philosophical arguments about the methodological nature of physicalism as the best kind of preface which a developed science of persons could have. If such a science were to emerge from such a preface, then we might be able to echo the thoughts of J. L. Austin, who himself subscribed to the incubation theory. As Austin put it: "Then we shall have rid ourselves of one more part of philosophy, in the only way we ever can get rid of philosophy, by kicking it upstairs". Ifs and Cans (In Philosophical Papers, eds. Urmson and Warnock p. 282).

general doctrine so as to give it a precise form, can be easily explained. One explanation can be traced to a dissatisfaction with the traditional forms of dualism: monism, in some form, has seemed a more attractive goal than dualism, simply owing to the greater economy and neatness of the former. And in a world in which physical explanations of natural and non-natural phenomena seem more and more readily available as time goes on, it becomes naturally tempting to suppose that a single system of thought, or a single body of theory, might suffice to explain every single phenomenon that human as well as natural life comprises.

But there is a different idea, which goes some way towards explaining why physicalism has seemed a tempting goal for philosophical theory, and this is the suspicion that mental language, in its various forms, plays a systematic and indispensable rôle in organising experience and in constructing theory. Of course we have, at present, no very precise way of saying what mental language is; although it may come as a surprise to learn that philosophers who share the suspicion that mental language plays an indispensable role are not, in general, very embarrassed by this fact. They know, or have well-founded intuitions, about some of the things mental language contains if it contains anything, and they simply see these paradigm or central linguistic constructions as being of a sort without which little progress could be made in speaking theoretically about phenomena in the way we want. But it does seem to me that one of the tasks of any really adequate appraisal of physicalism is at some stage to make this suspicion about the character of mental language exact - to find out what it is about this language that makes it indispensable. It follows from this that the lack of a clear view of how mental language differs from physical language, or alternatively of how mental phenomena differ from physical phenomena, ought to seem a good deal more of an embarrassment than it actually does. I shall try to reach some conclusions about both

these matters in the course of my essay. But at this stage I shall merely give an exposition, somewhat informally, of some of those theoretical problems which give rise to the suspicion that mental language plays a role in explanation which is both an importantly systematising one, and, in some sense of the word, a seemingly indispensable one as well. The general area in which those theoretical problems exist is the explanation of people's actions, where by "actions" I mean to include linguistic actions ("speech-acts").

Consider, for example, that part of linguistic theory whose aim it is to give a description of linguistic structure - the structure of sentences. This enterprise is part and parcel of the theory of action understood in the widest sense, for sentences are what people utter in their attempts to perform speech-acts.

The first thing to notice when considering this question of the description of linguistic structure is that what qualifies as an adequate description of something depends on the goal, or purposes, of the description. We cannot talk about the adequacy of a description unless we specify the task in question. Chomsky, in Aspects of the Theory of Syntax⁴ and elsewhere, has fathomed this question of adequacy in some detail, and has elucidated various different concepts of adequacy for a linguistic theory which depend on the task in hand. His proposal was that a theory is adequate to direct observation if it enables the theorist to draw up a descriptive list of the sentences and other linguistic units which are actually produced. This he called observational adequacy. He secondly

4. Chomsky: Aspects of the Theory of Syntax, Chapter I.

proposed that a theory is adequate relative to a descriptive task if it enables the theorist to "give a correct account of" (as he would put it) the linguistic intuition or competence of the idealised speaker-hearer.⁵ And thirdly, he proposed that a theory is adequate to the task of explaining this intuition if it enables the theory to select between any two different grammars which are both adequately descriptive of the speaker-hearer's linguistic competence. Now clearly what anyone wants the linguist ultimately to succeed in is not just the first and second of these tasks, but the third as well. It was Chomsky's suggestion that a theory adequate to this third task must contain terms for describing a corpus of innate knowledge of the language: both knowledge how and knowledge that, it seems. And here we have the introduction of mental phenomena into the explanation.

This may all seem a bit schematic. Chomsky's problem can be alternatively described in the following way. When a person speaks, what he utters is a sentence or group of sentences. And sentences, by definition, have meaning. But what physically happens when a person speaks is that some series of acoustic events takes place; and acoustic events are not things which can be said to have meaning in the way sentences can. Besides, although each sentence-utterance event is an acoustic event, it is obviously not true that each acoustic event is a sentence-utterance event. So the question arises: when does a particular acoustic happening constitute the utterance of a sentence, and when it does, how can it do so? What is the connection between the fact that a certain acoustic event took place and the fact that a certain sentence was uttered? Or, finally, what is the connection between the particular meaning which makes a sentence the sentence it is, and the particular acoustic configuration which forms its

5. In his Current Issues in Linguistic Theory p. 62 (In Fodor and Katz: The Structure of Language pp. 50-118). See also Aspects pp. 30-37.

physical basis on any occasion of its utterance?

Chomsky's answer was that sentences are structured not merely by simple phrase-structure rules, but by transformational rules too.⁶ This complex system of generative rules, both semantic and systactic, is said to form the content of a kind of innate or tacit knowledge which every speaker of the language has. Or as Katz puts it: a correct linguistic model is one which

"pictures the structure of the system and its unobservable components. In this way, a linguist can assert that his theory correctly represents the structure of the mechanism underlying the speakers ability to communicate with other speakers."

The mental capacities with which the speaker has to be ascribed are, presumably, the capacities needed to apply those generative rules, and that knowledge, in actual speech and the perception of speech. To many people it seems unclear just what the nature of the requisite linguistic knowledge is, and just what the nature is of the linguistic skills which enable a speaker to apply this knowledge. None the less, so generative grammarians argue, language-users must be ascribed with mental and cognitive powers of some rather complex kind in order for their linguistic abilities to be explained.

According to generative grammarians, therefore, there is no way of explaining the connection between acoustic events (which are events described physically) and the events consisting of the utterance of sentences, where sentences are items having meaning, without invoking rules and principles of an essentially mental character. However difficult it is to understand clearly in what sense those rules and principles do describe linguistic structure, Chomsky's attack on the purely physical and

6. Chomsky: Syntactic Structures, passim

7. Katz: Mentalism in Linguistics p. 128. *Language* (40) 1964. My underlining.

classificational approach to grammar, to which his theory provides an alternative, does seem to have been successful. It suggests the general implausibility of describing the phenomena of language-use in a way which does not involve the use of mental terms.

Chomsky's success in the negative enterprise of revealing the inadequacies of the non-mentalistic approach can be clearly seen, I think, if we look at some of the details of those non-mentalistic approaches themselves. There used to be an influential method in linguistics, originated by Bloomfield and extended by Zellig Harris, known as the classificational or taxonomic method. This method, as Katz later described it,

".... holds that every linguistic construction, at any level, reduces ultimately, by purely classificational procedures, to physical segments of utterance"⁸

The physical segments of utterance for the taxonomic linguist were sounds (phones). Classes of sounds, or classes of significant groupings of them, were classified as phonemes; so that the relationship of a sound or sequence of sounds to the phoneme was one of class-membership. Morphemes, similarly, were thought of as classes of sequences of phonemes; and sentences, ultimately, consisted of morphemes; so the relationship of phoneme-sequence to morpheme, and of morpheme to sentence, was again one of class-membership. This was the basic theoretical structure employed by Bloomfield. From this basis Harris developed a method of linguistic description which he called the distributional method. Like Bloomfield, he saw language as being describable as four levels: the levels of sound or phone, of phoneme, or morpheme, and of sentence; with the relation of a unit at one level to a unit at the next highest level being one of class-membership. But he went further than Bloomfield, and developed

⁸ Katz op cit p. 124.

what he thought of as a potentially mechanical method of uncovering ("a large part of") sentence-structure - based on the distribution of units at the lower level throughout a corpus of antecedently elicited grammatical sentences. Consider this example of the segmentation of a phoneme-sequence into morphemes.⁹ The phonemic representation of the sentence He's quicker is /hiyzkwiker/. Among the many methods of tracing the distributional structure of this sequence is to ask how many different phonemes occur after the n-th phoneme in the sequence; then a peak in the number of different successor-phonemes is supposed to mark a semantical or syntactical break, by indicating the independence of the phoneme at which the peak occurs from the phonemes which follow it. The peaks in the phonemic representation of He's quicker come at /y/, /z/, /k/ and /r/, thus correctly showing the morpheme or word-boundaries at (i.e. immediately after) those points. In a precisely analogous way to this, Harris supposed that a great proportion of sentential structure - syntax - could be described.

In fact, in the example I have described, it seems that assumptions about meaning do play a quite central role. It has been assumed, for instance, that a peak in the number of successor-phonemes at the n-th phoneme indicates the semantic or syntactical independence of the sub-sequence of which that phoneme is the terminal phoneme from the other sub-sequences within the whole. Objections aside, at any rate, Harris certainly saw his method as an essentially mechanical procedure for uncovering structure. "The method as a whole", as he described it,

"can be viewed as part of an orderly set of kindred methods capable of yielding a large part of language structure in terms of the relative occurrence of sounds, these occurrences being the physical events of language" 10

-
9. Cf. Harris: Phoneme to Morpheme, Language (31) 1955 p. 190. For the distributional method at other levels see Discontinuous Morphemes, Language (21) 1945; Morpheme to Utterance, Language (22) 1946; Distributional Structure, Word (10) 1954.
10. Harris: Phoneme to Morpheme pp. 212-3.

It was assumed by Bloomfield, as by Harris, that the distributional method, and the taxonomic method in general, did not require a stock of theoretical terms of any great variety or richness, and that it certainly did not require mental terms. But subsequent theory saw that the limited number of theoretical terms which they did allow themselves enabled the theorist to achieve Chomsky's first, observational level of adequacy, while falling short of the descriptive level of adequacy in several respects. For firstly, the method is not generative. It can only describe previously elicited well-formed sentences, and therefore contains none of the resources of a theory of sentence-structure in general. The un-uttered or unspoken sentences, for instance, are beyond its reach.

Secondly, Chomsky has argued that even the structural descriptions which the taxonomic method gives for the previously elicited sentences are inadequate. This is because an analysis into immediate constituents, which is what the taxonomic method provides, can never reveal all the so-called grammatical relations (e.g. Subject-Verb, Verb-Object, etc.) within any sentence. According to Chomsky, only a grammar containing a base component and transformation rules can describe all these relations.¹¹

These criticisms of Chomsky's, then, concern the programme of analysis into immediate constituents - which is the kind of analysis that grammar conceived along classificational or taxonomic lines was supposed to provide. When to Chomsky's critical attack on this programme is added the claim (which is often made by modern linguists) that transformational-generative rules are "psychologically real" while taxonomic rules are not, the result is an argument to the effect that mental or psychological concepts are an essential part of an adequate

11. For this famous argument, esp. for the case of John is easy/eager to please, see Chomsky Syntactic Structures (1957); Perception and Language (in Boston Studies in the Phil. of Science 1961/2 ed. Watofsky 1963) pp. 199-205.

theory of language. We may lack a good understanding of the concept of the psychologically real; but when this deficiency is made good, the inadequacies of taxonomic linguistics as compared to the transformational-generative approach will add up to a strong case in favour of a linguistic theory which is actually descriptive, in some clear sense, of mental operations and capacities.

Another argument for the apparent indispensability of mental terms ought to be mentioned briefly here. The argument, which is basically due to Grice, says that the meanings of language-elements (words, sentences) can themselves be best explained by reference to what a speaker means by producing an utterance on a particular occasion, and in such a way as not to presuppose the concept of word- or sentence-meaning. Utterers' meaning, in its turn, is to be explained in terms of the utterer's intentions in producing the utterance he does produce; and so the concepts of word-meaning and sentence-meaning are ultimately to be explained in terms of the utterer's complex intentions towards his audience.¹²

It seems initially that we will never be able to prove in a completely conclusive way that either Chomsky or Grice were right in what they said about the need for mental terms, or the need for principles which describe mental structure; for essentially they are trying to establish the negative claim that no theory not containing mental language could ever be produced which would be adequate to the subject-matter in question. But

12. Grice: Meaning, Phil. Review (66) 1957 pp. 377-88; Utterer's Meaning, Sentence-Meaning and Word-Meaning. Foundations of Language (4) 1968 pp. 1-18. Strawson, in Meaning and Truth, (Oxford, Clarendon Press 1970) has defended the view that the approach to sentence-meaning via semantic and syntactic rules which generate truth-conditions for a sentence must ultimately rest upon the notions of intention or belief, for, on his view, we cannot elucidate the concept of the truth-conditions of a particular sentence without reference to the speech-act of saying something true (or: making a true statement) - which is where intention and belief essentially reside. See also J. Searle: Chomsky's revolution in Linguistics, New York Review of Books, June 29th 1972.

we surely have to admit that both these approaches to the problem of sentence-meaning - Chomsky's approach which links sentence-meaning to sound-sequences via a description of mental structure, and Grice's approach which links sentence-meaning to the concept of speakers intention - do have a strong tendency to suggest that a vocabulary of mental terms, or more broadly speaking a vocabulary of terms for mental capacities and functions of one kind and another, is indispensable, in some sense, for an adequate theory of language and language-use.

Mental terms seem indispensable for any adequate theory of sentence-meaning. It ought to be possible to link this fact with the proposal which Brantano originally made, and which Chisholm has explored in more detail, that mental language as a whole is "holistic" and "irreducible". This proposal says, in effect, that we cannot provide necessary and sufficient law-like conditions for the truth of any mental sentence without employing terms which are themselves mental. I shall explain this phenomenon at greater length in the next section, when I give a more detailed exposition of the failure of the behaviourists' approach to mental phenomena. In this section I have tried to establish, by example, the plausibility of the general principle that mental terms are indispensable to the construction of adequate theories in many branches of science.

It used to be the case not many decades ago, that theories in general, and theories of behaviour in particular were derided if they were found to employ terms which referred to specifically mental capacities or the operations of specially designated faculties of the mind. And it seems that the derision which such theories received was due to the fact that general requirements of empirical verifiability (or falsifiability) were not and could not be satisfied. This critical phenomenon appeared as part of the general phenomenon of positivism in the sciences and philosophy.

By hindsight it seems a strange phenomenon, since the methodology of this part of positivism was never made quite as watertight as its proponents could have wished. Perhaps the criticism of theories on the grounds that they were mentalistic or psychologistic can be seen, again by hindsight, as a disguise for a different kind of criticism, namely the criticism that the particular mentalistic terms which were employed often had little or no real systematising power. That is to say, the criticism that the theories in question were of the wrong methodological sort may have become confused with the quite different criticism that they were just false, or inadequate as theories as far as predictive power was concerned.

But whatever was the case then is different now. Mentalism in science and psychology is back in the ascendant. It remains a task for philosophy to discover not only why it is that mental language poses such problems of analysis as it does; it also remains a philosophical task to discover whether there does exist some form of analysis in physical terms which is philosophically acceptable. In addition, it is a philosophical task to clarify what form such an analysis should take.

3. BEHAVIOURAL REDUCTION

So far the only statements of physicalism which we have before us are the statement of Smart which I quoted in the first section, and its slightly less technical relatives which I gave shortly afterwards. We now need to look at some formulations which are more detailed. However, there is a broad but important distinction which I believe must occupy a position at the centre of any such investigation, and so before introducing any of the more detailed hypotheses, I shall explain what this distinction is, and say how it divides such hypotheses into two exclusive categories. Only then shall I bring the hypotheses themselves into sharp focus.

I have already said that any really adequate appraisal of the theories of physicalism must involve an exact statement of what the difference is between mental phenomena and physical phenomena, or what the difference is between mental language and physical language. It would be natural to expect such a statement to precede the examination of the theories themselves. But this is not the strategy I shall adopt in this essay, for the reason that the philosophical problems inherent in tackling such concepts as physical phenomenon, mental phenomenon, physical or mental vocabulary, are exceedingly complex, and generally underestimated in modern discussions. So I propose to leave these important questions until later points in the essay (see Chapters III and IV), in the hope that the answers which emerge there will not alter the validity of the conclusions arrived at in these earlier sections.

The distinction which divides physicalistic theories of mental phenomena into two categories is the distinction between the reductive and the non-reductive. There are several accounts of the concept of reduction in the philosophical literature, but fortunately it is not important for the purposes of this essay to adjudicate their competing claims. Perhaps the most general and agreed characterisation of the reduction of one theory to another is given as follows. A theory T_1 can be said to be reducible to a theory T_2 , if all the theorems and postulates of the reduced theory T_1 can be deduced from a conjunction of the theorems and postulates of the reducing theory T_2 together with certain identificatory principles of "bridge laws", where these laws contain concepts belonging to both T_1 and T_2 . An example given by Nagel¹³ is the reduction of Thermodynamics to Mechanics, where the following manoeuvre (amongst others)

13. Nagel: The Meaning of Reduction in the Natural Sciences. In Danto and Morgenbesser (eds), Philosophy of Science, pp. 288-312.

takes place. There is a law of Thermodynamics known as the Boyle-Charles law, and there is a law of Mechanics which connects the pressure of a volume of gas with the average kinetic energy of its molecules; the former law can be deduced from the latter once the identificatory principle is accepted that temperature is mean molecular kinetic energy.

Taking the reduction of Thermodynamics to Mechanics as our paradigm, we can say that what a reduction of one theory to another is designed to do, amongst other things, is to enable a deduction of all the postulates, basic assumptions and general truths of one theory from those of another; the bridge laws or identificatory principles serving the purpose of enabling the deduction to take place.

Some explanatory remarks are in order about this notion of a bridge law. Firstly, a bridge law has the following important characteristics: its antecedent open sentence describes the same property as its consequent open sentence does (because property-identity is what reduction establishes), and the terms in which that property is described in the two open sentences belong, to put it not very precisely, to different levels of inquiry. These characteristics are present, for instance in the bridge law that

$$N(x)(x \text{ is water} \equiv x \text{ is H}_2\text{O})$$

where the terms "water" and "H₂O" belong to different levels of inquiry, and where the phrases "being water" and "being H₂O" determine the same property in different ways.¹⁴ A second important fact about bridge laws is that not every law of nature is a bridge law, and there are many statements of natural law from whose truth we cannot deduce that one property reduces to another. A simple example can show this. It is a law of

14. The prefix "N" stands for the words "It is a law of nature that ...", and is used throughout this essay in giving a statement of any natural law.

nature that if anything has just been born, then it will eventually die, and conversely; and yet the property of being born is not thereby reduced to the property of eventually dying. Nor can a bridge law be a causal law: a law which allows us to predict a phenomenon of the sort mentioned in the righthand side of the law from the occurrence of a phenomenon of the sort mentioned in the left-hand side. This is because, in a causal law, the phenomena between which the causal relation holds must be distinct phenomena, whereas in a bridge law one and the same phenomenon (in general a property) is mentioned under two descriptions.

A third characteristic of a bridge law, and this time one which is shared by natural laws in general, is that they assert a stronger sentential equivalence than that of mere material equivalence. A statement of material equivalence is written without the "N" prefix, thus:

$$(x) (x \text{ is } M \equiv x \text{ is } P)$$

since such a statement only asserts that the truth-value of the contained sentences (or their instances) happen to correspond. A statement of law, or a statement of nomological equivalence between sentences, on the other hand, asserts that the contained sentences (or their instances) are non-accidentally or systematically equivalent in truth-value. Statements of material equivalence are extensional, in the sense that co-extensive terms can be inter-substituted "salva veritate", while statements of natural law are not.¹⁵ Another well-known difference between statements

-
15. The difference between extensional truth and nomological truth can be easily demonstrated: a statement of the form
 $(x)(x \text{ is } M \equiv x \text{ is } P)$
 is equivalent in its truth-conditions to a statement of the form
 $(x)(x \text{ is } M \equiv x \text{ is } P \text{ or } x \text{ is a unicorn})$
 on the assumption that there are no unicorns. But for a law-statement
 $N(x)(x \text{ is } M \equiv x \text{ is } P)$ a similar equivalence
 $N(x)(x \text{ is } M \equiv x \text{ is } P \text{ or } x \text{ is a unicorn})$
 is clearly disallowed. Otherwise it would have to be concluded that the property of being M was reducible to the property of being P or being a unicorn.

of material equivalence and statements of nomological equivalence is that only the latter support counter-factual propositions.

While material equivalence is a weaker relation than nomological equivalence, necessary equivalence is stronger. A statement of the form " Np ", if true, is true in every nomologically possible world, while a statement of the form " $\Box p$ ", if true, is true in every possible world, which include those which are logically possible but not nomologically possible. So statements of nomological equivalence, written

$$N(x)(x \text{ is } M \equiv x \text{ is } P)$$

while stronger than statements of material equivalence, written

$$(x)(x \text{ is } M \equiv x \text{ is } P)$$

are weaker than statements of necessary equivalence, written

$$\Box(x)(x \text{ is } M \equiv x \text{ is } P)$$

When a statement of nomological equivalence is a bridge law, then we can infer from it a statement of property-identity: "being M = being P ". Whereas from a statement of material equivalence we can only infer a statement asserting the co-extensiveness of property-words: "'being M ' is co-extensive with 'being P '; while from a statement of necessary equivalence we can, I think, infer a statement asserting the synonymy of property-words. Thus $\Box(x)(x \text{ is male and unmarried} \equiv x \text{ is a bachelor})$ implies that the property-words "being male and unmarried" and "being a bachelor" are synonymous.

/order to

So much for the prefatory logical considerations which are needed in

fix the concept of a property-reduction. Unfortunately, not all cases where a reduction is actually proposed are quite so straightforward or well-organised as our paradigm. In the case of the physicalists' attempt at reducing the mental to the physical, the situation is typically as follows: we have a cluster C_1 of general truths about organisms expressed in mental terms, and we have another cluster C_2 of general truths about organisms expressed in non-mental terms. Clusters C_1 and C_2 , in the present state of our knowledge, hardly constitute theories, let alone axiomatised theories. All the same, the reductive exercise is to arrive at bridge laws or identificatory principles which, in conjunction with the truths expressed in physical terms, entail all the truths expressible in mental terms. So what is required are statements of the following form:

$$N(x)(x \text{ is } M \equiv x \text{ is } P)$$

where now we specify that the term "M" belongs to mental terminology and the term "P" belongs to physical terminology.

Now a reduction of the mental to the physical, understood in these terms, could be attempted in two ways. The first way is to find laws of the appropriate type linking mental phenomena to behaviour, and the second way is to find laws of the appropriate type which link mental phenomena to cerebral phenomena - phenomena of the brain or central nervous system. Reductions of the first type (I shall call them behavioural reductions) were advanced by members of that school of theory known as the behaviouristic school, while reductions of the second type (I shall call them cerebral reductions) have been contemplated in more recent years. There is a good reason for the appearance of cerebral reductions; for as I shall now show in some detail, attempts to produce behavioural reductions have one notable feature, and this is that they are bound to issue in

failure.

Efforts to effect a behavioural reduction began when behaviourism began. But although behaviourists sought connections ("conceptual" as well as nomological) between mental and physical terms, their avowed aim in doing so was to "translate" all the sayable things about an organism into the purified vocabulary of the physical language, so ultimately dispensing with mental language altogether. They generally explained their task in these terms rather than in terms of a mental-to-physical reduction; but since the laws they typically tried to find were of the bridge-law type, rather than, say, of the causal-law type, it was a reduction in the sense I have explained which in fact they were after. When we remember that bridge laws establish the identity of properties, we can see that the concept of a translation of mental language into physical language was an appropriate one to use, for if a mental property can be shown to be a physical property, then presumably any non-modal context containing a mental-property term can be translated into one with the physical-property term in its place.

There are different descriptions of the behaviourist's programme which make less immediate sense. J. A. Fodor has recently said, for instance, that:

"To qualify as a behaviourist in the broad sense of that term, one need only believe that the following proposition expresses a necessary truth: for each mental predicate that can be employed in a psychological explanation, there must be at least one description of behaviour to which it bears a logical connection" 16

But it is extremely unclear what it would be like for a mental predicate and a description of behaviour to be "logically connected". A predicate of any kind, strictly speaking, cannot have logical connections with

16. Fodor Psychological Explanation p. 51.

anything; and if we speak in terms of psychological open sentences, under universal or existential closure, then it is hard to see that any behavioural open sentence, under a corresponding closure, is likely to stand in a logical connection to it, if the phrase "logical connection" is taken in any of its usual strict senses.¹⁷ It is altogether more satisfactory to represent the behaviourist as one who engages in the task of translating mental language into physical language, and who does this, inter alia, by trying to establish bridge-law statements of the form " $\forall(x)(x \text{ is } M \equiv x \text{ is } P)$ ".

This certainly seems to have been Skinner's programme in his early book The Behaviour of Organisms. In that work he wrote:

"In approaching a field for purposes of scientific description, we meet at the start the need for a set of terms. Most languages are well equipped in this respect, but not to our advantage. In English, for example, we say that an organism sees or feels objects, hears sounds, tastes substances, smells odours, and likes or dislikes them; it wants, seeks and finds; it has a purpose, tries and succeeds or fails; it learns and remembers or forgets; it is frightened, angry, happy or depressed; asleep or awake; and so on. Most of these terms must be avoided in a scientific description of behaviour The important objection to the vernacular in the description of behaviour is that many of its terms imply conceptual schemes the vernacular is clumsy and obese; its terms overlap each other, draw unnecessary or unreal distinctions, and are far from being the most convenient in dealing with the data The ... criterion for the rejection of a popular term is the implication of a system or of a formulation extending beyond immediate observations" 13

In spite of his doubts about some of the ordinary mental predicates of English, Skinner clearly took it to be the task of psychology to explain behaviour in the quite ordinary sense of that term; as he expresses it,

17. It was once assumed by those who argued that desires could not cause actions that a predicate like "... desires to A" stood in a conceptual connection to a predicate like "...does A", where "A" is a phrase for an action. If this is the kind of "connection" Fodor has in mind, then all I can say is that it seems extremely doubtful whether an exact and precise description could be given of what sort of connection it is.

18. Skinner, B. F. The Behaviour of Organisms pp. 6-7.

"behaviour is what an organism is doing" (p. 6). So he clearly must have thought that a general explanation of behaviour could be achieved by employing just the relatively non-mental terms of English which conformed to his particular methodological standard. In fact he seems not only to have thought that the same explanation could be given with his reduced stock of non-mental terms as that which could be given with all the terms of English (both non-mental and mental); but he also seems to have thought that a greater degree of systematisation could be achieved by using only the former.

Other examples could be found in the literature of behaviourism to illustrate how methodological considerations of one kind or another led psychologists to restrict their own lexical resources to a certain minimum stock (on the whole a non-mental stock), while at the same time believing that they could accomplish any serious project in the behavioural sciences as effectively with limited resources as they could with all the resources of the language - mental and non-mental alike.¹⁹ Carnap was the first philosopher to explore this variety of behaviourism with any degree of logical rigour. As he said in The Methodological Character of Theoretical Concepts, behaviourism, together with the philosophical tendencies of early positivism,

"led often to the requirement that all psychological concepts must be defined in terms of behaviour or behaviour dispositions"²⁰

This was the view he explored in Testability and Meaning,²¹ once his earlier flirtation with a complete reduction of psychological concepts to a

19. Skinner's methodology was shared for instance by Neurath. See his Einheitswissenschaft und Psychologie (Vienna: Gerold and Co., 1933); Foundations of the Social Sciences (Univ. of Chicago Press, 1944). See Hempel's Logical Positivism and the Social Sciences (in The Legacy of Log. Positivism, eds. Achinstein and Barker) for other references to Neurath's views on physicalism.

20. Minnesota Studies I (pp. 38-76) p. 71.

21. Carnap: Testability and Meaning, Phil. of Sci. 3 (1936) pp. 419-71, and 4 (1937) pp. 1-40.

phenomenalistic basis had proved unsuccessful.²² The first exploratory proposal of Testability and Meaning was that for any psychological sentence there existed some non-psychological sentence describing behaviour having the same truth-conditions. This correspondence in truth-value would, Carnap supposed, either be guaranteed by natural law or else by the nature of the concepts themselves. In other words, the statements connecting the mental with the physical could be either statements of natural law, or they could be analytically or conceptually true statements in some sense of those battered terms. If such statements could be found, then Carnap supposed that we could translate any context containing the mental terminology into one containing the physical terminology, and without "loss of content".²³ Now I do not think we need to examine the goals of this programme in order to appreciate how peculiarly afflicted the means of achieving them turned out to be. For as Carnap himself came to see, this doctrine of the "explicit definability" of psychological terms in behavioural terms seems doomed to falsehood. On the face of it, the reason looks purely factual: for any psychological open sentence you choose, it seems that there simply does not exist a non-psychological sentence with the same conditions of truth, at least if the non-psychological sentence contains only terms belonging to what Carnap called the "thing language", i.e. "that language which we use in every-day life in speaking about the perceptible things surrounding us".²⁴ It seems wholly improbable that we could find a statement having the character of

22. The program of Der Logische Aufbau der Welt (Berlin-Schlachtense Weltkreisverlag 1928)

23. Carnap's view was that one statement could translate another "without loss of content" if they both implied the same observation-statements. See his Psychology in Physical Language, pp. 165-98 in Ayer, A. J. (ed.), Logical Positivism. (Originally published in Erkenntnis 1933 as Psychologie in Physikalischer Sprache).

24. Carnap: Testability and Meaning. Philosophy of Science 3 p. 466.

(1) (x)(t)(x wants a biscuit at t \equiv x's arm will extend towards
some biscuit shortly after t)

A statement of this character does not even approach to being true without an amendment to the effect that x's arm will extend towards something believed by x to be a biscuit - or without some amendment along these lines which includes a mental word. And if such an unamended statement as (1) is unlikely to be extensionally true (true by material equivalence), it is even more unlikely that a law-like version could be found.

It would be over-hasty to abandon the "explicit definition" approach without a consideration of some of its possible variants. The first variant comes from reducing the generality of statement (1). Instead of trying for a statement which connects a mental property with a behavioural property (in the case of (1), the property of wanting a biscuit and the property of extending the arm towards some biscuit), we could try for a statement which pertains to a single individual only:

(2) (t)(Fred wants a biscuit at t \equiv Fred's arm will extend towards
some biscuit shortly after t)

Such a statement has little to do with properties, as I shall explain later, but its defects parallel those of statement (1): for nothing is likely to prevent there coming a time when Fred has false beliefs about biscuits, but if this is possible then a mental qualification is again needed in the behavioural sentence.

Whether there is any system in these failures remains to be seen. To see how likely it is that there is, however, consider yet another variation on the "explicit definition" theme. The variation I have in mind here is one which Carnap did not consider; it consists of specifying not just a single non-psychological sentence on the right hand side, but a lengthy disjunction of non-psychological sentences: for it certainly is

the case that the physical motions accompanying a desire for a biscuit are exceedingly various. Could we devise a sentence having something like the following logical structure?

- (3) $N(x)(t)(x \text{ wants a biscuit at } t \equiv x\text{'s arm will extend towards some biscuit after } t \text{ or } x \text{ will salivate or } x \text{ will or)$

The question at issue here is partly the question of whether a statement of law could be disjunctive in form; but only partly. For the fact, if it is one, that there are laws in existence which contain short two- or three-termed disjunctions (I have in mind one formulation of Newton's law of motion: " $N(x)(\text{no resultant force acts on } x \equiv x \text{ remains stationary or } x \text{ moves uniformly in a straight line})$ ") would not provide much encouragement for the physicalist, because his disjunctions would almost certainly be exceedingly lengthy, and almost certainly open-ended. Newton's law of motion, for instance, which has the disjunctive form " $N(x)(x \text{ G} \equiv x \text{ H or } x \text{ J})$ ", is extremely compact compared with the type of disjunctive statement which the behaviourist would have to devise. It would seem that the hopes for this disjunctive version of the physicalist thesis are effectively dashed, therefore, not so much by the spectre of disjunction as such, as by the spectre of its lengthiness and its open-endedness.

None the less, the physicalist might still be encouraged by a realisation of the following fact: that every law-statement of the simple form which $N(x)(x \text{ is H}_2\text{O} \equiv x \text{ is water})$ has, can be arbitrarily converted into a statement which has a disjunctive aspect, and which, compared to the Newtonian example, is neither compact nor closed at its right-hand end, e.g.:

$N(x)(x \text{ is H}_2\text{O} \equiv x \text{ is parcel A of water or } x \text{ is parcel B of water or } x \text{ is parcel C of water or etc.)$

This fact might lead him to believe that the converse operation could be effected to re-convert a long-ish physicalistic disjunction into a single unified term. But the physicalist would, I contend, be wrong to put any faith in this hope, and for two reasons. The first is that for any law which is either disjunctive or convertible-to-disjunctive, there exists a certain kind of unity among the concepts represented in the disjunct. In the case of Newton's Law of Motion, for instance, the concepts of rest and uniform straight-line motion belong together (for according to the more recent view that motion is always relative, are even the same concept); while in the case of the disjunctive version of the law about H_2O and water the elements of the disjunct are conspicuously unified by the notion parcel of water. In the case of the long disjunction attempted by the physicalist, on the other hand, it seems extremely unlikely that there is a single concept which coheres the individual disjuncts and which is expressible purely in physical terms.

This brings us to the second reason why a disjunctive reduction is liable to fail; and that is that the best candidates with which a set of physicalistic disjuncts can be unified or completed with some degree of certainty are, at least at the present time, mental in character. Even if it is true, it is not quite sufficient to say (for example) that x's arm will extend towards some biscuit in his vicinity, for he might mistake the biscuit for a baloon or a bagel - or he might mistake a baloon or a bagel for a biscuit. It therefore has to be inserted that x's arm will extend towards something in his vicinity which he judges, perceives or believes to be a biscuit. Without a physicalistic reduction of judging, perceiving or believing, there seems no way of avoiding at least one of these psychological concepts in any attempt to analyse reductively what wanting a biscuit is.

Mental predicates seem to have no corresponding physical predicates

which are even co-extensive with them, let alone physical predicates which are co-extensive with them as a matter of law. So far I have only sought to illustrate this truth - a truth which Brentano was the first to suspect - and not until the end of Chapter IV will I try and say anything to explain or provide a context for it.

Carnap's own explorations of the "explicit definition" approach to psychological terms did not take him quite as far as to make him speculate a disjunctive reduction. The obvious falsehood of the most primitive approaches (e.g. those like (1) and (2)) led him to adopt a method of what he referred to as reduction sentences. His initial view was that if S_c was some sentence representing the physical conditions in which a certain organism found itself, and if S_m was some psychological sentence concerning that organism, then there was some sentence S_r , describing non-psychologically the response of the organism, which stood to the other two sentences in the following relation:

$$(4) \quad N (S_m \equiv (S_c \supset S_r))$$

of which one instance might be:

$$(5) \quad N (\text{Fred wants a biscuit} \equiv (\text{If there is a biscuit present, then Fred's arm will extend towards it}))$$

But the elementary truth-functional laws of ' \supset ' tell us that S_m would be true if S_c were false; or that, in the specific case, Fred wants a biscuit if there is not a biscuit present - an absurd result. This well-known and decisive objection leaves nothing further to be said about reduction-sentences formalised in this way.

Carnap later made an attempt to repair the defect in this method by

introducing schema to partially interpret psychological terms. This was an accent which consisted in a method of "bi-lateral reduction". Using again the terms S_c , S_m , and S_r , his new view was expressed as saying that:

$$(6) \quad N(S_c \supset (S_m \equiv S_r))$$

Sentences of this sort only provided partial interpretations for the mental sentence S_m , since even if a particular instance of (6) were true, there would remain the possibility of other conditions implying an equivalence between S_m and a set of response-sentences. An instance of (6) might be

$$(7) \quad N(P \text{ is asked how he feels} \supset (P \text{ feels terrible} \equiv P \text{ says "I feel terrible"}))$$

which specifies a necessary and sufficient condition of P's feeling terrible, but only in the circumstances that P is asked how he feels. In other circumstances the tests fail to apply.

The reduction-sentence (7) is equivalent to the conjunction of

$$(8) \quad N(P \text{ feels terrible} \supset (P \text{ is asked how he feels} \supset P \text{ says "I feel terrible"}))$$

$$(9) \quad N((P \text{ is asked how he feels and } P \text{ says "I feel terrible"}) \supset P \text{ feels terrible})$$

But this shows even more clearly that the example as a whole would not do as an attempt to give nomologically necessary and sufficient conditions for the truth of P feels terrible in non-psychological terms, since in order to make statements (8) and (9) true, it seems to be required that some explanation would need to be incorporated about P's understanding of

the question, and about his meaning a certain thing by saying "I feel terrible". And even if this additional information were built in to the example, the new assertion would most likely fail to record general truths about P - for, at least if he is like the rest of us, he would occasionally choose not to answer the question, and still feel terrible. But the example illustrates the method Carnap evolved for the partial interpretation of psychological sentences in non-psychological terms. As a method, clearly, it is not adequate to provide necessary and sufficient general and law-like conditions for the truth of psychological sentences, and indeed it was never supposed to do so. The reduction involved is a weak one; but its details give us additional evidence of the difficulty of nomologically connecting the psychological with the non-psychological.

In the case of both the naive but strong reductive theories, and the more logically sophisticated but weaker ones too, the failure seems to consist in the fact that, at the very least, some psychological sentence appears to be needed in the analysis of any given psychological sentence. This hypothesis of the behaviourally irreducible nature of mental language suggests itself at two levels: both at the level where we try to give a non-psychological analysis of a singular psychological sentence (e.g. P feels terrible) and also at the level where we try to give a non-psychological analysis of a general psychological sentence (e.g. (for any x) x feels terrible) - the task in the former case being to arrive at a general truth about P, and the task in the latter case being to arrive at a general truth about feeling terrible. But whether Brentano's irreducibility hypothesis is one which can be proved, or at least in some sense supported by a rigorous deduction is, to date, an open question.

It is also a question of the first importance to this branch of the philosophy of mind. Many philosophers have taken the fact that mental language is irreducible to behavioural language (in the sense explained)

to suggest that mental language is irreducible to any kind of non-mental language, and they have concluded from this that mental properties cannot be reduced in any way to non-mental properties. One of the themes of this essay will be to examine whether this line of reasoning is correct. We have seen in this section that the evidence for the irreducibility of mental language to behavioural language is extremely compelling, but the question of whether mental language is irreducible to cerebral language is, I shall eventually argue, a quite different and independent question. What I shall be challenging, in other words, is whether the systematic failure of any behavioural reduction of the mental is sufficient to support the claim that a cerebral reduction is impossible as well.

In the next section I present some physicalistic hypotheses of a sort which do not rely upon a reduction of the mental to the behavioural.

4. ALTERNATIVES TO A BEHAVIOURAL REDUCTION

One of the initial hopes of the reductive enterprise was to be able to so establish a connection between terms or sets of terms that the reduced set could effectively replace the reducing set, in any significant area of speech or theory. This connection, I explained, would have to be nomological: the mental sentence would have to be true whenever and only whenever the behavioural sentence was true, and not just in the extensional sense of the word "whenever". That nomological connectedness in a bridge law must be the standard for the reduction of one property to another can be easily appreciated if we survey the alternatives. Meaning-equivalence, or synonymy, between a mental predicate and a behavioural predicate must, on any reasonable interpretation, be too strict a standard for a reduction of one property to another, since not only are the predicates of our paradigm reduction ("temperature" and "mean molecular

kinetic energy") not synonymous, but synonymous predicates are not, in general, names of inter-reducible properties.

And while synonymy is too strict a standard, mere coextensiveness is too weak. Quine writes that:

"we have satisfactorily reduced one predicate to others if in terms of these others we have fashioned an open sentence that is co-extensive with the predicate in question i.e. that is satisfied by the same values of the variables" 25

But if this were correct, it would follow that the property of thinking about food would be reducible to the property of having saliva in the mouth, on the single condition that all and only those people who thought about food had saliva in the mouth; which would be in implausible result.²⁶

Reductive theories of the kind considered in the last section were "peripheralist" theories, in the sense that the things to which mental properties were to be reduced were behavioural properties - described, of course, in physical terms. Now as an alternative to theories of behavioural reduction there grew a suggestion that mental phenomena ought to be connected not with behaviour or physical motions but, in some way, with neurophysiological features of the organisms brain or nervous system. There is therefore the possibility of a "centralist" or (what I have called a) "cerebral" reduction, which, if actually carried out, would have the effect of establishing an identity of mental properties with cerebral properties, via bridge laws of the form $N(x)(x \text{ is } M \equiv x \text{ is } P)$, where this time "P" is not a behavioural predicate but a predicate of some physical language describing a property of the brain or central nervous system. Or, alternatively, at a less general level, it would consist of establishing

25. W. V. Quine, Ontological Reduction and the World of Numbers, p. 188. (In The Ways of Paradox and other essays. New York 1966)

26. See again footnote 15, above, and its attendant text.

bridge laws about particular individuals of the form $N(t)(A \text{ is } M \text{ at } t \equiv A \text{ is } P \text{ at } t)$, where the generality is only a generality with respect to different moments in the life of the same individual.

The general ideology of this reductive programme is much the same as that of the behaviourist programme: to demonstrate the ultimate dispensibility of mental language. Moreover the idea that mental properties and cerebral properties are one and the same has a certain amount of intuitive appeal. We know²⁷ that many of the phenomena of inanimate nature have no "analysis" in terms of outward visible features, whereas a suspicion exists that objects belonging to the same kind have a common internal microscopic structure. While objects subsumed under the concepts lemon or cat, for instance, have no jointly necessary-and-sufficient external features, it is more plausible to think that they have common structural features at the microscopic or genetic level. Advocates of the programme of cerebral reduction might ask what reason there is to suppose that mental phenomena might be different in this respect.

The reduction of the mental to the cerebral, either in the form of a reduction of mental properties to cerebral properties or in the form of nomological truths about particular individuals taken one by one, is a matter which I shall not try to treat fully until Chapter IV (section 2), where the crucial question will be whether the kinds of factors which tended to falsify all efforts at a behavioural reduction tend to also falsify any cerebral reduction. It is important at this point to explain the methodology of the other approach to physicalism, namely the non-reductive approach.

The essence of a non-reductive physicalistic hypothesis is that it

27 Thanks to Putnam's article Is Semantics Possible? *Metaphilosophy*, 1970.

suggests the identity of particular mental phenomena with particular physical phenomena, without any implication of generality at all. That is to say, it would be a non-reductive hypothesis that

Tom's being in pain at noon on Friday = Tom's being ϕ at noon on Friday

or that

Tom's remembering that p at noon on Friday = Tom's being ψ at noon on Friday

where " ϕ " and " ψ " are cerebral predicates, and where there is no suggestion that if Tom were to be in pain at noon on Saturday, then he would be ϕ at that time; or that if Fred were to remember that p at noon on Friday then he would be ψ at that time. The reason why a hypothesis of this very particular and non-generalisable kind is non-reductive is clear from the previous explanation of what a reduction itself consists of. No bridge law connecting the property of being in pain with the property of being ϕ would be necessary (although it would be sufficient (see Chapter IV, section 2)) to establish such a particular identity. In a non-reductive theory, mental "particulars", specified by a phrase of the form A's M-ing at T, are asserted to be identical one by one with physical "particulars", specified by a phrase of the form A's P-ing at T; but on grounds quite different in kind from those which motivate an identification of a mental property with a physical property.²³ Another feature of the

23. Such differences of level are sometimes expounded in terms of the difference between types and tokens. It is frequently said that the identity theorist has just two choices: either to identify a type of mental phenomenon with a type of physical phenomenon, or else to identify particular token mental phenomena with particular token physical phenomena, independently of what type they belong to. E.g. Fodor and Block use this terminology in What Psychological States are Not.

difference is that while a reductive theory is only properly advanced as a scientific hypothesis, to be confirmed or falsified by the facts, a non-reductive theory cannot be scientific in quite the same sense. It is for this reason, perhaps, that a non-reductive theory may be unmisleadingly called a philosophical or metaphysical theory.

The label "identity theory" is therefore not by itself sufficient to distinguish the reductive from the non-reductive; an identity theorist can suggest anything ranging from the identity of a mental with a physical property, at the most general and ambitious level, down to the identity of mental with physical "particulars", at the least general level. Only an identity theory at this least general level is non-reductive.

The credentials of identity theory (both reductive and non-reductive) will be a major theme of this essay. But before bringing this section to a close, I should like to mention a type of reduction outlined by Quine, which is called by him "ontological reduction", but which the distinction between the reductive and the non-reductive as I have just explained it would fail to categorise clearly either one way or the other. The concept which is perhaps fundamental to ontological reduction in Quine's sense is the concept of explication. Roughly speaking, explication consists of locating, for any object which needs analysis, or whose ontological credentials are unclear, some other object with which it can, for various purposes, be identified. But two items connected in this way need not be identical in the strict Leibnizian sense; and so the logical strictures imposed by the law of that name turn out to be simply irrelevant to whether an explication is successful or acceptable.

A simple example of explication would consist of considering a three-penny piece or a fly-button to be identical with a missing rook piece in a chess game, where neither object shares all the properties of the other and, clearly, where the chief motive for assimilating the two is to provide

a functional equivalent for a lost item. The apparatus of explication is more complicated than this simple example suggests, of course, and together with various elaborations and supplementations which Quine provides, is supposed to enable us to "eliminate" the entities of one theory in favour of the entities of another - thus ontologically reducing the one theory to the other. (Frege's analysis of numbers in terms of sets is supposed to be a paradigm case of ontological reduction.) The programme of ontological reduction is one which is clearly worthy of consideration in the general business of analysing mental phenomena, for its suggestion is that mental entities, if such there are, could be eliminated in favour of physical entities. I shall therefore consider it eventually. But since ontological reduction is a topic which is only tangentially connected in method to the main themes of this essay I shall confine my discussion of it, together with my sceptical conclusions, to an appendix.

5. PROGRAMMATIC REMARKS

Having now identified the main termini of my investigation, I must say something of what will lie in between them and the present point in the essay. I think it needs no argument that a really adequate physicalistic theory would have to take into account the various rather subtle differences which can be found within the entire corpus of what, up to this point, we have uncritically referred to as mental phenomena. Some systematic exposition would be needed, firstly, of what sorts or natural categories of mental phenomena there are; and secondly, of what differences the items belonging to these categories exhibit in respect of causal and explanatory connections with other phenomena, the conditions of their successful individuation, and so forth.

Now not all of these matters can be fully explored in an essay of

this length. For example, I shall leave largely undiscussed the question of whether we ought to speak of mental activities, mental processes, or mental features, etc.; and concentrate instead upon the two categories whose existence is generally thought to be less controversial, viz. the category of mental states and the category of mental events. It is the physicalistic analysis of items belonging to these categories which I shall devote most of my energy to; hoping, of course, that some of the details will be transferable to items of other categories, if and when they emerge at all clearly.

To begin with I shall deal with a certain non-reductive, physicalistic account of mental states, which has been put forward in recent years as an alternative to that version of the identity theory which asserts that for each type of mental state there is a type of physical state which is identical with it. The theory of mental states which I have in mind entails, as a conclusion, that a mental state can be identified with (in the explicative sense (see I, 4)) a class of physical states whose members are identified by certain functional tests. So there is the actual theory of mental states on the one hand, and there is a non-reductive physicalistic conclusion which follows from it on the other. My aim in the next chapter will be to conduct a critique of the theory itself, leaving an assessment of the conclusion until the appendix. In this sense I shall be first examining the route along which the theory travels, rather than its physicalistic terminus.

In order to understand exactly what is asserted and what is denied by this theory, and indeed in order to get the whole subject of physicalism into clear perspective, there is one more preliminary job which needs to be done, which is to lay out in detail the different levels of generality on which any physicalistic conclusion might be asserted.

Let us survey the question of generality for the case of mental states.

On the face of it, we can distinguish logically between three different levels of generality. On the first and least ambitious level, there is the theory that each particular mental state of a particular person at a particular time - i.e. each instance of a particular mental state of a particular person - is identical in the strict sense with a corresponding instance of a particular physical state of that person. Let us assume (what I shall later expose to criticism) that an instance of a particular state of a particular person is correctly nominated by a phrase of the form "P's being at T", where "P" names a person and where "T" names a time, and where the blank "...." is appropriately filled. Then this first theory asserts sentences of the following sort

Time-Specific Level

Tom's being in pain at T_1 = Tom's _____ at T_1

Tom's being in pain at T_2 = Tom's _____ at T_2

Fred's being in pain at T_1 = Fred's _____ at T_1

Fred's being in pain at T_2 = Fred's _____ at T_2

.
.
.

where none of the blanks in the right-hand side need be filled in the same way, although they must each be filled with a non-mental term. On the next most ambitious level, there is the theory that each instance of each particular mental state of a particular person is, at any time, identical in the strict sense with an instance of some particular physical state.

This second theory asserts sentences of the following sort

Person-Specific Level

(t)(Tom's being in pain at t = Tom's _____ at t)

(t)(Fred's being in pain at t = Fred's _____ at t)

.
.
.

where each one of these sentences generates, as instances, sentences like

Tom's being in pain at T_1 = Tom's _____ at T_1

Tom's being in pain at T_2 = Tom's _____ at T_2

Fred's being in pain at T_1 = Fred'sat T_1

Fred's being in pain at T_2 = Fred'sat T_2

where "_____" always contains the same non-mental filling, and where "....." always contains the same non-mental filling. Or equivalently, you could leave out the time specification altogether, and say, for instance in Tom's case,

Tom's being in pain = Tom's _____.

On the third and most ambitious level of all, there is the theory that without regard to time or person, for each instance of a particular mental state there is a particular physical state, an instance of which it is identical with. On this third level the theory asserts sentences like

Property Level

(x)(t)(x's being in pain at t = x's _____ at t)

which of course generates, as instances, particular sentences like

Tom's being in pain at T_1 = Tom's _____ at T_1

Tom's being in pain at T_2 = Tom's _____ at T_2

Fred's being in pain at T_1 = Fred's _____ at T_1

Fred's being in pain at T_2 = Fred's _____ at T_2

.
.
.

where "_____" is always to be given the same non-mental filling. This third theory can be represented as saying things like

Being in pain = _____.

although, as I shall show at a later point, the exact connections between such a statement of property-identity and any more particular statement

of the kind just listed (above) are complex.

The differences between these theories are differences of generality: the first contains no generality whatever, the second generalises with respect to time, and the third generalises with respect to time and individual. A fourth theory which generalised with respect to time, individual and psychological state, would be clearly over-general, and for that reason unworthy of serious consideration. I have chosen to display these matters explicitly because I want to explain what the theory of mental states which I shall be concerned with in the next chapter asserts and what it denies. That theory asserts the falsehood of an identity theory at the Person-specific level or the Property level, but has as its conclusion a series of sentences, which although superficially of the Property level form:

being in pain = _____

are in fact of a sort whose blank is not filled by a physical description. The most this theory can assert by way of a mental/physical identity-sentence is something explicative like

Being in pain is to be identified with the class whose members
are: Tom's being _____ at T_1 , Tom's being _____ at T_2 ,
Fred's being _____ at T_1 , Fred's being _____ at T_2 ,
Fred's being at T_1 , etc.

where the blanks can each be differently filled. So although the theory asserts the falsehood of any non set-theoretic identity at the Person specific level or the Property level (and on empirical grounds), it ambitiously seeks to find a different unitary analysis of the general state of being in pain.

The plan for the succeeding two chapters is as follows. In Chapter III, I shall address myself to the altogether more basic series of ontological questions concerning the existence of items belonging to the

so-called categories of mental events and mental states; and to the interpretative question of what kinds of thing phrases of the form "A's being G", "being G", "A's being G at T" are in point of fact appropriate to nominate. And in the final chapter I shall return to discuss, in the main, the possibilities left for a centralist reduction of the mental, as an alternative to both (i) the behavioural kind of reduction discussed in this chapter, and (ii) the standard non-reductive "identity theories" which identify particular mental events with particular physical events or particular mental states with particular physical states. Broadly speaking, my conclusion will be that a certain kind of centralist reduction is the single open possibility there is for physicalism, and that this is a possibility only because Brentano's hypothesis of the irreducibility of mental language is not exceptionlessly true.²⁹

29. While writing this essay I became aware of the conclusions worked out by D. C. Dennett in his book Content and Consciousness (London and New York, 1969), some of the more heterodox of which, although argued on different grounds, coincide with those of the present essay. I signpost these points of coincidence in Chapters III and IV.

Chapter II. Two Theories Examined

IIA. The "Functional" Theory of Mental States

1. Introduction
2. Exposition of the Turing Machine theory
3. Criticism of the Turing Machine theory
4. Relevance of the theory for physicalism

IIB. A Theory of Mental Events

1. Exposition of Davidson's theory
2. Causal interaction between the mental and the physical:
reasons as causes
3. Causal interaction between the mental and the physical:
mental states as causes.

IIA. The "Functional" Theory of Mental States

1. INTRODUCTION

In this chapter I intend to expose the deficiencies of two theories about the nature of the mind. These theories have not been arbitrarily selected: they have both been quite widely endorsed in recent philosophy, and they are both designed, by their exponents, to support certain physicalistic conclusions. The first theory I shall consider has come to be known as the functional state theory, or occasionally as the functional theory of mental states. According to the most general form of this theory, mental states are those states of the organism which have some function in the genesis of behaviour, and indeed are those states of the organism which are, in a sense I shall seek to explain, identified by what their functional relationships are. Now the interest of this theory for the doctrine of physicalism is considerable: for it not only suggests a way of distinguishing the mental from the physical, but it also contains important implications for both reductive and non-reductive forms of the physicalist doctrine.

But the theory has been held with more or less conviction by a number of people, and because of their varying convictions, in a number of forms. Stuart Hampshire, without actually using the word "function", once suggested¹ that feelings are characterised by the fact that they give rise to certain inclinations, e.g. pain (of the type occasioned by bodily damage) is characterised by the inclination it gives rise to, to withdraw the damaged part. Chomsky has spoken about mental concepts in a similar vein in the context of the theory of grammar:

1. Hampshire: Feeling and Expression. London 1961.

"The mentalist need make no assumptions about the possible physiological basis for the mental reality he studies. In particular, he need not deny that there is such a basis. One would guess, rather, that it is the mentalistic studies that will ultimately be of greatest value for the investigation of neurophysiological mechanisms, since they alone are concerned with determining abstractly the properties that such mechanisms must exhibit and the functions they perform" 2

And Gilbert Harman writes that

".... psychological states and processes are functionally defined" 3

"Just as states of an automaton are to be defined in terms of functional relationships rather than in terms of the exact nature of their physical realisation, so too for psychological states (the same psychological states may be differently realised in different people, or even in the same person at different times). These states are defined in terms of their functional relationships to other psychological states, as in reasoning; to input, as in responses to observation; and to output, as in intentional action" 4

However the views expressed in these quotations can hardly be regarded as specific enough for a philosophical critique to be based upon them.

They do not embody any indication as to how the difficult word "functional" and the difficult phrase "functional relationship" should be understood; and they do not make totally clear the relationship which any mental state bears to a physiological or physical state. In these respects they are only slightly more specific than Armstrong's general formula:

"The concept of a mental state is primarily the concept of a state of the person apt for bringing about a certain sort of behaviour" 5

-
2. Chomsky: Aspects of the Theory of Syntax. p. 193.
 3. Harman: Knowledge, Reasons and Causes. *Journal of Philosophy* 1970, p. 851.
 4. Harman, loc cit. p. 849.
 5. D. M. Armstrong: A Materialist theory of the Mind p. 82.

Fortunately, there exist more exact formulations of the theory that mental states are functional states, the most specific of which is Putnam's view that the way in which mental states are organised can be compared to the way the Machine Table of a Turing Machine codifies the functional states of such a machine. It is this formulation of the functional state theory on which I shall therefore concentrate my attention. But there is another relatively specific formulation of the functional state theory which has been presented by J. A. Fodor, and whose shortcomings I think it is instructive to see, before progressing to the details of Putnam's view.

Fodor's view is, it seems, that when giving a psychological description of a state of a person, we thereby say (or, as I shall put it, make manifest to our audience) what its specific function is in relation either to other psychological states, or to the behaviour of the organism concerned. At least I think that this is a reasonable interpretation of many of the things Fodor says about psychological states. For according to Fodor, part of our theory of a persons behaviour consists in giving descriptions of

"The internal states of organisms in respect of the way they function in the production of behaviour The properties of these states are determined by appeal to the assumption that they have whatever features are required to account for the organism's behavioural repertoire"⁶

What this aspect of our theory tells us about these internal states, according to Fodor, is

"what role they play in the production of behaviour ... Theory construction proceeds in terms of such functionally characterised notions as memories, motives, needs, drives, desires, strategies, beliefs, etc. with no reference to the physiological structures which may, in some sense, correspond to these concepts If I say "he left abruptly upon remembering a prior engagement" I am giving an explanation in terms of an internal event postulated in order to account for behaviour"⁷

6. Fodor: Explanations in Psychology p. 173

7. Fodor: Explanations in Psychology p. 172. (My underlining)

But this theory of Fodor's, interpreted to mean that a psychological description says or makes manifest what specific function the state in question has, can be seen to make the functional state theory immediately false. For when we describe a person's state (either his state or the state of his mind) as the state of believing that rain is imminent, say, we do not, in describing him in those terms, say (or even by implication make manifest to our audience) what the specific function of that state is in relation to further states of mind or to his actions. It may be a fact about that person that subsequent to believing that rain is imminent he remembers leaving his umbrella behind on the train - and it may also be a fact that we can explain his remembering leaving his umbrella behind by saying that he believed rain was imminent. (He would not have remembered about his umbrella if he had not acquired the belief about the rain). But this does not mean, or require, that in describing his psychological state as one of belief that rain is coming we explicitly (or even implicitly) relate it either causally or as a possible explanation to any subsequent psychological state. And the same goes for the relation between a state psychologically described and some subsequent action. Suppose the man suddenly runs for shelter - and suppose that, in explaining this action, we refer to his state of belief that rain is on the way. Then, even though that may be a correct explanation in that situation, it does not mean that in describing his state as a certain kind of belief-state we thereby say what its specific functional role is in relation to some action he later performs. The words "he believes that rain is on the way" may plausibly be said to describe a state; but the words by themselves do not tell us what specific subsequent psychological states or subsequent actions of his can be explained by them.

The correctness of my interpretation of Fodor's functional state theory, and the falsehood of the theory under that interpretation, is

also clear from an analogy which Fodor gives. According to this analogy, a psychological description stands in the same relation to a brain state as the term "valve-lifter" stands to a camshaft.⁸ In both cases, he thinks, the description gives the specific function of a structural mechanism or part, and it is called a functional description because the description itself informs us as to what specific function the part performs. The valve-lifter is specifically: that thing which lifts valves; in the same way, a psychological description is supposed not merely to describe some state of the organism or its brain, but to describe it functionally. But it is clear that the analogy does not hold; since, psychological terms in themselves simply do not describe the specific functions of brain-states in the way the term "valve-lifter" admittedly does a camshaft. Nor can we patch up the analogy by construing a psychological description as somehow having the same meaning as a phrase like "action-causer" or "state appropriate for the production of behaviour". We cannot construe psychological descriptions in this way for the simple reason that individual psychological states do not invariably cause actions; there are plenty of mental states which can be perfectly "idle", and which need have no obvious bearing on a person's behaviour at all. A man may be indulging in a fantasy or a daydream about how he might have behaved differently on some occasion which is now past. He might be running through the details of an amusing story in his head; he might be basking in the memory of some former public glory of his; and yet none of these facts need show themselves in what he subsequently does. Psychological information of this sort can be valued for its own sake. If, in Fodor's example of "he left abruptly upon remembering a former engagement", the mental event mentioned is one which is "postulated in order to account for behaviour", then of course, so specified, it must account for the

8. Fodor: Explanations in Psychology pp. 177-9.

behaviour. But it seems implausible to say that all mental states owe their existence to such a postulation.

It is not even true that mental states are all and only those which are suitable for explaining their owner's behaviour; or, to mimic Armstrong's words, that mental states are all and only those which, on suitably chosen occasions, are apt for the explanation of their owner's behaviour. Although it is quite true in a general way that actions very often can be explained by reference to previous or concurrent states of mind, it is not by any means true that this is the only way to explain actions, or even the best way. Actions can be explained by reference to the state of mind of a person other than the agent, as when it is said that a certain school-boy did what he was told because his teacher wanted him to; or an action can be explained without reference to a state of mind at all, as when it is said that a dinner guest ate with just his fork because etiquette required it; and there are other kinds of cases, in which the explanation for what someone did does not lie in the agent's mind, but elsewhere. This is not to deny that in a case where an action needs to be explained, there usually is some state of mind of the agent which is relevant, for I think there usually is; it is to deny that when we explain actions, we always or necessarily do so by citing a mental state of the agent.⁹ But if actions do not necessarily need to be explained in terms of a mental state of the agent, and if it is also true that an agent's mental state need not have any causal bearing on his behaviour, then it cannot be true that mental states are states of a sort which are apt for bringing about, or explaining, behaviour.

9. And an action can be explained by the absence of a desire, as in "I put my umbrella up because I did not want to get wet". This is perfectly explanatory, but different from "I put my umbrella up because I wanted not to get wet". The first implies the counterfactual "If I had wanted to get wet I should not have put up my umbrella"; but the second does not.

In as much as mental states are not functional in the sense of that word used to interpret Fodor's theory, then mental states cannot be distinguished from physical states by their functionality (conceived in that sense). And in as much as mental states seem not to be especially apt for bringing about behaviour, in the sense of that phrase employed in my interpretation of Armstrong's view, then mental states cannot be distinguished from physical states by using that concept either. Clearly, an adequate distinction between the mental and the physical must begin with a characterisation of the mental which is itself adequate.

If mental states are said to be functional in the weaker sense of providing necessary conditions for the truth of a statement about behaviour, then although this is more plausible, it fails to provide a strong enough test with which to distinguish mental from physical states. Having a pair of legs is a state without which it is impossible to go for a walk; having a brain in one's skull is a state without which it is impossible to work out a problem or write a letter; and yet neither of these states can plausibly be described as mental states. It remains to be seen whether there is any other sense which can be assigned to the word "functional" which makes it capable of adequately characterising just those states which are mental. Only if this is so will it be possible to draw the mental/physical distinction in those terms.

2. EXPOSITION OF THE TURING MACHINE THEORY

I now move on, as promised, to examine the detailed opinions of H. Putnam on the nature of mental states and their organisation. The quotations from Chomsky, Harman and Fodor which were given in the previous section all contained the idea that the nature of a psychological state is to be explained independently of how it comes to be "realised" in

matter; although Chomsky and Fodor thought it possible, none the less, that a mental description might have the role of describing the function of a physiological mechanism. These two intuitions, that there is no physical realisation that is necessary to any particular mental state, and that mental states are in some sense "functional" states, both have a place at the centre of Putnam's theory.

In considering pain, Putnam writes:

"... pain is not a brain state, in the sense of a physical-chemical state of the brain (or even of the whole nervous system), but another kind of state entirely. I propose the hypothesis that pain, or the state of being in pain, is a functional state of the whole organism" 10

This hypothesis Putnam calls the "functional state hypothesis".¹¹ Many of the states which Putnam describes as psychological are those which others would be inclined to think of as being physical: the state of pain, the state of hunger, the states of thirst or aggression.¹² Apart from the specific case of pain, there are signs¹³ that he considers all the other more obviously cognitive states to be functional states as well: states like being able to recite the Ancient Mariner or knowing the chemical composition of sugar, and others of the same degree of complexity. The functional state hypothesis, then, can be taken as the hypothesis that psychological states in the broadest sense are functional states of the organism.

Putnam explains that a large part of the motive for adopting this theory derives from the clear empirical falsehood of that version of the "identity theory" which identifies every type of mental state with a particular type of physico-chemical state of the brain. Such a theory

10. Putnam: Psychological Predicates, Art, Mind and Religion (ed. Capitan and Merrill) p. 41.

11. Putnam, op cit. p. 44.

12. See Quine for example: Word and Object. § 54.

13. Putnam: Psychological Predicates pp. 43, 45. The Mental Life of Some Machines p. 211.

would be one which belonged to the Property level, in the sense I explained earlier (Chapter I, section 5). Putnam regards such an identity theory as implausible, partly because a person who is in pain on two successive occasions can have very different brain states (described in certain physico-chemical terms, at least), and also because two animals can both be in pain (the same state) even if they belong to different species, and have, as a matter of fact, quite dissimilar brain structures. Putnam argues that, in order for the identity theorist to be right, the physical-chemical state with which pain is to be identified

"must be a possible state of a mamallian brain, a reptilian brain, a mollusc's brain (octopuses are mollusca, and certainly feel pain) etc. At the same time, it must not be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain" 14

Since this is implausible, he reasoned, a method of collecting brain-states must be arrived at which does not make small differences in physical-chemical structure matter.

The functional state theory replaces the view that a psychological state is a physico-chemical state, and says that the psychological states of two creatures (or two successive states of the same creature) are the same providing that they have identical functional relationships to other states, to "input", and to the organisms behavioural repertoire. To put it summarily, the theory says that psychological states are the same which have the same function.

But we are not yet in a position to critically assess this theory, for we still need to know in more exact detail what is meant, in the context of the theory, by the phrase "functional relationship". Putnam provides these details by comparing a body of psychological information about a person with the information represented on a machine table for a

14. Putnam: Psychological Predicates p. 44.

Turing Machine. The concept of a Turing Machine, and the concept of a machine table for such a machine are exact and precise concepts; and so the way to understand Putnam's view on the nature of the mind is to understand what a Turing Machine is, and how its machine table describes its operations.

A Turing Machine is

"a device with a finite number of internal configurations, each of which involves the machines being in one of a finite number of states, and the machines scanning a tape on which certain symbols appear. The machine's tape is divided into separate squares on each of which a symbol (from a finite fixed alphabet) may be printed" 15

Such a machine is defined by its machine table, which essentially specifies how the machine will change its state when a certain "input" is received, and how it will discharge a certain "output" when a change of state comes about. So whether or not any actual physical mechanism can be described as a Turing Machine depends upon whether or not its operations can be described by a machine table.

Since, as I shall illustrate in a moment, a machine table can be either probabalistic or deterministic, it is possible for an actual physical device to be described as a probabalistic Turing Machine and as a deterministic Turing Machine at the same time. A machine is a deterministic Turing Machine if each state specified by the machine table has only one state named as its successor, and only one as its predecessor. It is probabalistic if each state specified by the machine table has several states listed as its possible predecessors, and several listed as its possible successors - the exact probabilities of transition between that state and its several possible predecessors and successors also being given in the machine table. This shows clearly why it is that if a physical system or type of physical system is a probabalistic machine, it can be

15. Putnam: Minds and Machines, p. 140.

a deterministic machine too. To call it a probabilistic machine is to say that its machine table records the probabilities of transition between various states; which is compatible with there being some different machine table which only mentions probabilities 0 and 1 - that is, one which gives a deterministic description. Conversely, if something is a deterministic machine, it can also be a probabilistic machine, for similar reasons.

To call a device a deterministic Turing Machine, or to call it a probabilistic Turing Machine, is just to say what kind of machine table it has. These notions can now be illustrated as follows. The simplest machine worth describing¹⁶ is one which has two states, A and B, which is deterministic, and which only has two possible inputs: only two letters can appear on its input tape. Then the machine table might look like this:

	<u>state A</u>	<u>state B</u>
input 1	B2	A2
input 2	A1	B1

(Figure 1)

where the instruction "B2", for example, means "change to state B and print the symbol "2" in place of the symbol now in front of you; and finally, scan the next space on the input tape". The instruction to scan the next space of the input tape is naturally part of every instruction - but in every other respect the instruction is specific to the state which the machine is in and to the input which it is receiving. Now in the case of this simple, deterministic, two-state machine, the "functional" relations are strictly causal - for they are simply the determining relations which the machine table specifies for each state. The causal

16. For a description of a four-state machine see Putnam: Minds and Machines in Hook (ed.) Dimensions of Mind.

relations for state A, for example, are two in number: state A determines the printing of symbol "2" and a change to state B when the input is input 1; and it determines the printing of symbol "1" and no change of state when the input is input 2. To say that the states of such a machine (states A and B) are identified functionally would be to say that the identity of each state is fixed by the machine table column. If two separately named states have identical machine table columns under them, then it would follow that those separately named states are one and the same.¹⁷

The case of a probabalistic machine is in principle just the same: but a causal or determining relation here is not a causally deterministic

-
17. This is an example of what Quine calls the "identification of indiscernibles". See From a Logical Point of View pp. 70-3; Word and Object p. 230. It might be useful to have an example of this procedure at work, in order to see the interplay between the identity of a state and its machine table column. Suppose the finest individuation of states which the theorist can adopt, independently of functional considerations, yields states a, b, c, and d. Suppose the machine table for these is

	<u>a</u>	<u>b</u>	<u>c</u>	<u>d</u>
i1.	?b	?a	?b	?b
i2.	?a	?c	?a	?c
i3.	?c	?b	?c	?a
i4.	?d	?d	?d	?d

(where the use of "?" summarises the fact that every instruction in any one row contains the same input letter, and each state has the same probability - it does not matter what they are for the purposes of the example). Then since the column under a coincides with the column under c, state a = state c, and the a column can be eradicated from the machine table. But if a = c, then each "c" in every remaining column has to be re-written as "a", which yields another identity, viz that of the column under a with the column under d. This leaves us with

	<u>a</u>	<u>b</u>
i1.	?b	?a
i2.	?a	?a
i3.	?a	?b
i4.	?a	?a

one - only a probabalistic one. A machine table for a three-state, probabalistic, three-input machine might look in part like this:

	state A	state B	state C
input 1	$B2\frac{1}{2}$ $A1\frac{1}{2}$
input 2	...	$B2\frac{1}{6}$ $C3\frac{1}{3}$ $A1\frac{1}{2}$...
input 3	$C1\frac{1}{8}$ $A3\frac{1}{8}$ $C2\frac{1}{2}$ $B2\frac{1}{4}$

(Figure 2)

where the instruction corresponding to any input and any state specifies a set of probable changes to a range of other states. When the input is input 2 and the machine is in state B, for instance, the machine table says that the probability of printing "2" and not changing state is $1/6$, the probability of printing "3" and changing to state C is $1/3$, and the probability of printing "1" and changing to state A is $1/2$.

With this general description of a Turing Machine before us, we can now see clearly what it means to say that a psychological state is a functional state. For according to Putnam, a person is a Turing machine, in the sense that the organisation of his mental states can be set out on a machine table.¹⁸ A person's psychological states are functional states, according to this view, in just the sense in which states A, B or C of

18. A machine tape is an essential part of any Turing Machine. It is a tape on which symbols from a finite alphabet appear, and which the machine can "scan", erase from, or print onto. A Turing Machine is essentially an effective computing device: but to make it realistic as a model for persons capable of perceiving and acting, certain additional specifications have to be made. Putnam adds that the machine has to be thought of as equipped with a sensory system that scans the machines environment and which prints symbols into the machines tape, and also with motor organs which are such that when the machine itself prints certain symbols onto the tape, the motor organs execute certain "actions". See The Mental Life of Some Machines pp. 178-9.

the previously described machines are functional states: it is an essential feature of them that they are related to other states in various ways, and their identities as states are fixed precisely by what these ways of relating are. Two psychological state-descriptions describe the same psychological state if and only if their respective machine table columns coincide: this is the sense in which psychological states are the same which have the same function.

Now it is important to appreciate that is included in this theory that a person is a Turing machine, and what is denied by it. In particular, it is important not to underestimate it. What I have in mind here is the fact that the characterisation of a Turing Machine which Putnam gives, and in which I have followed him, allows that every object can be trivially described as a Turing Machine by thinking of it as having only one state. A stone is a Turing Machine if its sole state is thought of as the state of being a stone; and a brick is a Turing Machine if its sole state is thought of as the state of being a brick; and so forth. And only slightly less trivially, any human being can easily be described as a Turing Machine if he or she is thought of as having only two states: the state of being conscious and the state of being unconscious. A machine table for a person thought of in this way could be easily written, for it would simply contain a specification of the transition-probabilities between those two states in the face of various inputs; and in this case the catalogue of relevant inputs would be relatively short, since there are few types of stimuli which are catastrophic enough to cause a change from consciousness to unconsciousness, or from unconsciousness to consciousness.

This trivialising thought-experiment is to be resisted if the theory is to be taken seriously. In its serious form, it says not simply that a person is a Turing Machine, but that a person is a Turing Machine of a sort whose machine table lists all and only the person's psychological states and their mutual functional

connections. This is the first major claim that the functional state theory makes. The other is that each psychological state owes its identity to what its functional connections are. This is to say that it is essential to the identity of every particular psychological state that it has the functional connections it has: that it is not merely accidental or usually the case that it has those connections; and that any psychological state would not be the psychological state it is, if its listed functional connections were different.

This second claim is clearly a strong one. It is a consequence of the view that a person is a Turing Machine taken in conjunction with the view that any machine state listed on the machine table owes its identity to the details of the instructions contained in the relevant machine table column (a view which was explained in the last section as being definitive of what a machine table is). We saw that for any machine table, two separately listed states must be identified with each other if and only if those separately listed states have machine table columns which contain identical instructions; and so it follows from this that, on the view that a person is a Turing machine, two differently specified psychological states are different if and only if their machine table columns contain sets of instructions differing in at least one detail.

An analogy might be useful in understanding this point. The announcement that any psychological-state concept is a functional-state concept might be compared with the truth that the concept of age is a historical concept. Just as it is essential to my age's being what it is that I was born in a particular year, it is said by the functional state theory to be essential to any psychological concept's being what it is that it has certain connections and relations with other states, and with behaviour. What specific psychological state a person is in is fixed by these connections and relations, in just the same way as what my age

is, is fixed by the year of my birth. The year of my birth is criterial.

The fact that the year of my birth is criterial does not entail, however, that no other tests or procedures can be used to find out how old I am. Clearly they can. And in an analogous way, the fact that there are certain things which are criterial of what psychological state a person is in does not entail that there are no tests or procedures of a non-criterial sort which can be reliably employed to determine a persons psychological state. So while it is of the essence of the functional state theory to stress the criterial features of psychological states, it does not make the mistake of denying that, in everyday life, the identity of a psychological state can be effectively discovered by non-criterial means.

3. CRITICISM OF THE TURING MACHINE THEORY

In the form in which I have just explained it, the functional state hypothesis is certainly much more specific and ambitious than any simple formula to the effect that psychological states have a role to play in determining behaviour. Now whatever the final merits of these more simple formulae, the functional state hypothesis in its fullest and most elaborate form is, I believe, defective in a number of respects. Accordingly, I shall devote this section to attempting to explain exactly how.

It has to be conceded at the outset, however, that the principle of individuating psychological states which the theory supplies has a certain amount to recommend it. The theory suggests, in effect, that psychological states are one and the same which have the same function, where the word "function" is understood in the prescribed sense. We can see the element of plausibility in this claim if we consider how certain facts about the history of psychology might be adduced to rebut a critic of the theory who claimed that psychological states are to be individuated solely or primarily by their linguistic specifications.

This critic does not consider that the concept of function need be introduced in order to individuate psychological states, for his suggestion is that the language we use to describe the mind supplies all and only the discriminations we need; so that each linguistically different specification of a psychological state specifies a different psychological state. "How else can such linguistic differences be explained?", this critic asks.

This critical question only presents a real challenge to the functional state theorist on the assumption either that there is a pair of psychological state-descriptions, say " P_1 " and " P_2 ", which are linguistically different but which are such that the two different states they describe have the same function, or else that there is a pair of psychological state-descriptions, say " P_3 " and " P_4 ", which though linguistically different, specify the same psychological state. Otherwise there is no conflict. But the critic can be answered in a general way without finding out whether such a conflict in fact exists, for in his own defence the functional state theorist can show that his theory reflects with a certain amount of accuracy the way in which theories of the mind have historically developed.

An example, which is in the nature of a parable, will suffice to illustrate how. Before the concept of the unconscious acquired a place in our thought, a single discrete psychological state was specifiable with the predicate

".... believes that fire is hot"

Then the concept of the unconscious was introduced, so that where there was previously only one state, there were now two, namely those specifiable with the predicates

".... unconsciously believes that fire is hot"

".... consciously believes that fire is hot"

Forging a difference between conscious and unconscious beliefs, the functional state theorist explains, was a process based on, and justified solely by, the fact that there is a functional difference between beliefs of those newly specified kinds. The meanings of the words "conscious" and "unconscious" was even explained and made learnable by reference to these functional differences.

If this short example shows what is generally the case when a psychological word is introduced into the vocabulary, then the advocate of the use of functional criteria is individuating psychological states possesses an argument with which he can answer his linguistic critic. However, in point of fact, the functional state theorist is claiming much more than that functional considerations are sometimes or usually relevant, or that they play some role in determining the identity of a psychological state, for he insists that the identity of any psychological state is completely determined by its causal and probabalistic connections with other such states and with behaviour. Moreover, what the functional state theorist wants us to accept is not just that causal or probabalistic connections are the factors which totally determine the identity of each psychological state, but that these facts can be captured by the resources of the theory of finite (deterministic or probabalistic) automata. The central question, therefore, concerns the credibility of these more ambitious doctrines.

Let me firstly discuss, in order to remove two objections against the theory which, in one way or another, impute circularity. In each case, the circularity is said to reside in the claim that the identities of psychological states are determined by their links with each other, and with behaviour.

There is one sort of circularity which need not detain us. One formulation of the theory says that a particular psychological state is identified by its functional (causal and/or probabalistic) connections

with other such states; so it might therefore be objected that a reference to the thing identified appears in the statement of the conditions which do the identifying, a reference which appears in the form of the pronoun it (see the first clause of this sentence). But this is not serious, for the pronoun in question does not have to appear. The theorist could alternatively say that a particular psychological state p is just that state which has such-and-such relations to other psychological states and to behaviour. For instance he could say such things as this: "the state of not knowing that p is just that state which causally precedes the state of knowing that p, in the event that the subject sees or understands for the first time that p". So at least it can be asserted in a non-circular way that a particular psychological state can be identified by what other psychological states it either follows or gives rise to (when certain specified stimuli impinge on the organism).

The second imputation of circularity is more interesting because, unlike the first, it is actually veridical. That is to say, it points to a certain circularity in the functional state theory which actually exists. The imputation runs as follows. If any psychological state is identifiable only in terms of the functional links it has with other psychological states, then clearly we have to say the same for those: so that we have a situation in which each particular psychological state is identified in terms of something which is identified in terms of something which is identified in terms of what? Now one natural reply to this imputation is to suggest that the causal claim which runs forward in time may have an end in behaviour, and that the causal chain which runs backwards in time, so to speak, may have a beginning in some stimulus-event. Behaviour and initial stimuli are the points at which attempts to identify psychological states anchor themselves.

But in point of fact this reply will not do. Behaviour and stimulation

cannot be the non-mental places at which mental identifications anchor themselves, because it is simply an error to suppose that behaviour and stimulation can themselves be specified non-mentally. Stimuli, whatever they are, have to be seen or judged or accented by the agent or organism to be of a certain kind, and behaviour, in so far as this means intentional action, has to be connected in some way or other with the intentions and beliefs of the agent. The presence of the mental cannot be eliminated by talking hastily of stimuli and behaviour; and so it appears that each psychological state, if identified in terms of its forward and backward causal links through other psychological states to the organisms "input" or "output", is really being identified in terms of things of the same sort as itself. It seems to follow from the functional state theory that there is a sense at which attempts to identify psychological states cannot completely escape the mental sphere. As a matter of fact, this circularity seems to be nothing other than what Brentano is credited with having noticed (and which I argued in favour of in the first Chapter), that the mental is irreducible to the behavioural.

In the last few paragraphs I have been concerned to show that the functional state theory cannot be defeated by either of the imputations of circularity which I described. However, this does not render it impregnable. In fact, I now want to concentrate on the central question of whether the truths which need to be captured about psychological states can, as the functional state theory says, be captured within the framework of the theory of finite automata. For a number of reasons, the answer which I want to suggest is in the negative.

An organism describable by a machine table can only be described

as having a finite number of internal states. In Putnam's words, a Turing Machine is

"a device with a finite number of internal configurations, each of which involves the machines being in one of a finite number of states"¹⁹

The question which arises here is whether a human organism can be adequately represented as having only a finite number of possible psychological states, and as being subject to only a finitely various input.

The facts suggest that the answer to this question is No. But we must be careful to distinguish the number of psychological states a person can be in at any one time, from the number that are possible for him altogether. It is a common assumption in philosophy that a person can, at any particular instant, be in more than one psychological state; but this is an unwarranted assumption, and only serves to confuse the issue. It is a mistake, in the first place, to infer from the fact that a person had more than one belief or desire, that he was in more than one mental state - for in fact there is only one mental state which a person (more accurately: his mind) can be in at any particular time. Supposing someone were to say: knowing that the earth is round is a psychological state, and believing that the universe is finite is another, so they will appear as two distinct states on the machine table, with a different set of connections (either deterministic or probabilistic) specified for each; and such that a person who had that knowledge and that belief at the same instant in time was in both states at once. In that case, the suggestion might continue, a deterministic machine table will specify two instructions corresponding to the next input (whatever it is), one for the knowledge state and one for the belief state. But then the problem will arise that they cannot both specify the next psychological state for the person, or the next most likely, for the chances are that the

19. Minds and Machines, p. 140. Already quoted; see footnote 15 above.

instructions corresponding to each of the two states will be different. Perhaps one instruction takes priority, but which? And Why?

But surely anyone who reasoned in this way would have made a basic mistake about how many psychological states a person can be in at any one time. It may be a mistake which ordinary language encourages, but it is a mistake all the same - since the fact is that a person can only ever be in one state at any one instant of time. If a person knows that the earth is round, and if he also believes that the universe is finite, then we have not described two states, but have only given a partial description of one state, the state of his mind. Nor is this a definition.

Consider a common object like a lawnmower, by analogy. Its blades may be rusty, its box may be structurally unsafe, and its carburettor, if it has one, may be out of order. These are states which the blades, the box, and the carburettor respectively are in; but it does not follow, just because these are parts of the lawnmower, that the lawnmower itself is in three states. It is in one complex total state, a state which can be described by saying, among other things, that part of the lawnmower is rusty, part is structurally unsound, and part is out of order. So too for the mind. Although in describing a person as thinking about Vienna or desiring a drink we may for all we know be giving descriptions which apply to just a bit or a portion of his brain or nervous system, the fact remains that in giving these descriptions we are not describing anything other than his mind: we are describing the single state his mind is in at the time. The mistake of thinking that a person can be in more than one psychological state at any particular instant in time is made by Block and Fodor, in the course of their discussion of Putnam's theory:

"behaviour can be the result of interactions between simultaneous mental states ... the functional state identity theory can provide for the representation of sequential interactions between psychological states, but not for simultaneous interactions. Indeed the functional state identity theory even fails to account

for the fact that an organism can be in more than one occurrent psychological state at a time, since a probabilistic automata can only be in one machine table state at a time" 20

If a person believes that p and believes that q, then we contribute to a total description of his single psychological state by saying that he does (being careful to avoid saying that he believes p and q). A machine table description can capture this fact without difficulty.

What the machine table description cannot capture, however, is the fact that the total number of different psychological states which are possible for a person is infinite. It is true that there is not an infinite amount of time in the life of an organism, but this only shows that the number of different psychological states the individual will actually pass through is finite.²¹ The number possible for him altogether

20. Block and Fodor: What Psychological States are Not pp. 170-1.

21. In spite of the following two arguments. The first is the argument from the divisibility of states like belief, knowledge, and so on; and the second argument is the argument from the divisibility of change in general. Both can be dealt with quite quickly. The first argument is this: Suppose I believe that the sky is blue, in the sense of believing that each part of the sky is blue. Then it follows that for each part, I have belief concerning it, that it is blue. But then a division of the sky into infinitely many parts makes the number of my beliefs infinite. The argument rests on the simple mistake, however, of confusing a single belief about an infinitely divisible thing with an infinitely divisible belief. It cannot be inferred from the fact that I have a single belief about A, together with the fact that A has an infinite number of parts, that I have a belief about each of the parts.

The second argument is similar, and faulty for much the same reasons. It is this: suppose I see a thing changing colour from blue to green, and that I believe what I see. Then since the number of colour gradations between blue and green is theoretically infinite, like the number of points on a finite line, then it follows (so the argument goes) that I can be said to have a belief corresponding to each of the colour-states the thing passes through. Or take the following argument: I see a thing which is blue at the left-hand end and green at the right hand end, and I inspect it from left to right, believing what I see; therefore I must have a distinct belief corresponding to each of the infinitely many colour-tones between the left hand end and the right hand end. In the first example there is a gradual change in an object, and in the second example there is a

(footnote continued overleaf)

is infinite, since what he believes or desires can be specified by attaching any declarative sentence to the words

He believes that

or

He desires that

(Etc.)

and we know, because of the recursive and generative nature of syntax, that the number of such declarative sentences is infinite. This would be enough to trouble the theory indeed, even if it were not for the infinite number of stimuli that can impinge upon a person and determine how his internal states change. These stimuli are, for the linguistic adult, themselves largely linguistic, for any sentence in English is sayable in his company. It seems clear that finite machine table is too small to accommodate the potentially infinite number of psychological states that a person can be in, and also that a finite machine tape is too small to accommodate the infinite variety of linguistic stimuli that a person might receive.

In making this point, I might just as well have used the notions of competence and performance which Chomsky employed in explaining how a finite language-user could have some conceptual grasp of a generative

(Footnote 21 continued from page 63

gradual change in what I see.

But in both cases we again have a violation of the evident principle that belief is not divisible. When I see a thing change colour, and believe what I see, then what I believe is that the thing changes colour. But it does not follow from this that for each minute colour-change, I have a belief concerning it. Or even if I do, the fact that I do would not be logically connected with my belief that the object changed colour.

language. A machine table description must go further than giving an account by hindsight of the finitely various performances of one or a group of organisms, for it must give a complete predictive account of what psychological states an organism would attain in the event of any one of an infinite variety of linguistic stimuli. But the description which the theory of Turing machines offers contains no recursive or generative devices which enable anything other than a simple enumeration of a finite number of psychological states to be listed. To this extent, a description of an organism which is given on a machine table is bound to be inadequately rich.²²

The objections described so far have concerned the claims made by the functional state theory about the way in which psychological states must be individuated from each other and captured by the resources of the theory of finite automata. The second group of objections, which I come to now, concerns the claim that shared psychological states share their machine table columns, in the total machine tables for the organisms concerned. Again, I think it can be shown that no plausibility whatever attaches to this claim, and therefore that the functional state theory as a whole is defective on this count as well as on the first.

-
22. Computing engineers tell us that, at least for the case of digital machines, any binary computing machine can be described as a Turing Machine, i.e. is a Turing Machine. Our problem about representing the linguistic capacities of an adult speaker of English on a machine table could be formulated, therefore, in terms of the following question: can an actual person's speech capacities be represented as a digital computing machine's capacities for printing out (binary) sequences of 0's and 1's?

I suspect the answer to this question is again no. I suspect also that the answer lies in the capacity of a binary print-out to represent only languages structured in a simple phrase-structure way. If this guess is correct, then Chomsky's demonstration that a phrase-structure grammar is (by itself) inadequate as a description of English would be additional evidence, or anyway evidence from a different quarter, that no binary computing device can model the linguistic capacities of a speaker of English.

It is a consequence of identifying psychological states functionally that two animals of different species can have the same psychological state, although their brains are in different states from a certain structural, or physico-chemical point of view; the same goes for two different animals of the same species, and the same goes for two different time-stages of the same animal. And it is because of this consequence that the theory seems to some extent attractive. Plausibility also attaches to the theory because, as I explained earlier, it seems to reflect some facts we know about the process of theory-construction in psychology.

But we must weigh the aspects of the theory which make it seem attractive against other aspects which make it seem less plausible. Consider the following case: a person (call him P) who is in pain will, with a probability of 70%, change his state to the state of being in pain and having a readiness to shout "I'm in pain" when some sympathetic person comes near.²³ Part of his machine table presumably will then look like this:

	State A	State B
	= pain	= pain + a readiness to shout "I'm in pain".
input = some sympathetic person is seen to approach	B -

(Figure 3)

23. Concentrating on the case of a probabilistic machine rather than a deterministic one is in one respect relatively realistic. We simply do not know of any deterministic laws connecting one type of psychological state with another; nor can we be sure that there are any. But we are better at assigning probabilities to various transition-events between one psychological state and another in the face of various different stimuli.

where the blank "_" in the left-hand instruction is filled by whatever "symbol" the machine is to print. But if this connection between states is part of what defines the state of pain for that person, then it becomes impossible for a dog, say, to share the state of pain, since, lacking a language, he can never be described as being ready to shout "I'm in pain". The functional state theory requires that shared psychological states have to be described as having shared functional connections with other states and with actions, but the state of pain for P seems to require a response which quite clearly the dog could not produce.

This is a particular instance of a general difficulty. A second person, person Q, might have a readiness to shout something in French or Dutch or Singhalese, if these were his languages. Or he might have a readiness to shout something in English using different words from those which person P has a readiness to use. One way of overcoming this general problem, it seems, might be to re-describe state B in a way which would enable a dog to share it, and which would also avoid any dependence upon the relevant expressions being in English. Suppose State B was to be re-described as the state of being in pain and having a readiness to signal (or call) for help. Clearly this is a possible dog-state, and clearly it contains no mention of an expression in English (or in any other particular language).

I do not know whether the particular re-description just suggested would find acceptance among functional state theorists. Perhaps it would not. In any event, the problem which the theory has to confront at this point is clearly one of generalising from the machine table description for one individual of a particular species or kind, to machine table descriptions for other individuals of the same species or kind. One of the demands it must be reasonable to make of a theory of the kind put forward is that machine tables for individuals visibly fall into

psychological kinds in just the way that the individuals do - in other words, psychological state descriptions, stimulus-descriptions and response (or action) descriptions must, it seems, be generalisable.

Suppose our concern was not with the state which the state of pain produced, but with the action which the organism performed as a result of his being in pain. We could describe one human action of this type as yelling that it hurt, but once again this is an action which, clearly, no dog could perform. So here too the problem exists of whether a suitable re-description of the human action could be found, which could be applied to a dog as well.

Finally I want to discuss a similar type of difficulty which the functional state theory presents. In the case where the machine table is probabalistic, the law-statements it contains are a fortiori not deterministic but probabalistic. The laws say things like: In the event of a certain type of stimulus, an organism in state X will change into state Y with such-and-such a probability. But the statement must not merely record a statistical generalisation (like "If anyone is an Englishman, then there is an 80% likelihood that he lives in a city".), but a statistical law (like "If a coin is tossed, then there is a 50% chance of its coming down heads".).

Now a difficult problem for the theory appears at this point, since it seems that a situation which the theory ought to allow is one in which there are two psychological states, state A and state B, which are such that the probability of a change from state A to state B for one person is slightly different from the probability of a change from state A to state B for another person. But can the functional state theory actually allow that a change from state A to state B for the first person occurs with probability 70%, while only with probability 65% for the second person? It cannot; for according to the theory, any probability

difference implies a difference of state. To put the point more clearly: the different probabilities of change in the situation just described would be registered on the machine tables for the organisms concerned. But since states are different when the associated machine table columns are different, it would follow from the theory that not both people could have been in state A to begin with. Indeed the theory disallows any situation of the kind described.

But this consequence of the theory is completely counter-intuitive. Normally speaking, we should either regard the two individuals as having different personalities (but broadly speaking the same range of psychological states) or else we would account for small probability-differences in some other way - by appeal to such everyday concepts as idiosyncrasy. Or perhaps the differences would be quite unimportant.

We want the theory to preserve our belief that psychological states are shared even when there are slight differences in the transition-probabilities; how could it do so? There seem to be two possible ways. Rather than count the psychological states of two subjects as different in type, as a consequence of probability-differences, the theory might propose to count them as being of the same type, but different in degree of intensity. A probability-difference of the sort I mentioned, for example, could be seen as an indication of a difference of degree of belief (desire, pain, etc.) between the two subjects rather than a difference in the state itself. All that needs to be said about this course, however, is that a slightly higher probability of response to a stimulus is not normally a guide to a greater degree of belief, pain, or whatever the state is. How much or how strong the belief, pain or desire is which a person suffers is not measured by any such simple probability measure, but by an enormously complicated set of different expressive and conventional features of his behaviour.

A different way for the theorist is for him to discount small

numerical differences in the transition-probabilities from one state to another, and to erect clusters of probabilities, such that if two or more numerical probability-values are close enough so as to fall into the same cluster, the differences between them are discounted for the purposes of identifying psychological states. But again, this is a hopelessly arbitrary suggestion, and the theory as formulated gives no guide whatever as to how to re-structure it in a less ad hoc way.

Problems of this sort are problems which arise in generalising from the case of one individual and one machine table. It must be a constraint on the theory that for a range of individuals (say people), the machine tables for each individual must fall into types or kinds in just the way the individuals do, in so far as their mental life is concerned. Putnam recognised this problem when he said:

"the difficulty of course will be to pass from models of specific organisms to a normal form for the psychological description of organisms" 24

And yet without a solution to this difficulty the functional state theory can hardly be said to live up to its promises. It is not sufficient for the functional state theory simply to propose that each individual organism can be described by a machine table which lists its psychological states: any adequate version of the theory must show us how the machine table description for an individual organism can be generalised in such a way as to reveal the similarities with other organisms who share its states - for this, after all, was one of the original pretences of the theory.

24. Putnam: Psychological Predicates p. 43. Until recently it used to be fashionable in philosophy to speak of "pain-behaviour". Resorting to this expedient to overcome the generalisation-problems raised here would be trivial and worthless (and as confusing as it was in its original context). Is there "belief-behaviour" and "desire for an apple behaviour" too?

Similarities in machine table descriptions must occur where psychological similarities exist between organisms. Even if the theory can provide for each individual singly, which I have argued is doubtful, it seems unequipped to make the psychological similarities between individuals explicit.²⁵

4. RELEVANCE OF THE THEORY FOR PHYSICALISM

I end my investigation of the functional state theory with some remarks about its physicalistic and anti-physicalistic pretensions. The original insight which to a large extent prompted the theory in the first place was that two different creatures could be in the same psychological state although from some narrowly structural or physico-chemical point of view their brains were in qualitatively different states. Putnam assumed that this is a simply empirical fact, and I follow him in this assumption²⁶ if it means that under some suitably detailed type of physico-chemical description there is a difference to be found in the cerebral matter of two psychologically similar organisms. Moreover Putnam deduced that this empirical fact spelt the falsehood of any identity theory asserted at

-
25. A completely different kind of problem, and an additional one, concerns the extent to which the so-called mental events are implicitly accounted for by the theory. A machine table lists only mental states; the question therefore is whether a given mental event can be identified with the event of change or transition between one appropriately selected mental state and another. Perhaps this can be done, but there are likely to be difficulties. Learning that p (an "event") is not simply a matter of coming to be in the state of knowing that p, since if it was, it would be indistinguishable from remembering that p or seeing that p (which are both also "events").
26. In spite of the fact that the theory that psychological states are machine table states, to which the assumption leads, is almost certainly wrong.

the Property Level: that is, that for any identity theory which asserts anything having the generality of:

being in pain = being _____.

there is a physico-chemical term suitable for occupying "_____" which is of a sufficiently detailed kind as to make a statement of that generality false.

It seems that Putnam also assumed it to be an empirical fact that one and the same organism could, on two different occasions in its life-history, be in the same psychological state although his brain, under a suitably detailed kind of physico-chemical description, was in a different physical state on the one occasion from the state it was in on the other. And Putnam concluded from this assumption that any identity theory on the Person-Specific level is also empirically false; or in other words, that a statement of identity having the generality of:

Tom's being in pain = Tom's being _____

is falsified by the existence of a physico-chemical description D suitable for occupying "_____" such that Tom is D on the first occasion of pain and not-D on the second. I again follow Putnam in supposing that if the assumption is correct then such a conclusion would follow. But here it seems even less likely than in the previous situation that there is no kind of physical description under which his brain on one pain-occasion is in the same state as his brain on any other pain-occasion. Whether or not this is so is a difficult question, and until I return to it in Chapter IV (section 2) I intend to leave it unanswered. Let me say here that ~~it~~ it seems more likely on the face of it that the same

organism on two occasions (or two different organisms of the same species) will fall under a single physical description of some kind when in a psychologically similar condition, than that two organisms belonging to different species will. In point of fact even this latter possibility will turn out to have some justification.

To an identity theory asserted on the Time-Specific Level, similar facts and deductions cannot be supplied, for there is no generality at all in the identities asserted at that level. They must be of the form

Tom's being in pain at T_1 = Tom's being _____ at T_1

Whoever Tom is, and whatever his brain is like (whether its mechanisms are made of "grey matter" or balsa wood), it must be conceded even by an advocate of Putnam's theory that whatever is said about psychological states being functional states precipitates no obvious clash with an identity-statement of this particular form. In other words, the functional state theorist must concede that everything is, on the face of it, consistent with an identity theory of this non-general and rather uninformative kind.

Whether such an identity theory is likely to be true, and what it would exactly mean if it were, are also questions I reserve for later. I also leave until later any assessment of an idea of a different kind which has been advanced by Fodor on the basis of the agreed empirical facts just mentioned. Appreciating that the known facts about the brain and nervous system rule out the possibility of a certain kind of "identity theory" on the Property Level and on the Person-Specific Level, but not obviously one on the Time-Specific Level, Fodor suggested that

"the objects appropriate for identification with psychological states are sets of functionally equivalent neurological states"²⁷

27. Fodor: Psychological Explanation p. 118.

As I explained in the last section of the last chapter, this appears to mean that being in pain, for instance, can be identified with the set whose members are all the neurological states possessed by any organism at whatever time it is in pain: in other words, the set whose members comprise such things as Tom's being _____ at T_1 , Tom's being _____ at T_2 , Fred's being at T_1 , Joe's being at T_3 , and so forth, where each blank is filled by a (may-be different) physico-chemical description, however precise or detailed. But as I also said in that section, the Appendix of the essay is the place in which an assessment of such a view will appear.

IIB: A Theory of Mental Events

1. EXPOSITION OF DAVIDSON'S THEORY

In the first part of this chapter I concerned myself firstly with the theory that mental states are, in a certain sense, functional states of the organism who possesses them; and secondly with the more precise theory that a person is a Turing Machine. A conclusion of both theories is that any mental state can be identified with a set of functionally equivalent states of the brain: this is the physicalistic truth which would appear to follow if either theory were true. But I hope to have established that a great deal stands in the way of the theories' truth. The unelaborated theory that mental states are functional states seems not to account accurately for the facts, and the more specific view that an adequate description of a human being's mental capacities can be captured by the resources of the theory of finite automata seems to leave many problems dangling, and unsolved.

In the second part of this chapter I intend to conduct a similar examination of another theory which has a physicalistic conclusion. The previous two theories were both theories of mental states: the theory which now comes under scrutiny is one about mental events. Essentially it is due to Donald Davidson, and its conclusion is that many particular mental events are strictly identical, one by one, with particular physical events. So this is a theory at what I called the Time-Specific Level (see Chapter I, section 5). I have chosen to examine this theory for two reasons. The first is that Davidson's work in the philosophy of mind is a good deal more comprehensive in its scope - not to say subtle - than that of many authors. And like the functional theory of mental states, it is a theory which has a physicalistic doctrine as a conclusion rather

than, as in the original work of Place and Smart, as a context-free hypothesis. To this extent it is advanced by Davidson as a proof of the physical nature of some mental events. The second reason is that the arguments with which I believe Davidson's theoretical framework can be challenged happen to be exactly relevant to some of the main contentions of the functional state theory as well. That is to say that several of the points at which I think Davidson's theory is defective are just those where the functional state theory is defective too. To this extent my critical programme has a certain unity.

Davidson's argument is this.²⁸ First, that particular mental events and particular physical events can be causally connected: particular physical events cause particular mental events in perception and perhaps knowledge, and particular mental events cause particular physical events when an agent acts intentionally (intentional actions are mental events, for Davidson). Second, that for any particular singular causal sentence there must be a general causal law which it instantiates, in the sense that where there is a true singular causal statement, there exists some re-description of the relevant events in such a way as to make the statement explicitly an instance of the law.²⁹ Thirdly, that there are no strict deterministic laws connecting events when both are described psychologically, and that there are no strict deterministic laws connecting events described psychologically with events described physically. Fourthly, and therefore, the general laws of which particular mental-physical causal statements are instances must be framed in physical terms; but it follows from this that any particular mental event which interacts causally with a physical event must have a physical description, which is to say

28. Expounded in Mental Events (In Experience and Theory, Eds., Swanson and Foster).

29. Argued in Causal Relations JP 1963.

that those mental events are physical events.

Now for my own part I regard the steps in this argument as not all of equal plausibility. I should accept the second point, because it to a very large extent defines the sense of the word "cause" which is being employed in the argument. Step three concerns Brentano's controversial doctrine of the "irreducibility" of mental language, and the question of whether mental predicates can occur with others in a statement of law; both of which I discuss at some length at other points in the essay (see Chapters I and IV). And stage four represents the conclusion. But what I believe has to be seriously questioned is the doctrine expressed and the presuppositions implicit in stage one, which asserts that particular mental events interact causally with particular physical events.

The view that the mental and the physical interact causally is, as we have seen, one part of what the functional state theory says. According to that theory, there can be complex and long causal chains which, beginning with some stimulatory event, consist of the range of successive mental states the organism passes through, and end with behaviour. A particular mental state may be causally or functionally related to another mental state, which may in its turn be causally or functionally related to a further mental state, and so on. The initial mental state in such a sequence may be causally related to an antecedent physical event "outside" the organism, and the terminal mental state may also be causally related to a physical event "outside" the organism (namely his action). Those who speak of causal interaction between the mental and the physical may therefore be represented as speaking of those initial and terminal links: in the first of which the physical causes the mental, and in the second of which the mental causes the physical.

I shall concentrate here on Davidson's treatment of the terminal links, in which, in his view, the mental causes the physical when a person

acts for a reason. The sense of the word "cause" which is relevant to this discussion is, it must be emphasised, that deployed by Davidson himself. According to him, causality is an extensional relation between particular events, and the existence of a true singular causal statement implies that there is a general causal law (although we may not know what it is) under which the singular statement can be subsumed. To interpret the word "cause" differently is to change the subject.

2. CAUSAL INTERACTION BETWEEN THE MENTAL AND THE PHYSICAL: REASONS AS CAUSES

Davidson's argument appears in his classic paper "Actions Reasons and Causes",³⁰ which sets out to show how actions, which in his view are physical events, have causal antecedents which are mental; and that an account of those mental antecedents can be given in terms of an agent's reason for his action. In Davidson's view, the kind of reason for an action which is the cause of an action is what he calls a primary reason. The explanation of what a primary reason consists of is as follows: an agent's primary reason for an action consists of a "pro-attitude" towards actions of a certain type, and a belief that the agent's action is an action of that type. Davidson explains that a pro-attitude can be a desire, a want, an urge, a belief, in a fairly loose sense, provided it can be interpreted as an attitude which is appropriate for an agent to direct towards actions of a certain kind. The belief which enters into a primary reason is simply a belief to the effect that the action concerned is an action of the specified kind. When combined with the fact that a

30. Journal of Philosophy 1963. The doctrine that the physical causes the mental in perception and possibly knowledge has its modern origins in Grice's The Causal Theory of Perception.

reason is only a reason for an action when the action is described in a certain way, we can, according to Davidson, give a necessary condition for reasons which are primary, in the following terms:

"R is a primary reason why an agent performed the action A under the description d only if R consists of a pro-attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description d, has that property" 31

Some examples might help to clarify how primary reasons are supposed to contain a pro-attitude and a belief. Suppose that

(1) I flipped the switch because I wanted to turn on the light

then "I wanted to turn on the light" would be said to give the primary reason for my flipping the switch; it would be said to consist of a belief that my switch-flipping action is an action of the type which causes lights to turn on, and it would be said to consist of some sort of pro-attitude, towards actions of that type (actions which cause lights to be turned on); In this example, I think, it is correct to assume that if the agent did flip the switch because he wanted to turn on the light, then he must have believed that his flipping the switch was an action of the kind which causes lights to turn on, for it is only in the light of this assumed belief that the reason he gives is intelligible as a reason. Consider a second example. Suppose that

(2) I went into the shop because I wanted to buy the watch in the window.

In this case also it does seem that we can explain how the fact that I wanted the watch in the window was the primary reason for entering the

31 Actions Reasons and Causes p. 699.

shop by pointing to a belief that my entering the shop was an action of the type which leads to situations in which watches can be bought. So the thesis that a primary reason for an action is connected with a belief is a plausible one. There are two minor difficulties, however, which it is worth mentioning in passing. The first of these concerns the notion that a primary reason is connected in a certain way to a pro-attitude of the agent's, in addition to a belief. Pro-attitudes, Davidson explains, include

"desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed towards actions of a certain kind" 32

The inclusiveness of this list suggests that a pro-attitude is not necessarily a desire for a certain goal - although it may be. If the agent's pro-attitude on any occasion of action was simply a desire for a goal, then it would be a simple matter to specify the pro-attitude given a specification of the primary reason. In the first of my examples the pro-attitude would be specifiable as a desire for the light to be turned on, and in the second example the pro-attitude would be specifiable as a desire to buy the watch in the window. But if a pro-attitude can also be an urge or a prompting or an economic or moral or aesthetic prejudice, then there is some difficulty in automatically constructing the agent's pro-attitude from a mere specification of the primary reason alone. But whether the statement of the pro-attitude should be so intimately connected with the statement of the agent's primary reason is not altogether clear.

A second minor difficulty, and a related one, concerns the concept of constitution. A primary reason is said to consist of a belief and a pro-attitude; but what exactly does this mean? In the case of the doctrine

that a primary reason consists of a belief, we might interpret "consists" in terms of entailment, for it seems true that a sentence like

(3) P did A because he wanted to bring about situation g.

cannot be true unless it is also true that

(4) P believed that doing A would help to bring g about.

which is to say that (3) entails (4) - although (4) does not entail (3), even when P did A. But whether this interpretation of "consists" is the intended one is again not altogether clear. Nor is it clear whether saying that a primary reason consists of a belief and a pro-attitude means the same as saying that a person's having a primary reason consists of the person's having a belief and a pro-attitude.

But let us waive these minor problems, and attend to the crux of Davidson's argument, which concerns the notion of cause. Davidson gives various arguments designed to show that the primary reason for an action is its cause, the main one of which can be introduced as follows. The first step is this: that to know that a person had a reason R for doing A, and that he had certain beliefs and attitudes which are appropriate to any situation in which R is actually the reason for A, and to know also that he did A, does not allow us to infer that he did A for the reason R. The mere having of a reason for an action does not explain the action, so it is alleged, even when the action is performed while the agent had the reason. The reason must not merely be had; it must be efficacious. It must be the reason which explains the agent's action.

The next step in the argument is to ask: what is the nature of the efficacy which a reason has, when the reason is the one for which the

action is done? This situation is alleged to be illustrated by an example of the following kind. Suppose that a man wants to buy a watch, and suppose that he knows that going into the watch-shop will enable him to buy the watch. And suppose he goes into the shop. Now merely on the basis of this information we cannot explain why he went into the shop by saying that he wanted to buy the watch, since, although (allegedly) we can say that he had a reason for entering the shop, the reason in question might not have been efficacious. It might not explain why he went into the shop, since some other reason might have been operative at the time. He wanted to insult the owner, say. So the question arises: what is the link between a reason and an action when a reason which the agent had is one which explains the action?

Davidson's answer to this question is that the link must be causal; but the final steps to this conclusion are weak. One reason he gives is simply that the word "cause" is the only respectable synonym we have for the word "explain"; so explaining an action by citing the reasons for it must be understood as a kind of causal explanation. "One way we can explain an event is by placing it in the context of its cause; cause and effect form the sort of pattern that explains the effect, in a sense of "explain" that we understand as well as any". He ends his argument with a challenge:

"If reason and action illustrate a different pattern of explanation, that pattern must be identified" 33

And he concludes

"If causal explanations are wholly irrelevant to the understanding we seek of human action then we are without an analysis of the "because" in "He did it because", when we go on to name a reason" 34

This then, is Davidson's main argument. To explain an action by citing

33. Actions Reasons and Causes; p. 692.

34. Actions Reasons and Causes; p. 693.

the reason for which the action was done is to explain it in the light of it's cause. And to redescribe an action in terms of the effective reason is, like redescribing a conflagration as "what the short-circuit produced", to redescribe it in terms of its cause.

This argument is not without ingenuity; but it is, I believe, faulty. It derives its apparent strength from seeming to offer relief from two related pressures; one pressure is the pressure to explain the difference between cases of acting on or for a reason and cases of merely having a reason and acting; and the other pressure is the pressure to elucidate the sense of the word "because" in a sentence like

(5) He closed the door because he wanted to stop the draught

and the answer which it gives to the first problem, that acting on a reason is acting as a causal result of the reason, is really identical with the answer to the second problem, that "because" in (5) means something like "was caused by" (although contextual adjustments have to be made if the one phrase is to be substituted for the other). And yet it seems to me that the positive reasons for adopting this conclusion are very weak indeed. For in the first place, as I hope to show, the idea of merely acting and having a reason, with which cases of acting on a reason are supposed to be contrasted, is not at all clear, and certainly not contractive in quite the way Davidson supposes. And secondly, it seems hardly sufficient, as an argument, to suggest that "because" in action-sentences has a causal sense, on the sole grounds that no other clear sense for the word is available.

Let us take each of these points in turn. I have said that the causal analysis of reasons derives a good deal of its support from the contrast between cases of acting on a reason and cases of merely having a reason

and acting. But what needs to be questioned is whether the difference between these two types of case is significant;³⁵ I shall begin to do this by questioning what it means to say that a person had a reason for doing something in the situation in which the something he did was not done for the reason in question.

Let us take the following case as an example. Suppose that on some days of the week I visit the University Library, but when I do so, my reasons for doing so are not always the same. On Mondays my reason for visiting the library is that I want to return the previous week's books. On Wednesday, my reason for visiting the library is that I want to read the new journals, and on Fridays my reason is that I want to chat to the librarian. Let us suppose this is a regular pattern. Now someone could, if they wished, describe me as having three reasons for visiting the library - one to return books, one to read the journals, and one to chat to the librarian. But in so describing me, he need not say that on each of my particular library visits I had three reasons for making that particular visit, one of which happened to be operative, depending on the day of the week. Nothing would oblige him (or us) to speak in this way; and this is the extent to which the distinction between "operative" and "inoperative" reasons is somewhat artificial. Instead of saying that on

35. Davidson is not the only person who believes this contrast to be significant. Charles Taylor, in Ch. 1 of The Explanation of Behaviour, says: ".... that something is an action in the strong sense (i.e. of being directed towards bringing about a certain condition as an end) means not just that the man who displayed this behaviour had framed the relevant intention or had this purpose, but also that his intending it brought it about. That is, it is not a sufficient condition of an action's occurring that a man intend to do something and that behaviour answering to the relevant description occur. For it is perfectly conceivable - and, indeed, happens in rare cases - that the two be unconnected, and that behaviour occur for some other reason" (p. 33).

the occasion of any particular library-visit I have three reasons but only one reason for, it would be far less artificial to say that for each particular visit I have one and only one reason for making it; and to leave the "inoperative" reasons out of the picture altogether. Monday's reason is different from Wednesday's reason, and each of these is different from Friday's reason - and the library-visits I make, although actions of the same kind (according to one obvious way of classifying them) are particular individual actions having particular individual reasons of their own.

So in this situation, if I am described as having three reasons for visiting the library, this does not mean that three reasons are had by me on the occasion of each particular library-visit, only one of which I select and somehow make operative depending on the day of the week. Saying this might lead us to assert that reasons can be had for an action which are inoperative reasons for that action, but this would be otiose. The best non-otiose description of a situation of this kind involves saying that for each particular action I have a single-(sufficient) reason, so that different particular actions are attended by the having of different reasons for doing it.

The supposition that a certain type of reason has to be classified as "operative" in order to distinguish them properly from those which are "inoperative" (merely had but not acted upon) may derive from the fact that very often we say of a person whose does, or plans to do, a certain thing, that "he had every reason for doing it", and yet where, in the event, he does not do it for the reason we have in mind. Let us examine a typical situation of this kind. Suppose we know that a certain person has always wanted to practice his hand at trout-fishing, and the man in question suddenly makes a visit to Scotland in the middle of the trout-fishing season. He knows, and we know that he knows, that Scotland is where you can get trout. Without knowing why he went, we might reasonably

say, on the basis of our knowledge about his aspirations as a fisherman, that "he had every reason for going" - but when he returns we learn that his Scottish aunt was taken ill, and that he made the journey not to fish trout, but to look after her while she was unwell. Can we say in this case that he had a reason for making the journey, but that the reason was not operative?

Again, there is nothing to prevent us saying this; but there is nothing to compell us to either. If we say in this situation that "he had every reason for going", having in mind his relatively long-term or permanent desire to practice his fishing technique, we do not mean that he actually had that reason as he went, but merely that he had that permanent or long-term desire, and that the desire might, for all we know at the time, have been connected with his reasons for making the journey. Before we knew the truth about the reasons he did actually have, we simply entertain it as a hypothesis that his wanting to fish trout was a reason of his; but this is not a hypothesis which we can go on entertaining once we know that he went in order to see his aunt. It is a hypothesis which, if he did not go for the trout-fishing reason, is just false. In the light of the facts as they turn out to be in this story, we cannot truly say that he had a trout-fishing reason as he went; but merely that he had a long-term or permanent desire to fish for trout. But this, according to the story, is something we know about him in any case. The statement "he had every reason for doing A, namely R" is used to mean something like "he very probably did A for reason R".

The standard situation is, I think, this. When we correctly say of a person, before \underline{t} , that he has a reason R_1 for doing A at \underline{t} , and if we then observe that he does A at \underline{t} , and if, further, he sincerely cites a different reason (R_2) for his doing A at that time, then we would standardly draw either of two conclusions. We would either conclude that both R_1 and R_2 were sufficient, so that R_1 needs no mention if

R_2 is given. Or else we would conclude, I think, that the agent had abandoned R_1 as a reason for that particular action by the time he came to do it - although it may be a reason for his doing some other particular action of the same type on another occasion. "He had a reason for killing her, although when he killed her that wasn't the operative reason" means something like "Given his state of mind, he might well have killed her for that reason - but in the event he did not".

I have been arguing that the concept of an operative or effective reason derives a good deal of its intelligibility from a supposed contrast with reasons which are had by an agent but which are not operative when he acts. But I have criticised this idea on the grounds that the notion of a reason which is had by the agent but not acted upon is one which is, at best, grounded in a certain way of speaking about behaviour which there is no necessity to adopt; and that because of this, the contrast between the two types of reason is an artificial one. We must now deal with the other part of Davidson's argument, which is to the effect that the word "because" has a causal sense when used in a statement like "He did so-and-so because he", where we go on to give the agent's reason.

We saw that the argument to this conclusion consisted of two suggestions: the first being that "because" in "he did it because ..." has a causal sense because "was caused by" is the only respectable synonym we have for the word "because"; and the second being, to quote Davidson again, that if "causal explanations are wholly irrelevant to the understanding we seek of human action, then we are without an analysis of the "because" in "He did it because ...". But this second claim is false, since although we have no reason to suppose that actions do not have causes in the same way that other events do, it does not have to be the case that the agent's reasons are identifiable as the causes; the actual causes might be events which the agent has no knowledge of whatever. And

as to the first claim, it is hardly a sufficient argument to say that the word "because" in a sentence like (5) must have a causal sense on the grounds that no other sense seems immediately available. For it seems that if this was a sufficient argument, then presumably it would be equally sufficient to show that the "because" in

(6) He collected his pay because yesterday was Thursday

is causal. But how could yesterday's being Thursday be the cause of anything?

The word "because" occurs in many contexts in which it is almost impossible to construe it as being synonymous with "was caused by"; for example:

(7) Peter is in London because Paul is out of town

(8) My umbrella is up because it is likely to rain

(9) $2+2=4$ because $1+1=2$

Here, the truth of the sentence preceding the word "because" is explained in some degree by the truth of the sentence following it; but the explanatory connection is not of the kind which holds between one event and another event which it causes.³⁶ Indeed it cannot be, because the sentences following the word "because" in these examples are not descriptive

36. Nor does the word "explain" necessarily mean "cause", as Bromberger's example shows: a pendulum's period being T can explain its length's being L, but it obviously cannot cause it. See S. Bromberger, An Approach to Explanation (in R. J. Butler, ed.).

of events.

It is of course one thing to say that "because" need not everywhere have a causal sense, and another thing to say that in a particular and well-defined kind of context it does not have such a sense. With regard to the well-defined type of context in which an action is connected to an agent's primary reason:

"He did A because he"

I suggest not only that the causal sense cannot be foisted on "because" in virtue of some general theory about that word, but also that there is a specific reason why it should not be; namely, that the causal relation as understood and expounded by Davidson is a relation between particular events, whereas the sentence following the word "because" in a statement of an agent's reason for acting does not often describe, or "apply to", an event. Davidson's arguments for the causal sense of "because" in such a statement are uncharacteristically inconclusive.

The persuasiveness of the doctrine that the agent's reason for his action is its cause is perhaps engendered by the illusion that reasons are entities. This is an illusion which is itself engendered, perhaps, by the use of such idiomatic phrases as "He gave a reason for A", "The reason for A is". But in the giving, offering, accepting or rejecting of reasons, nothing is literally given, offered, accepted or rejected, let alone entities of which the word "reasons" could be the generic name. I am in sympathy with Alan White, when he writes:

"the words "reason", "motive", "cause" do not refer to anything that could be a factor in an explanation of conduct, in the way that an antecedent event, a feeling, a habit, or an instinct could operate as such a factor. And this despite the fact that there is a sense in which the agent himself can be correctly said to "have" a reason or a motive. Reasons, motives and causes, unlike events, do not happen at particular times or places ..."³⁷

37. Alan White: Philosophy of Mind (Random House, NY. 1967) p. 135.

In spite of this, however, the idea that the mental interacts causally with the physical when a person acts intentionally admits of a restatement. This restatement, which I shall examine in the section which follows, involves saying not that the agent's reasons are causes, but that the desires and beliefs of which the reason (purportedly) consists are the causes of action.

3. CAUSAL INTERACTION BETWEEN THE MENTAL AND THE PHYSICAL; (b) MENTAL STATES AS CAUSES

The argument considered in the previous section relied to some extent on the importance of an alleged difference between operative and inoperative reasons. This approach is not uncommon. Charles Taylor adopted it, and applied it to intentions and purposes, erecting the same distinction between operative and inoperative intentions and purposes as the one Davidson suggested for reasons. Indeed Taylor actually treats the presence of an operative or productive intention or purpose as one of the distinguishing marks of action:

"the distinction between action and non-action hangs not just on the presence or absence of the corresponding intention or purpose, but on the intention or purpose having or not having a role in bringing about behaviour. Within action, we might say, the behaviour occurs because of the corresponding intention or purpose; where this is not the case, we are not dealing with action" 38

But I shall not rehearse fully my arguments for thinking this view defective, if it means that intentions or purposes are causes. It does not follow from the fact that the word "because" appears in the situation, that there is a causal link: and the distinction between operative and inoperative intentions and purposes is an illusion. And as is the case for reasons, intentions and purposes are not events by any stretch of the

38. Charles Taylor: Expl. of Behaviour p. 26.

imagination. It follows that their role cannot be strictly causal.

But any argument of this type, whether couched in terms of reasons, intentions or purposes, can be given a restatement according to which it is desires and beliefs themselves, (or desirings and believings, if this phrasing is preferred) which can be the causes of action. The doctrine that reasons, intentions, purposes (and things of this same peculiar logical type) can be causes drops behind with this readjustment, and the actual mental phenomena themselves come to the fore. The previous distinction between operative and inoperative reasons, for instance, becomes transposed into a distinction between operative and inoperative desires and beliefs.

In the next chapter I hope to show how little evidence there is for supposing that there are any mental events in any strict sense, but the traditional assumption is certainly that there are, so for the space of the argument I shall proceed as if the traditional assumption was correct. Given this proviso, it seems to me that the main difficulty which the restated view has to face is one of reconciling the necessity of locating events as causes (as opposed to mental states and other phenomena) with the notion that the phenomena normally cited or referred to in explaining actions are not events at all.

Let me explain this problem in a little more detail by referring back again to Davidson's theory that reasons can be causes. Davidson's theory is that primary reasons are (or consist of) beliefs and attitudes. Now he appreciated that there might be an objection to the view that these primary reasons are causes, and he expressed it by saying that beliefs and attitudes, being states of mind, are not events - and hence not things of the right kind to be causes. He then attempted to provide a refutation of this objection; but the attempt was, I think, suspect.

Davidson is one who certainly believes that causes have to be events;³⁹
and so it

39. See his Causal Relations. JP 1967

was incumbent on him to reconcile this assumption with the fact that many explanations which seem to be causal mention only the states of an object. He attempted to achieve this reconciliation by saying that wherever a causal explanation is conducted in terms of an object's states, there must always be an event "closely associated" with the state in question. On the face of it, this can mean either the event of coming to be in that state, or else the event which itself caused the object to be in that state, or else some other event which caused the object-in-that-state to behave in whatever way it did. Davidson applied this method of reconciliation to the case of primary reasons, as follows:

"In many cases it is not difficult at all to find events very closely associated with the primary reason. States and dispositions are not events, but the onslaught of a state or disposition is. A desire to hurt your feelings may spring up at the moment when you anger me; I may start wanting to eat a melon just when I see one; and beliefs may begin at the moment we notice, perceive, learn, or remember something. Those who have argued that there are no mental events to qualify as causes of action have often missed the obvious" 40

Now in the first place it is hard to see how Davidson can both acknowledge that causes must be events and continue to maintain that primary reasons (beliefs and attitudes) are causes. For even if an event closely associated with a primary reason was successfully located, then it would be this, and not the beliefs and attitudes embodied in the primary reason, which would have to qualify as the cause.

Let us examine the possibilities. As I said just now, there are three possible candidates for the "closely associated event": either the event of coming to be in that state, or the event which caused the organism to get into that state, or else some other event which causes the object-in-that-state to behave in the way it does. And yet it seems that in none of these cases could the event in question constitute the reason, or even plausibly be cited in the giving of a reason. If the primary reason

for taking the left-hand fork is that the traveller wanted to get to Katmandu, then beginning to want to get to Katmandu, or getting into the state of wanting to get to Katmandu, are likely, if they provide reasons at all, to provide reasons which would rationalise some action other than taking the left hand fork. These events may even have occurred long ago in the agent's past. The actions which "I wanted to get to Katmandu" is capable of rationalising are not necessarily the same actions which the expression "I began to want to get to Katmandu" is capable of rationalising - if there are any of these. Nor are we likely to rationalise the action which "I know he was depressed" rationalises with the expression "I learnt that he was depressed". Indeed, statements reporting the beginnings of mental states are very seldom used in reason-giving at all.

What about the event which caused the organism to get into the state in question - the event which itself caused the event of the organism's coming to be in the state in question? Davidson says "a desire to hurt your feelings may spring up at the moment you anger me; I may start wanting to eat a melon just when I see one", as if to suggest that his being angered caused the nearly-simultaneous event of his beginning to want to hurt the other person's feelings, or that his seeing the melon caused the nearly-simultaneous event of his beginning to want to eat it. If the caused event of these event-pairs is supposed to be the cause of the ensuing behaviour, then the considerations of the last paragraph apply. But the other alternative - that the causing event (his being angered, his seeing the melon) is the cause of the ensuing behaviour - cannot be correct, for the simple reason that the effect of those events has already been given, but not as the event of behaviour itself. Ex hypothesi, the effect of his being angered is his beginning to want to hurt the other person's feelings, and not the ensuing insulting action; and the effect of his seeing the melon is his beginning to want to eat it, and not the ensuing

action of attempting to get it.

In point of fact, when we explain an action we very seldom do so by citing a mental event. And when we cite a state of mind (a mental state), there is seldom a "closely associated event" which, in the context, occupies the right causal role in relation to the ensuing behaviour. If the relevant state of mind itself explains the behaviour, then indeed it is not surprising that no other phenomenon, however "closely associated", does so as well, for in general different phenomena tend to explain different things. So I conclude that even if mental phenomena themselves, rather than reasons, intentions or purposes are put forward as the explananda of actions, then there is little plausibility in the view that these play a causal role in the production of behaviour. States of mind and mental events do enter into explanations of behaviour, but not into causal explanations. States of mind are of the wrong logical type to be causes, whereas mental events, while of the right logical type, never occupy the right causal role in relation to the behaviour itself.

A typical reaction to any argument which purports to show that the mental does not causally interact with the physical is to protest that it appears to support a view of mental phenomena and physical phenomena according to which they occupy different realms and have distinct, although may be parallel, existences. The mental must affect the physical, Charles Taylor appears to argue, for

"... how could we even exist as rational life if the realms of mind and matter functioned independently of one another?"⁴¹

But that the mental does not affect the physical in any sense only follows from the fact that the mental does not causally interact with the physical if several other premises are supplied; but these premises, if listed

41. The Explanation of Behaviour, p. 53.

carefully, can be seen to have little plausibility. For instance they must at least comprise (a) the claim that there is such a thing as the mental realm and such a thing as the physical realm, or alternatively that there are mental phenomena and physical phenomena, as well as (b) the claim that the only explanation of the word "affect" is in terms of a specific and restricted concept of "cause". The first of these premises needs spelling out and carefully examining (see Chapter III); and the second premiss is false. There are several types of relation between one kind of phenomenon and another which, while not equivalent to the causal relation, are quite sufficient to prevent the conclusion that the mental and the physical "function independently". The existence of even the weakest sort of explanatory relation between mental statements and physical statements is enough to show this. Another kind of relation which would falsify the strange parallelist view is that mental statements can play a classifying or taxonomic role in relation to the physical. Since I suspect that this is in fact the relation between a mental statement and a statement of action, in the situation where the mental statement is the statement of the agent's reason for his action, I shall end this section by saying a brief word about it.

Suppose that Henry did something whose intentional description is "Henry poisoned his mother". Now we can complete the story in any number of ways, by imagining any of the following statements to be the answer offered to the question "what was his reason?"

1. He wanted some time in jail
2. He wanted to use up the poison
3. He wanted to harm the person he hated
4. He wanted to revenge for his father's death
5. He wanted to obey M's instructions
6. He wanted to indulge a momentary whim

7. He wanted to end her unhappiness
- 1'. He likes going to jail
- 2'. He didn't want poison hanging about the house
- 3'. He hated her
- 4'. He believed she killed his father
- 5'. He was instructed to by M.
- 6'. He was subject to a momentary whim
- 7'. He knew she was unhappy

Sentences 1-7 are quite different in kind from sentences 1'-7': the former contain a reference to some future state of affairs which the agent wants to bring about, while the latter merely mention the agent's state of mind. Sentences of the first kind, it seems to me, have a central place among reason-giving statements in general, which sentences of the second kind do not. What seems to be generally true about sentences of the first group is this: that when they are put forward by a speaker to an audience as reasons for an action A, they are put forward in a way which invites the audience, or makes it possible for the audience, to re-describe the action in question as one seen by the agent as a means to a certain end. This is possible, since the sentences used mention a future state of affairs which the agent wishes to bring about by the action.

The invitation implicitly offered by the speaker to his audience is to re-describe the action in terms of the effect which the agent believes (hopes, expects, etc.) it will have. The sentence "He poisoned his mother" names the same action as the sentence "He poisoned his mother to use up the poison", or as the sentence "He poisoned his mother to avenge his father's death", or the sentence "He poisoned his mother to end her

unhappiness". The sentences describe the same action, namely the thing the agent did, but all except the first ("He poisoned his mother") re-describe the action in a way which displays the agent's purposes, hopes, or goals.⁴² To this extent the reason-statements of the first group have a classifying force: they enable the action they explain to be classified in terms of the mental fact which the reason-statement gives. This is even a conclusion which is not wholly unacceptable to the causal theorists, for they can agree that a reason-statement enables a classification of the action in question; but the difference of opinion emerges when the causal theorist goes on to say that the reason-statement enables an action to be classified in the light of, or re-described in terms of, its cause. This additional piece of theory is one for which, as I have indicated, there exist only bad arguments.

If I am right, then we can see clearly the sense in which an agent's reason does explain his action. The explanation provided is of the taxonomic kind; to illuminatingly group an action with others, to place it in the category to which it belongs, is to explain it by displaying its links with its neighbours, rather than by displaying its temporal antecedents. Taxonomic explanation may be explanation in a weak sense only, but it is this weak kind of explanatory connection which typically binds an action to a reason.

To summarise the main points of this section. Davidson's proof that some mental events are physical requires the premiss that causal interaction takes place between mental events and physical events. This premiss is a hypothesis which is also central to other theories of mental events, which is why I have chosen to be carefully critical of it. That mental phenomena in general have an essentially causal or productive role

42. Cf. Melden: "citing a motive is giving a fuller characterisation of the action" Free Action p. 93

is part and parcel of the view that mental phenomena are "inferred entities", a theory which relies upon a certain analogy with the way objects are "posited" in the sciences; and an important part of the functional state theory is to advance the view that mental states are effective in causing certain actions when other stimulatory events take place. My arguments have been designed to show that if we take causality to be an extensional relation between events, then none of the ordinary cases where behaviour is explained by reference to the mental phenomena of the agent are at all appropriate to have this model of causality applied.

In fact, there are good reasons for suspecting that causality is massively more complex than this simple model suggests. If it is, then it remains to be seen whether the explanatory connection between mental phenomena and actions can be assimilated to it. But all this is hypothetical. I suggest that the assimilation of explanation by reasons to the simple causal model involves an undue amount of adjustment in the actual facts of explanation as they occur in situations. Reasons cannot be causes because they are of the wrong logical type; mental states cannot be causes because they are of the wrong logical type, although a different one; and mental events, are never, as far as I can see, involved in the right way in the business of giving primary reasons. Arguments for the physicalistic conclusion will therefore have to come from a different quarter.

Chapter III. Grammatical and Logical Complexities

1. Description of the problem
2. Events in general: the evidence from logic
3. The ontology of the mental
4. Actions and causes again: some remarks on their grammar

1. DESCRIPTION OF THE PROBLEM

In the last chapter I discussed two theories - a theory about mental states and a theory about mental events - whose aim was to establish a physicalistic conclusion. The first theory sought to establish that mental states could be identified with sets of functionally equivalent cerebral states, and the second sought to establish that any mental event which interacted causally with a physical event is itself a physical event. My general strategy was to expose some of the weaknesses of the arguments which led towards these conclusions. From this point onwards I want to attend more carefully to the details of those conclusions themselves, and of those like them.

The kind of scrutiny which I believe all physicalistic conclusions deserve is from the general point of view of their intelligibility. By this I mean that considerations which can be broadly described as grammatical or logical must be brought to bear on the actual statements which physicalists offer as the conclusions to their theorising.

This methodological point might be expanded in the following way. Most philosophers who seek to establish a physicalistic conclusion do so because they see a distinction, however unclearly, between physical phenomena on the one hand and mental phenomena on the other, and so the point of their theorising is to reconcile this appearance of dualism with a monistic view of the world which, usually for independent reasons, they find attractive. One modern method of achieving this reconciliation is to identify mental phenomena with physical phenomena; in other words, to identify mental states with physical states, mental events and actions with physical events, mental processes with physical processes, and so on. Now the plausibility of this "identification" approach has been the subject of endless discussion in the last decade or two, and most of the controversy has surrounded the

question of whether, in accordance with the logical requirements of "=", the properties possessed by phenomena on the mental side are possessed by phenomena on the physical side. Certainly questions of this sort have to be answered if they can be coherently posed; but the effect of the present chapter will be to suggest that very many prior questions of a quite different kind have to be raised and settled before the technicalities of Leibniz's law become relevant. To put it bluntly, we must take the utmost care in deciding first what mental phenomena there are - and this is partly, if not wholly, a semantical or logical matter.

For example, unless we introduce and understand the mass of complex semantical and logical evidence which bears on the question of whether actions exist as a category of individuals, I do not see how we can even entertain saying many of the things the physicalist does say: I do not see how we can assert either that there are actions, or that actions are identical to physical events, or indeed any other physicalistic doctrine of this sort. And unless we understand the referential mechanisms of mental language, I do not see how we can say (except in the same eliminable sense in which we say that there are reasons for action, or so many miles between here and some other place) either that there are such things as mental events, or that mental events are (say) identifiable with physical events. An understanding of these mainly semantical questions is not, that is to say, something with which physicalism, if true, will provide us; for physicalism, whether true or false, presupposes this understanding.

That this general problem is important, and that the answers to it are not so far conclusive, can be seen from the bewildering number of suggestions which physicalists have actually made in the short modern history of the subject. This is not however to say that no progress has been made: Place's original 1956 view¹ that consciousness itself was

1. U. T. Place: Is Consciousness a Brain-process? British Journal of Psychology, 1956.

identical with a process or set of processes in the brain was replaced, or amended, by Smart, who suggested² that visual sensations, auditory sensations, tactual sensations, aches and pains, were all in their several ways identical with some brain process or other. But this idea of Smart's suffers from limitations of much the same sort as those which Place's theory suffers from. For while Place's suggestion has to confront the fact that consciousness is in no intelligible sense an item or individual, let alone an individual falling into the category of processes, Smart's own suggestion errs in supposing that individual aches and pains and sensations are themselves ontologically fit or appropriate for identification with physical processes in the brain. But they are not. The confusion is similar to that which is involved in supposing that thoughts, beliefs, desires and memories (etc.) are themselves appropriate for identification with physical events or states rather than that the events or states are which occur or obtain when people think that so-and-so, or believe or desire or remember that such-and-such. If a person thinks or believes that p, then in a loose and derivative sense we can speak of the thought which is thought (viz, the thought that p), or the belief which is believed (viz, the belief that p) - but it makes a big difference to the intelligibility of the theory itself whether it is the objects of these psychological attitudes or the fact (or the event) of their being had which are the items which it is the purpose of physicalism to analyse. There is nothing to suggest that accusative or intentional objects like beliefs or thoughts themselves exist in any clear or serious sense.

Something even worse in Smart's suggestion is the fact that the central class of mental phenomena, namely those reported in speech by means of a

2. J. J. C. Smart: Sensations and Brain Processes. Philosophical Review, 1959.

psychological verb followed by an embedded proposition, are not considered by him at all. What similarity is there to be found between whatever such expressions as "I ache", "I am in pain" etc., can be used to report and the goings on which propositional attitude sentences can be used to report?

The problem at issue here, as I explained, is to arrive at a clear understanding, may-be even by example, of what mental phenomena there actually are; and to do this via inquiries of a grammatical or semantical kind. Nagel eventually advanced a rather plausible-seeming adjustment to the views of Place and Smart just mentioned when he wrote:

"Instead of identifying thoughts, sensations, after-images and so forth with brain processes, I propose to identify a person's having the sensation with his body's being in a physical state or undergoing a physical process. Notice that both terms of this identity are of the same logical type, namely a subject's possessing a certain attribute"³

Now an attribute, for Nagel, is to be distinguished from what he calls a particular instance of an attribute. An attribute is "signified" by an open sentence with a free variable, and a particular instance of an attribute is "signified" by the gerundive noun-phrase which is obtained by filling the variable of an attribute-specification and nominalising. So in general an open sentence like "x ϕ " specifies an attribute, and a noun like "A's ϕ -ing" specifies an instance of that attribute. The open sentence "x is thinking about Vienna" specifies an attribute; and the noun-phrase "the stone's thinking about Vienna" specifies a particular instance of it. What the identity theory connects, for Nagel, are particular instances of attributes.⁴

We might want to ask whether the expression A's ϕ -ing names a different attribute-instance from the expression A's ϕ -ing at T, in the case where A

3. Nagel, Physicalism. In Borst (ed.) The Mind/Brain Identity Theory p. 216.

4. Nagel, op. cit. fn. 13., p. 220.

does happen to ϕ at T. Perhaps Nagel would say that either expression would do, providing the open sentence matches the gerund. But for our purposes it does not matter greatly whether Nagel would say that the Time-Specific Level (see Chapter I) is the proper level on which to assert an identity theory, or whether he would say that the Person-Specific Level is. For the factor of greatest importance concerns the general conditions under which particular instances of attributes are identical. No answer to this vital question is forthcoming from Nagel's own essay on the subject, and so this is another question which we must eventually explore for ourselves if we can.

There is a welter of "phenomena" whose ontological credentials we could discuss, and more than one way in which we could discuss them. Since my interest in this essay has focussed upon two categories only - the alleged category of events and the alleged category of states - it will be the evidence as it concerns these that I shall confine my attention to in this chapter. Moreover I shall divide the evidence into two kinds: one kind I shall call the logical evidence, and the other kind I shall call the grammatical evidence (though these are not independent of each other, as we shall see). In the sections which follow, I shall examine, in turn, the logical evidence as it effects events in general; the grammatical complexities of mental events; and finally, though somewhat tangentially to the main theme of the essay, some of the problems involved in the ontology of action.

2. EVENTS IN GENERAL: THE EVIDENCE FROM LOGIC

We must examine the logical evidence for the category of events. Others have discovered that the main task in doing so is, to give a preliminary description of it, to adjudicate the logical priority of

event-sentences over "categorical" existential sentences beginning "There is/was an event which is/was", where an event-sentence is conceived as a sentence which can be used to answer a question like "What happened?" or "What occurred?", or alternatively, as one which can be grammatically joined with another sentence of the same type by temporal phrases like "before" or "after". According to some, an event-sentence itself, whose logical form only contains a variable for a material object, displays the meaning of a categorical sentence of the sort which begins "there is/was an event which is/was"; while according to the converse view, the logic of event-sentences can only be given in terms of such a categorical sentence. The controversy as to the priority, in this logical sense, between event-sentences and categorical sentences, really boils down to this: given an event-sentence, do we need to employ a categorical sentence to display its logical form, or is it enough to rest with the assumption that its logic merely contains a material object variable? Is a sentence of the form "A ϕ 's" to be elucidated logically as merely having the form

"There is an object x such that x is A and x ϕ 's".

or as having the form

"There was an event e such that e was a ϕ -ing of A "?

Geach once opted for the first alternative. He proposed that for the most natural and primitive event-language,

"we need to get events expressed in a propositional style, rather than by using name-like phrases (what Karttunen has called onomatoids) any sentence in which an event is represented by a noun-phrase like "Queen Anne's death" appears to be easily replaceable by an equivalent one in which this onomatoid is paraphrased away; we could use instead a clause attaching some part of the verb "to die" to the subject "Queen Anne" Cutting our onomatoids in this

way, we get a manner of speaking in which persons and things are mentioned but events do not even appear to be mentioned .." 5

So for Geach, a sentence like "A ϕ -ed" is logically prior to the noun-phrase "A's ϕ -ing" in as much as the elucidation of the truth-conditions of "A ϕ -ed" do not depend on there being an event-noun to fill the blank in an allegedly equivalent categorial sentence of the form "there is/was an event which was".

Geach's answer to the problem of how a sentence is identified as being event-reporting in the first place was also in terms of whether it can be conjoined with other sentences with temporal sentence-conjuncts of the "before", "after", "happened at the same time as" variety, so that temporal relations between one sentence and another sentence, or between one sentence and a part of itself, are, for Geach too, the chief identifying characteristics of those which are potentially event-reporting.

One of the reasons Geach gave for his view was that

"nobody ever talked or is going to talk a language containing no names of people or things but only names of events, and the claim that our language could in principle be replaced by such a language is perfectly idle" 6

But obviously the question of how we decide to logically "regiment" parts of our language is only partially dependent on what sentences people actually use in speech (although I shall argue that it is dependent, in a sense, upon what sentences people could use); and the point about the replaceability of English by event-language seems to me not strictly pertinent either. None the less, I still believe that Geach's general view of the matter is defensible. Or if not defensible absolutely, then more defensible than its main rival, which is the doctrine espoused by

5. Geach: Some Problems about Time. In Studies in the Philosophy of Thought and Action (ed. Strawson) pp. 136-7.

6. Geach, op. cit., p. 136.

Davidson, Wallace and others that sentences of the sort which we have been considering are event-reporting sentences just because their logical analyses have to proceed in terms of an explicit reference to an event. I shall now describe this rival doctrine, and then attempt to show its defects.

The rival theory was originally expounded as a theory of the logical form of event-sentences, and was then extended to sentences of action - actions, according to Davidson, being a species of events. It was Davidson's view that in order to represent the logic of sentences of action, variables had to be introduced to range over events, with the consequence that verbs normally thought of as n-place predicates became predicates of n+1 places, with the extra place occupied by an event-variable.⁷ The original example was the sentence

(1) Shem kicked Shaun

which became regimented under logical analysis as the sentence "there is an event e such that e is a kicking of Shaun by Shem". No singular term picking out an event is used in this analysis (aside from the variable); although an indefinite singular term does make an appearance in the predicate.

Now although no definite singular term appears explicitly in the analysans sentence, the analysans sentence is none the less said to be true if and only if some individual event is such that it satisfies the open sentence "e is a kicking of Shaun by Shem". Trivially the description of this event is "the kicking of Shaun by Shem at t", where t is the time when the event took place. Or any other description would do, providing it satisfied the open sentence.

7. Davidson: The Logical Form of Action-sentences, pp. 81-95 (In The Logic of Decision and Action, ed. N. Rescher).

So Davidson's analysis proposes that the sentence (1) is true if and only if there was (is) an event which was (is) a kicking of Shaun by Shem. In opposition to Geach's view, a sentence with a verb of action is thus said to be logically posterior to the noun-phrase which picks out the event in which the action consists. There are reasons why I believe this analysis to be suspect as a general principle for the logic of event-sentences, and I must now explain, or try to, what they are.

But in order to carry out this explanation, I must introduce two hypotheses about the general nature of logical analysis. Both suggestions concern the constraints upon those things which some philosophers call logical regimentations of English sentences. I cannot completely satisfy myself that my suggestions on the nature of logical analysis are final or absolutely correct, and indeed the literature on this subject contains several well-argued proposals which are different from mine. None the less, I shall do the best I can to state what seem to me to be two minimal and basic constraints on the programme of the logical regimentation of English. Fundamentally the issue is nothing other than the nature of grammar.

The first constraint is not controversial, and needs little explanation. It simply says that the regimenting sentence and the regimented sentence say, in some sense of the words, the same thing. Moreover - and this is vital - this requirement must be satisfied otherwise than by purely stipulative or definitional means. Suppose S is the sentence to be regimented and S' is the regimenting sentence. Now if the requirement is satisfied, then S' means whatever S means. But the meaning of S' must clearly be specificifiable in some other way than by alluding to the meaning of S; for if this were not a constraint, the hypothesis that S' does regiment S would be unfalsifiable. (The point here is that the regimented and the regimenting sentence must, in the words of J. M. Hinton, "speak for themselves"). I call all this the same-saying requirement.

The second constraint is presupposed by the first, but it is one for which those philosophers who place heavy emphasis on the logical regimentation of ordinary English tend not to have a great deal of sympathy. The constraint is this: the regimentation of an English sentence, the thing which logically regiments an English sentence, must itself be either a grammatical English sentence, or a string of symbols for which mechanical rules exist for reading it as a grammatical sentence of English. I call this requirement the requirement of grammaticality. It can be illustrated in a fairly simple case, as follows.

Traditionally, logical regimentation transposes an English sentence into the terms of standard first order predicate logic, for this is a system within which an iterative explanation can be given of the truth conditions of sentences in terms of the satisfaction, by sequences, of the variables of quantification; and the point of the overall programme is to display in a uniform way the conditions under which sentences are true. So for an ambiguous sentence like

(2) All girls love a sailor

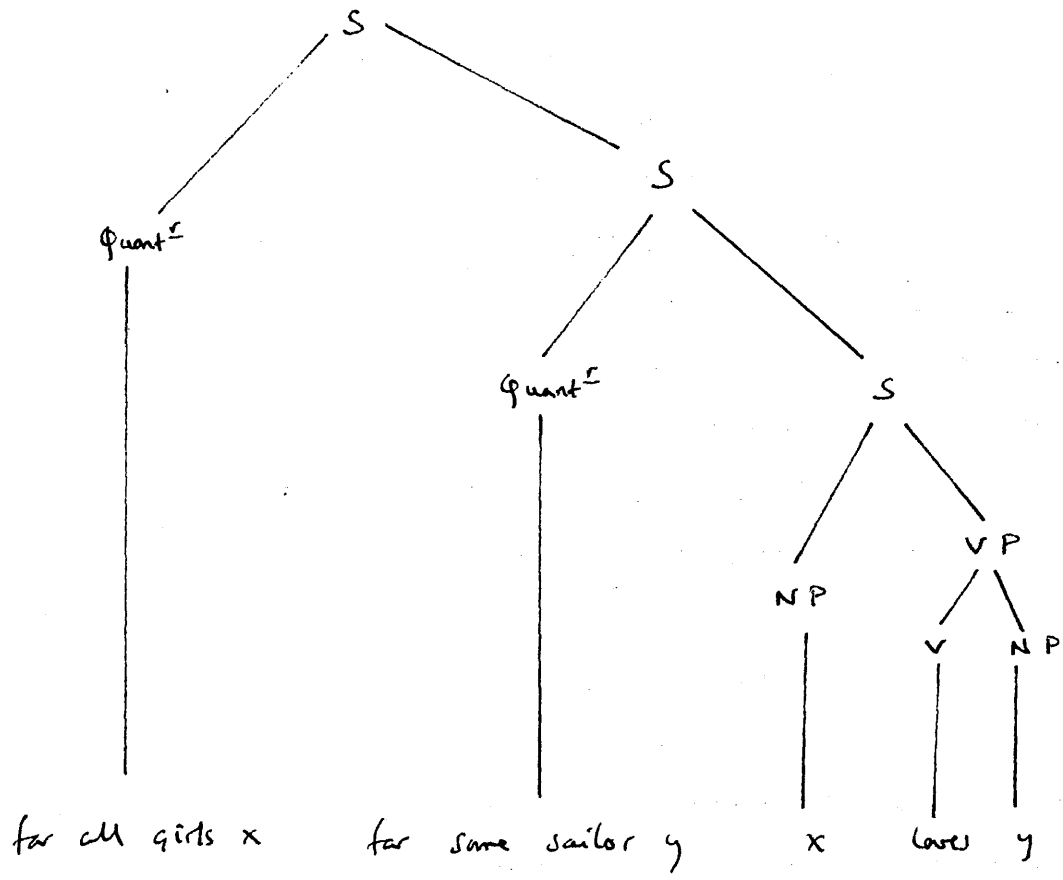
there are two logical regimentations, namely,

(3) $(x)((\exists y)(x \text{ is a girl. } y \text{ is a sailor. } x \text{ loves } y))$

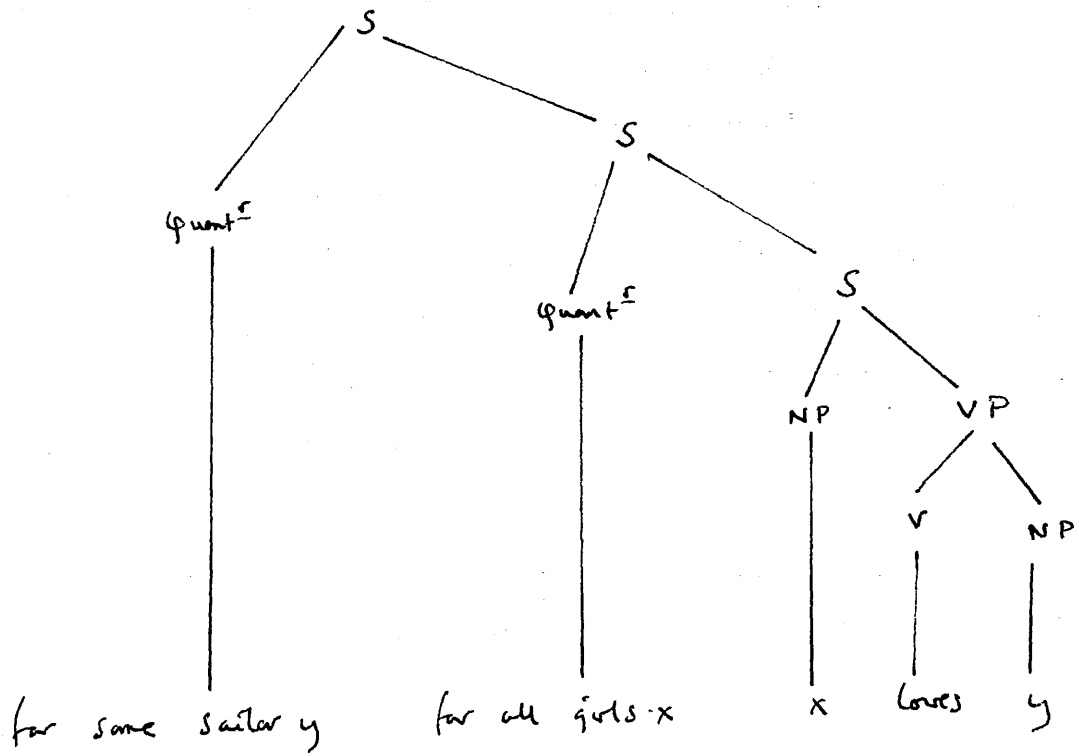
(4) $(\exists y)((x)(x \text{ is a girl. } y \text{ is a sailor. } x \text{ loves } y))$

where the bracketing conventions indicate scope, or, in grammatical terms, the domination of some sentence-elements by others. We could re-express (3) and (4) by tree-diagrams, thus:

(5)



(6)



Now of course a large part of the purpose of bracketing or tree-diagram conventions is to express syntactical relations within the sentence - this is an uncontroversial point which needs no emphasis. But what does need emphasis, and what indeed it is the point of the grammaticality requirement to emphasise, is that the string of terminal elements in (5) and (6), while not constituting sentences as they stand, can be so read as to constitute English sentences. The sentences are "All girls are such that there is a sailor such that she loves him" and "There is a sailor such that all girls are such that she loves him". These are also the English sentences which (3) and (4) represent; and mechanical rules exist for reading (3) and (4) as such.

My illustrative point is that if (3), (4), (5), and (6) did not represent sentences, then they would not be the sorts of things to which meanings or truth-conditions (in whatever terms) could attach; and if they are not the sorts of things to which meanings or truth-conditions can attach, then they cannot effectively regiment the ambiguous sentence "All girls love a sailor". For it is in the nature of a hypothesis that this sentence has certain truth-conditions (and so, like all hypotheses, it must be falsifiable). But this seems to entail that the diagram or string of symbols which is said to reveal the logical structure of some English sentence must itself be sentential in nature, for, as I said earlier, sentences are the only things for which truth-conditions can be given.

This might be thought to produce a paradox, or a circularity, in the programme of regimentation as a whole.⁸ If regimentations are sentences (call them regimentation-sentences), then all we have constructed to

8. Cf. the point made by John Searle in Chomsky's Revolution in Linguistics (New York Review of Books, June 29 1972), in connection with the notion of a semantic paraphrase derivable from a sentence's deep structure.

exhibit the logical structure of any English sentence is another English sentence; so that the problem arises again of the logical regimentation of this - and so on. However there is really no paradox or circularity here at all. The programme of logical analysis consists, in essence, of transposing each English sentence into a pre-logical or semi-logical but none the less sentential idiom which is itself part of English (or some language), and which consists of sentences directly accessible to a mechanical exposition of their truth-conditions; or which, we could say, wear their truth-conditions on their sleeves. For this set of sentences, couched in terms of phrases like "such that" and the apparatus of pronouns and/or variables, we have a direct and clear statement of truth-conditions in terms of the theory of truth offered by Tarski. But in the total project of showing how the meaning of a sentence depends on the meanings of its parts and on its syntax, at least in the programme which consists of doing this via the notion of truth, the logical regimentation of ordinary English sentences into a system of regimentation-sentences is only the preliminary step. Only once this has been completed comes the enterprise of connecting the truth-conditions of a sentence with its meaning.⁹ It is with reference to this programme that my requirement of grammaticality is to be defended against circularity.

Let me now return to Davidson's theory of the logic of event-sentences.

Davidson himself tells us that he

"dreams of a theory which makes the transition from the ordinary idiom to canonical notation purely mechanical, and a canonical notation rich enough to capture, in its dull and explicit way, every difference and connection legitimately considered the business of a theory of meaning" 10

9. See Strawson: Meaning and Truth for some preliminary suggestions.

10. Logical Form of Action Sentences, p. 115.

I agree with the requirement that the transition be mechanical, but I do not see how the canonical notation can capture anything unless its formulae can be read in grammatical English.

Now in his exposition of the logical form of "Shem kicked Shaun", it seems quite clear that Davidson fails to meet the requirement of grammaticality: for it is immediately obscure how the regimented version of "Shem kicked Shaun" should be read in a grammatical way. Here is the relevant passage from The Logical Form of Action Sentences:

"I suggest that we think of "kicked" as a three-place predicate, and that the sentence, i.e. Shem kicked Shaun" can be given this form:

LF₁ Ex. Kicked (Shem, Shaun, x)

If we try for an English sentence that directly reflects this form, we run into difficulties. "There is an event x such that x is a kicking of Shaun by Shem" is about the best I can do, but we must remember "a kicking" is not a singular term" 11

Davidson appears to treat the fact that there is no obvious English reading for LF₁ as a peripheral difficulty, whereas I should prefer to emphasise it. Looking briefly at the ways in which the deficiencies of LF₁ might be repaired will, I believe, lead us to my second point about Davidson's analysis.

One of the problems with LF₁ as it stands is obviously that it contains the word "kicked", whereas its English reading is said to contain the word "kicking". The presence of "kicked" seems to be a hangover from the allegedly inadequate analysis of "Shem kicked Shaun" as

LF₂ Kicked (Shem, Shaun)

11. Davidson, op. cit. p. 92.

where "kicked" is a two-place predicate, and where LF_2 is translatable into English in accordance with the following rule:

RR₁ For "G-ed(A,B)", take "A" as the subject, "G-ed" as the verb, and "B" as the object. In other words, rearrange and delete brackets, to get "A G-ed B".

In the case of LF_1 , the suggestion which obviously presents itself at this point is two-fold: that its word "kicked" be replaced by the word "kicking", and that the resultant version of LF_1 be associated with the following reading rule:

RR₂ Regard " $\exists x (G\text{-ing}(A,B,x))$ " as being always equivalent to the sentence "there is something which is a G-ing of B by A".

Now I mention this suggestion, not in order to appraise it immediately, but because it affords an opportunity to mention a set of problems which is purely grammatical in nature (and to which the next section of the essay will be devoted). For we might reasonably insist that a third important constraint on a regimentation is that its noun-phrases, if any, correspond in their classification to the entities, if any, which the original idiomatic sentence is "about". If the original sentence is about a material object, then any variable in the regimented sentence purporting to take the object as value must be replaceable by a material-object noun. And similarly, if the original sentence is about an event, or if it reports an event, or the occurrence of an event, then any variable in the regimented sentence purporting to take an event as value must be replaceable by an event-noun. (This is a requirement on which we are entitled to insist if logical variables are likened to pronouns).

Now the word "kicked" is not even a general term, and cannot be transformed into a noun-phrase by attaching a determiner. Replacing "kicked" by the noun "kicking" satisfies the evident demand for a noun phrase, but

it raises an additional question: is it a noun which can be used for reference to an event?

We have now arrived at a purely grammatical question, namely: is there a noun-phrase associated with the verb "kick" which we can use to refer to what took place when Shem kicked Shaun? There may be plenty of perfectly good nominal expressions which can describe what took place - but what they are is partly a factual matter, for it depends on what actually took place. Perhaps an act of aggression or an act of retribution is what took place when Shem kicked Shaun; but now we are using nominal expressions which identify acts, and which have no grammatical or logical link whatever with the verb "kick". I think we are now in a position to see what is wrong with the noun "kicking". The problem is that kicking is an activity, and it is an invariable feature of activity-words that the indefinite article cannot be prefixed. Activity-words are like mass nouns in this respect; the sentence "A kicking took place" is as grammatical as the sentence "A footwear was donned" is - which is not at all. Amending both so as to get "A spate of kicking took place" or "An article of footwear was donned" has the effect of turning nonsense into sense, but only at the price of introducing totally new nouns.

So it is certainly open to question whether there are such things as kickings. What I have said shows how much I doubt that there are, since it seems that on the best interpretation of the word "kicking", it cannot be grammatically preceded by the indefinite article, as Davidson's analysis would in this case require.

I have discussed the grammaticality requirement as it concerns the single example of the logic of "Shem kicked Shaun" in order to illustrate the kinds of considerations about sense, nonsense, interpretation etc., which the application of that requirement suggests. The same sorts of

considerations are, by implication, just as important to any other example. But before leaving the case of "Shem kicked Shaun", let me make one more remark about it in the same vein.

Perhaps it will be suggested that the difficulties surrounding the word "kicking" can be avoided by introducing the word "kick" in its place. This replacement would give a representation of (1) as

(7) E e. Kick (Shem, Shaun, e)

But even if this analysis satisfies the grammaticality requirement, it does not quite succeed in satisfying the same-saying requirement. For (7) reads "there was an event which was a kick of Shaun by Shem". But (1) says something subtly different, since the assertion that Shem kicked Shaun does not restrict Shem to having given Shaun just one kick. Two or more kicks might have been delivered.

The further we discuss such matters the more obvious it becomes that the choice of nouns in the regimentation of any sentence is an important one. The difficulty however in passing from particular cases to some general doctrine about this choice of nouns is aggravated, unfortunately, by the fact that English is not entirely systematic in the way its noun-phrases form; witness the analysis of

(8) The emperor died

Here, the right analysis would, I think, be

(9) There was an event which was a death of the emperor

which satisfies both the grammaticality as well as the same-saying requirement

despite the fact that people normally die only once. If we had failed to remember this we might have been tempted by "There was an event which was the death of the emperor". Another clearly mistaken analysis would be

(10) There was an event which was a dying of the emperor

It is mistaken because a dying is not anything we can easily make sense of at all (and a fortiori not as an event); and if what is meant is some dying, then the same-saying requirement lapses. (10) would not then say the same thing as what (8) says, for it does not follow from the fact that there was some dying that someone actually died. The death might have been prevented in the nick of time.

This example is therefore dissimilar to the kicking case, in as much as the language contains one appropriate noun. One suggestion in the general case is that when no appropriate noun is available, we are not entitled to postulate an event; for if making up new words was admissible, then all sorts of strange analyses could be proven. Another general point is that -ing words are not invariably suitable for event-reference; but this is not to say that nominals not of this form always are so suitable. As I shall argue in the next section, they are clearly not. But if we are to have sentences which assert the identity or difference of events, then it seems to me that ultimately we must have noun-phrases. The tenor of my grammatical explorations here has been to suggest, paradoxically, that for many event-sentences, we cannot say that there is an event which the sentence reports, in the strictly ontological sense of the phrase "there is"; and that this is due to the problems of specifying the event in a suitable way. There may be events like deaths, flashes, bangs and births, for which intransitive verbs are used in the event-sentence. But these, if they do exist, are about all there are.

Problems encountered with the grammaticality requirement are not the only problems which aggravate the analysis of sentences like (1). As Romane Clark has shown,¹² the Davidsonian analysis as applied to the problems of predicate-modifiers of almost every sort give bizarre results both from the point of view of grammar and from the point of view of same-saying. The sentence

(11) I flew my spaceship to the Morning Star

for instance, is represented as

(12) $(\exists e)(\text{Flew}(I, \text{my spaceship}, e) \ \& \ \text{To}(\text{the Morning Star}, e))$

which is said to read "there is an event e such that e is a flying of my space-ship by me, and e involved a motion towards the morning star". But something is clearly wrong here, since the truth-conditions of the last quoted sentence are not the same as the truth conditions of (11): not all objects towards which my space-ship moves are those towards which I fly it. And some analyses, to labour the point, are just incomprehensible without a prior understanding of the analysandum sentence; witness the sentence Davidson proposed for

(13) Brutus stabbed Caesar in the back in the Forum with a knife

namely "there exists an event that is a stabbing (of Caesar by Brutus event, it is an into the back of Caesar event, it took place in the Forum, and Brutus did it with a knife"¹³

12. Romane Clark: Concerning the Logic of Predicate Modifiers Nous 1970.

13. Davidson: On Events and Event-Descriptions (In Fact and Existence ed. Margolis) p. 84.

This and other examples (see Clark, op. cit.) suggest the analysis which depends on event-quantification to be wrong as a general strategy. This is not to deny that there are events, but it is to deny that a quantification over events is what underlies any and every event-sentence. One final point ought to clinch the dispute in Geach's favour. According to the view that the sentence-form

there was an event e which was a V-ing by P

is logically prior to the sentence-form

P V-ed

in the sense that the truth-conditions of the second sentence-form are explained by those of the first (instead of by some object's satisfying the predicate "...V-ed"), it must be the case that the variable e in the first sentence form is replaceable by an event-noun which designates the event which makes the relevant event-sentence true. Suppose that difficulties of the sort which I have been discussing were to be overcome, and that there was such a noun. Then the theory ought to be able to tell us under what conditions the noun in question does actually designate an event. But it seems that the theory cannot do this, without renouncing one of its premisses. For example, it cannot say that "the emperor's death" designates an event if and only if the emperor died, since this would involve renouncing the view that the sentence is logically posterior. But what else could the theory say? I share with Geach and Romane Clark the view that the sentence is logically prior.¹⁴

The logical evidence for events as a general and comprehensive category

14. Clark, op. cit., p. 319.

of individuals is, it seems, rather poor. This is not to say that there are no events, but simply that attempts to formalise event-sentences by introducing an explicit event-variable in very many cases fail to comply with the constraints which govern what qualifies as an adequate regimentation. The simple case of "Shem kicked Shaun", for instance, illustrates how, very often, the satisfaction of the grammaticality requirement will lead to a violation of the same-saying requirement: and there are all the problems about adverbial modification which have been aired exhaustively by others. The requirements of grammaticality and same-saying are stringent indeed; but the fact that they are so difficult to comply with establishes a presumption, at least, that Geach was right in supposing that events are most naturally reported by sentences; or, as he put it, that "we need to get events expressed in a propositional style rather than by using name-like phases".

I next adduce extra evidence in favour of Geach's view, but this time specifically as it concerns events which are mental. The arguments will proceed in a different way, although similarities will be noticed between them and some of the remarks made in the present section. But they converge upon a similar conclusion, which is that the assumption that there strictly are such things as mental events is extremely difficult to maintain.

3. THE ONTOLOGY OF THE MENTAL^{14a}

There are more ways than one in which to establish a presumption that there are objects of a certain category. One way is to discover where the logical variables fall in the formalisation of ordinary sentences into

14a. Much of the material in this section is to appear under the title "Mental Events: Are There Any?" in the December 1973 issue of the Australasian Journal of Philosophy.

the canonical terms of some chosen logic. But this is not the only way. A rather more basic, but no less intricate, strategy, is to examine what noun-phrases there are in the language, and, where they form systematically, to ask what kind of object they can be used to refer to.

The mere existence of a noun is not, of course, always evidence which is relevant to ontology. Nevertheless, reference to an item is effected in many simple cases by the use of nouns or noun-phrases, together of course with other particles; and so I think that ultimately it should seem unsurprising if some evidence relevant to ontology can be gathered from this quarter. A fact which is slowly emerging from the work of some linguists is that noun-phrases, although extremely varied when taken as a single group, do fall into a number of distinct grammatical categories. It would clearly be convenient to be able to reveal the differences amongst these categories, and the parallel differences, if any, amongst the object-categories which correspond to them. Our interest in this subject is naturally aroused by questions of the following kind: does the noun

John's thought

refer to an event or not, when appropriately prefixed to a predicate of the right type? And if it does, what does the noun

John's thinking

refer to? Or to repeat a familiar question: do we refer to an event or a process or something else when we use the noun

the kick which he received

Can we find any system in the way these noun-phrases form, or can nothing general be said about their relation to the verb? With these sample questions before us, let us investigate the ontological credentials of the language of "mental-events".

It is an almost universal assumption in modern discussions of the mind-body problem that there are such things as mental events, and that these comprise noticings, perceivings, assumings, rememberings, judgments and the like. It is true that arguments are sometimes advanced in support of this assumption, but in the main it is thought to be too obvious to need any support at all. All the same, I believe that a number of arguments can be given which suggest that this ontological assumption ought to be regarded with considerable suspicion.

D. C. Dennett is almost alone among modern philosophers in having recognised the great difficulties involved in arriving at a satisfactory view of what "mental entities" there really are. In the first chapter of Content and Consciousness¹⁵ his counsel was for caution. He recommended a policy of "tentatively fusing" mental sentences, so as not to "assume, from the start....that....certain sorts of parities exist between physical entity nouns and mental entity nouns" (p. 16) - a policy which he adopted in order to by-pass the difficulties of both the identity theory and its denial, and in order to allow himself to concentrate instead on the question of whether mental sentences as wholes can be "correlated in an explanatory way" with physical sentences as wholes. This replacement programme does, I think, contain difficulties of its own. But the overall conclusion of the arguments I advance in this section is that these difficulties ought to be faced, for the arguments in question point to the conclusion that the cautiously sceptical posture which Dennett adopted in relation to the

15. New York and London, 1969.

ontology of the mental not only deserves to be adopted, but deserves to be adopted with greater confidence than Dennett himself thought was justifiable.

My method of argument is as follows. Outside the ontology seminar we happily and unmisleadingly talk as if there were such things as miles, propositions, thoughts, voices, numbers, and a whole array of other things which, when the seminar is in progress, we find it appropriate to doubt the existence of. Quine's slogan "to be is to be the value of a logical variable" and its variants, whatever its actual merits, is very often used in such situations as the tool or instrument of doubt. My strategy in relation to the specific question of whether there strictly are such things as mental events will be to employ three different, but related, devices. The first device consists in examining whether the English language contains noun-phrase constructions of a sort which are appropriate, in a sense I shall explain, for specifying mental events. The second consists in examining how, if there are such noun-phrase constructions, their referents are modified or qualified in actual speech. And the third device consists in examining whether there are any sentences in which pronouns or pro-nominal constructions appear in such a way as to either replace such noun-phrases or else to repeat a reference originally made by one of them.

Each of these devices is obviously grammatical in nature. The reason why each of them is appropriate to questions of an ontological kind derives from the fact that the very notion of an individual or range of individuals which cannot be described or nominated or specified in any systematic way at all is, on the face of it, just incoherent. The fact that there are such things as unspecifiable numbers, assuming that it is a fact, has no real power to deflate this principle. Ontology as applied to mathematics is in a notoriously obscure state in comparison with ontology elsewhere. And the general proposition that all things of the number kind are unspecifiable would, for common-sense reasons, be hard to make sense of, unless it

was understood as an obscure way of saying that, strictly speaking, there are no such things.

Before turning to the grammar of mental sentences, let me firstly re-explain which mental sentences I shall be concerned with, and let me re-phrase the basic question which I shall be asking about them. We roughly demarcate event-sentences as those sentences which can be grammatically joined with each-other by temporal phrases like "before" or "after"; or alternatively as those sentences which can be used to make a statement in answer to questions like "What happened?" or "What occurred?". And we then re-phrase our basic ontological question by asking whether, for any or for every true mental event-sentence, there exists a way of specifying which event is reported by the sentence as having taken place. By the phrase "specify which event took place" I do not have in mind the type of specification which is achieved by naming an event (thus: there occurred a certain event, namely E), nor the type of specification in which an event is specified as uniquely having certain properties (thus: there occurred the event which is the ϕ), but the type in which an event is assigned to a kind (thus: there occurred an event of the F kind). The question now before us is therefore whether, for any or for every true mental event-sentence, there exists a way of specifying which kind the reported event belongs to.

Let us take as an example of a non-mental event-sentence the sentence

(14) The star exploded

and let us take as an example of a mental event-sentence the sentence

(15) John remembered that the ship was sinking

Evidence for the view that sentence (14) reports an event when used to make

a true statement is that there exists a noun, namely "explosion", which can be predicated of an event in such a way as to specify which kind of event took place. There is a natural conversion of sentence (14) which reads:

(16) An explosion of the star occurred

Here, it seems, we have a natural relation between an event-sentence and an event-noun. Moreover, this natural relation is undisturbed when an adverb of manner - to take just one type of adverb - attaches to the event-sentence itself, as in

(17) The star exploded violently

for here we can change the adverb to an adjective and attach it in predicative position to the event-noun which appeared in the conversion of (14) to (16):

(18) A violent explosion of the star occurred

The existence of such sentences as (16) and (18), and of their natural relations to sentences (14) and (17) respectively, is what strongly suggests that sentences (14) and (17) can be used to report an event; and, when they are so used, that there is an event which each sentence respectively can be used to report, namely, in this case, an event of the type explosion and an event of the type violent explosion.

This ontological conclusion is also strongly suggested by the existence of sentences in which pronouns occur, for example:

(19) The star exploded, and it occurred at noon

(20) The star exploded violently, and we know what caused it.

where the plausible way of explaining the function of the pronoun "it", in each case, consists in supplying a description of the event itself. The question "What occurred at noon?" is answered by "An explosion of the star". The question "What do we know the cause of?" is answered by "A violent explosion of the star". If the pronouns in (19) and (20) do refer, then it seems clear that an event is the right type of individual for them to refer to.

These grammatical patterns can be found in many, but not in all, event-sentences which are non-mental; and where they can be found they support the appropriate ontological conclusion. But the interesting fact is that none of these patterns apply to event-sentences which are mental. Consider sentence (15). It might be said that for the verb "remember" we have the noun "memory"; but for the larger noun-phrase which stands to (15) in the relation in which (16) stands to (14), what do we have? We might try the noun-phrase

(21) A memory that the ship was sinking by (of?) John

but this, if we can make sense of it at all, like

(22) John's memory that the ship was sinking

surely specifies the content of what John remembered, rather than any event. A similar interpretation has to be offered for the noun "perception", in relation to the verb "perceive"; while for other mental verbs there is no noun of this particular grammatical type - Chomsky calls them "derived nominals"¹⁶ - at all. (The verb "realise" has the noun "realisation",

(Footnote 16 printed overleaf)

"assume" has "assumption", and "judge" has "judgement"; but what do the verbs "notice", "see" or "learn" have?).

Our task is to discover how to specify the sort of event, if there is one, which would be reported by saying truly that (for example) John remembered that the ship was sinking. The derived nominal in this case specifies the wrong thing. Nor can we specify the type of event in question as "that type of event which is reported by saying truly that John remembered that the ship was sinking", because this answer merely prompts the "what type?" question all over again. Nor, of course, can we specify it as "that type of event which took place when, or as, John remembered that the ship was sinking", for this lets in any physical event which happened to take place at the same time.

How then can appropriate specification be achieved? It might be suggested that our efforts to systematically specify mental events with derived nominals like "memory", "perception" etc., were misguided from the beginning, for the reason, firstly, that derived nominals in English have extremely irregular formation-patterns right across the board, for mental

(Footnote 16 brought forward from page 126)

Remarks on Nominalization. pp. 134-221 in Jacobs and Rosenbaum (eds.): Readings in English Transformational Grammar.

Apart from the derived nominals, Chomsky describes the formation of "mixed" and "gerundive" nominals. For a sentence like

John refused the offer

to take one example, the gerundive, derived and mixed nominals from respectively as follows:

John's refusing the offer

John's refusal of the offer

John's refusing of the offer

as well as for non-mental verbs,¹⁷ and that secondly, there is a different type of noun, namely the gerundive noun, which forms in a perfectly regular way and which we might just as well use instead. There are several remarks to be made about this suggestion. In the first place the apparent irregularity of derived nominal formation might turn out to have a systematic explanation when investigated fully. It seems to be a matter of controversy among linguists as to how likely this is.¹⁸ Secondly, and more importantly, the positive suggestion that the gerundive nominal be used to specify events is, in point of fact, extremely important to assess; for several philosophers, following Nagel's precept in Physicalism,¹⁹ have suggested that an identity theory of the mental and the physical can be best formulated by using nouns of just this type.

Gerundive nouns are those which form from any verb by the addition of

-
17. The nominals in the following list show clearly that a blanket interpretation cannot seriously be attempted (cf. Chomsky p. 189)

laughter	(v: laugh)
marriage	(v: marry)
construction	(v: construct)
belief	(v: believe)
conversion	(v: convert)
qualification	(v: qualify)
house	(v: house)
box	(v: box)
conveyance	(v: convey)

18. See Chomsky, op. cit., pp. 187-9:

"the semantic interpretation of a gerundive nominal is straightforwardly in terms of the grammatical relations of the underlying proposition in the deep structure.

Derived nominals are very different in all (of these) respects. Productivity is much more restricted, the semantic relations between the associated proposition and the derived nominal are quite varied and idiosyncratic, and the nominal has the internal structure of the noun-phrase. (These matters) raise the question of whether the derived nominals are, in fact, transformationally related to the associated proposition"

See also Chomsky, loc. cit., footnote 11.

19. Philosophical Review 1965

"ing".²⁰ But it is clear that the type of specification which they afford is different from the type discussed so far. The gerundive noun related to an event-sentence appears not to specify a type of event, but to provide a kind of description of the event itself, with some implication of uniqueness; and I suspect that those who advocate the gerundive strategy would say that since gerundive nouns always form, an event-description can be formed from any event-sentence; so that for any true event-sentence (mental or physical) we can specify the reported event by using a gerundive description, and saying that an event having that description occurred. For instance, instead of converting sentence (14) to read "An explosion of the star occurred", we could convert it to read:

(23) The star's exploding took place (occurred, happened, etc.)

It is certainly an advantage of this style of event-specification that gerundive nouns can be formed in such a regular and mechanical way. However, there is an interesting and peculiar set of facts about the gerundive construction itself which suggests, I think, that the referential function of the gerundive noun is distinctly unusual, and that the type of specification

20. That is, it seems to be an exceptionless general truth about our language that any sentence of the active intransitive form

S V-ed

can be associated, via a simple transformation, with a noun-phrase of the form

S's V-ing

and that any active transitive sentence of the form

S V-ed P

can be associated, via a similar transformation, with a noun-phrase of the form

S's V-ing P

which it can be used to achieve is not event-specification at all. These facts concern its internal structure. If we compare the internal structure of the gerund with the internal structure of the complex possessive noun ("John's hat", and the like), we find differences which I think can only mean that the gerund fails to operate as a mechanism of reference in quite the same relatively straightforward way in which the complex possessive noun does. The important difference can be summarised by saying that while a complex possessive noun of the form P's F has an internal structure which can be represented as the F which is P's or the F which belongs to P (where the qualified phrase the F serves to identify the kind of thing eventually designated), the gerundive noun does not.

Let me now explain this difference more fully.²¹ While a gerundive noun has the superficial appearance of a complex possessive noun, its underlying structure cannot be the same. Any sentence containing a complex possessive noun can be re-phrased without loss of meaning with the possessive noun unravelled, so that for instance

(24) John's hat is dented

can be re-phrased as

(25) The hat which is John's is dented

In addition - this is a second feature of possessive nouns - any sentence of this latter form can usually be existentially generalised, so that from (25) we can infer

21. See Chomsky, loc. cit., p. 138ff.

(26) A hat which is John's is dented

By contrast, your average gerundive noun will not unravell; and nor, as a consequence of this, will the sentence which would be said to unravell it submit to existential generalisation, in the manner in which (25) did to yield (26). Both these failures beset mental as well as physical gerunds. Take the sentences:

(27) (The star's exploding) G

(28) (John's remembering that the ship was sinking) H

where "G" and "H" are predicates. We cannot, with any confidence, rephrase these sentences to give:

(29*) (The exploding which is the star's) G

(30*) (The remembering that the ship was sinking which is John's) H

for very real and reasonable doubts can be entertained as to whether these sentences are well-formed; and the same goes for the sentences which would be said to follow by existential generalisation from these, namely:

(31*) (An exploding which is the star's) G

(32*) (A remembering that the ship was sinking which is John's) H

I think we must say that these last four sentences, unless invested with meaning by pure stipulation, are literally speaking deviant.²² Perhaps

(Footnote 22 printed overleaf)

their deviant nature can best be appreciated by looking at parallel cases which contain the gerundives corresponding to the verbs "to be" and "to have". If we do not count

(33*) The being in pain which is John's

(34*) A being in pain which is John's

(35*) The having of a toothache which is John's

(36*) A having of a toothache which is John's

as deviant, it is difficult to see what deviance is.

I shall later suggest what conclusions should be drawn from the fact that complex possessive nouns have a different internal structure from the gerund, and what the referential function of the gerund exactly is. I want now to examine the two other grammatical phenomena which I suggested were relevant to the ontological question about mental events. We saw that there were patterns of adverbial modification for non-mental event-sentences in which the thing modified was some event; but when we consider how adverbs attach to mental sentences, it is hard to escape the conclusion that the thing modified is always something other than an event. It is

Footnote 22 brought forward from page 131.

This fact would have to be ignored by anyone who thought that the logical form of "John remembered that the ship was afloat" contained a logical variable for an event; for the regimenting sentence would be "There was an event which was a remembering that the ship was afloat of John's", which I would maintain is deviant. (Cf. previous section). And again, if it was stipulated that it had the same meaning as the regimented sentence, then of course the hypothesis that the second sentence does regiment the first would be unfalsifiable.

difficult to argue decisively about a subject which is still being explored, but the main ways in which adverbs invade mental event-sentences can, I think, be tentatively laid out as follows.

The first pattern is where the modifier in fact modified the agent, so that a sentence of the form²³

P Adj-ly V-ed that p

can be explained as having the form

It was Adj of P to V that p

(or perhaps: It was Adj of P that P V-ed p). This is the pattern which adverbs like "perceptively" and "thoughtfully" fit, together with "intentionally", "deliberately", "mistakenly" or "clumsily", as these occur in sentences of intentional action (supposing these sentences to be mental). It is obvious that if John noticed perceptively that the ship was sinking, the item which "perceptive" qualifies is not some event, but John himself. The point here is not that "John was perceptive" is entailed, for as a remark about John-in-general this might well be false - but that "It was perceptive of John to notice that the ship was sinking" is.

The second main pattern is that whereby

P Adj-ly V-ed that p

must be explained as having the form

23. Here I use the shorthand of Adj for an adjective, Adj-ly for an adverb, V for a verb, and P for a person-name.

It was Adj that P V-ed that p

(or perhaps: That P V-ed that p was Adj). This is the pattern which pertains to adverbs employed in a "factive" sense. "Alarmingly", "surprisingly", "predictably", "annoyingly" etc., can have this role.²⁴ If John surprisingly noticed that the ship was sinking, then it was the fact that he noticed what he did that was surprising, and not some noticing-event.

A third pattern is exemplified by sentences like

(37) He remembered $\left. \begin{array}{l} \text{vividly} \\ \text{vaguely} \\ \text{dimly} \end{array} \right\}$ that the building collapsed

but here the item which was vivid or vague or dim was not some remembering-event, but the memory itself.

There are other adverb patterns apart from these which can be interestingly investigated in the same ground-level way. No special expertise in the sciences of linguistics or logical form is needed to see that none of the sentence-patterns in which an adverb occurs in a mental event-sentence is such as to suggest that underlying the sentence itself is an event-description which the related adjective has the force of qualifying. The moral which emerges from these investigations is that although we can and do say such things as that John suddenly or surprisingly or perceptively or vaguely or predictably remembered (or noticed, etc.) that such and such is the case, it is wrong to infer that there must be

24. The example of "surprisingly" shows that one and the same adverb can on different occasions have different roles. "It was surprising of John to notice that the ship was sinking" (pattern one) means something different from "It was surprising that John noticed that the ship was sinking" (pattern two).

events of a mental sort on the grounds that the adverbs in question have the role of saying how they occurred, or what they were like. The most cursory examination of the meanings of the relevant sentences shows that they do not.

Let us finally consider whether any conclusion can be drawn from the most ordinary patterns of pronominal occurrence in mental event-sentences. Pronouns, it seems, very often have a referential function. Very often we can expand a sentence which already contains a referring phrase in such a way that a pronoun occurring in the added part repeats a reference made in the original. Indeed the role played by a pronoun in the expansion of a given sentence can often be helpful in deciding just what phrases of the non-expanded sentence are being used referentially. This is not always the case, for while it seems to hold for a sentence like

(38) I picked up my hat and put it on the peg

where there is little doubt that "it" refers to whatever "my hat" refers to, it cannot be said to hold for a sentence like

(39) John shut the window upstairs and Peter did it downstairs

where it seems that "it" has the function of replacing the phrase "something of the kind which John did" rather than referring directly to what John did. But despite these variations, there is still an argument which is of relevance to our ontological question about mental events, for we can surely say that if there strictly are such things as mental events, then there will exist possible expansions of mental event-sentences that contain backwardly referring pronouns. The existence of such expansions would support the ontological hypothesis that mental events occur, while the absence of

such expansions would support a contrary hypothesis.

In fact there seem to be no convincingly grammatical examples in which a pronoun in an expanded mental event-sentence does have this event-referring role. The nearest we can get is, I think, a sentence like

(40) John noticed that the ship was sinking, and it surprised (shocked, startled) us

However, the pronoun "it" in this example, if it backwardly refers to anything at all, can only be taken to refer to a fact, namely the fact that John noticed that the ship was sinking.

I shall say more about the difference between facts and events later. The negative conclusion that has emerged up to this point is this: not one of the three types of grammatical phenomena which I have considered provides any evidence for the supposition that there is such an ontological category as the category of mental events. Indeed the evidence can be taken to suggest the opposite. Admittedly, the case has been argued for the category of propositional mental events - those purportedly expressed by a sentence of propositional attitude - but it would be surprising if similar arguments could not be extended to mental sentences not of the propositional attitude form.

Before closing this section I must emphasise again that nothing I have said up to this point suggests that there are no events at all. It is obvious that there are; for apart from explosions, which I mentioned and recognised earlier, it seems pointless to deny that there are also events which are bangs, flashes, collapses, marriages, deaths, births and so forth.²⁵ The conclusion to which the grammatical evidence points is not that there are no events, but that not every sentence properly described

25. Cf. Strawson: *Individuals*, Ch. I.

as an event-sentence is such that there is an event, or type of event, which, when used to make a true statement, it reports. The argument has been that mental event-sentences which are of the propositional attitude form are of this type. For these, Geach's view that events must be expressed in a propositional style rather than in nouns holds good.

Probably the most interesting and important group of grammatical facts that has come to light so far concerns the gerundive noun. Its interest lies in its un-noun-like structure; its importance derives from the use to which it has been put in recent discussions of the mind-body problem. Nagel once tried to make use of the fact that every English sentence has a gerundive noun corresponding to it to avoid an objection to the theory that thoughts, beliefs, pains and sensations could be identified with physical processes in the brain or central nervous system. The objection was that physical processes or activities in the brain had spatial location, whereas neither thoughts nor beliefs nor pains nor sensations did. In recognition of the importance of Nagel's ingenious way out of the difficulty I quote his statement for a second time:

"Instead of identifying thoughts, sensations, after-images, and so forth with brain processes, I propose to identify a person's having the sensation with his body's being in a physical state or undergoing a physical process. Notice that both terms of this identity are of the same logical type, ... namely, ... a subject's possessing a certain attribute" 26

(Where in general an open sentence like " $x\phi$ " specifies an attribute, while a gerundive noun like "A's ϕ -ing" specifies an instance of that attribute).

But I need hardly re-emphasise how important it now becomes for Nagel's approach to physicalism to contrive identity-conditions for instances of attributes. When is A's ϕ -ing identical with B's ψ -ing, and when

26. Nagel, Physicalism (In C. V. Borst, ed.: The Mind/Brain Identity Theory) p. 216.

distinct? There is one possible answer to this question whose defects I think it is instructive to have clearly before us. For it shows what is wrong with Nagel's proposals for interpreting the gerund, and thereby leads us to see what entities his version of the identity theory really concerns. Suppose we were to say²⁷ that particular attribute-instance A's ϕ -ing is identical with particular attribute-instance B's ψ -ing (where ϕ -ing is a mental word and ψ -ing is a physical word) if and only if both

$$(i) A = B$$

and (ii) ϕ -ing = ψ -ing

or, perhaps even more fundamentally, that particular attribute-instance A's ϕ -ing at T_1 is identical with particular attribute-instance B's ψ -ing at T_2 if and only if

$$(i) A = B$$

$$(ii) \phi$$
-ing = ψ -ing

$$(iii) T_1 = T_2$$

Now there are a number of things wrong with an answer in these terms. The first is that it makes the identity of particular attribute-instances dependent upon the identity of attributes themselves (clauses (ii)); whereas the theory which seeks to identify each instance of Tom's being in some psychological condition with an instance of Tom's being in some physical

27. An answer very similar to the one I criticise was given by J. Kim (On the Psycho-Physical Identity Theory, American Philosophical Quarterly 1966), who suggested that an event a's being F and an event b's being G are the same event iff either a is F and b is G are logically equivalent, or else iff $a = b$ and the property of being F (F-ness) = the property of being G (G-ness). My criticisms do not pertain to the logical equivalence condition.

condition is supposed, and rightly, to be less ambitious than a theory which seeks to identify each psychological condition with some physical condition. To take the case of pain: it is rightly supposed to be easier, or less ambitious, to establish the truth of statements like

(iv) P's being in pain at t = P's ψ -ing at t

for some P and for some t, than it is to establish the truth of statements like

(v) P's being in pain = P's ψ -ing

for some P; and easier in turn than to establish that

(vi) being in pain = ψ -ing

But if the identity of attribute-instances is made to depend upon the identity of attributes themselves, then the order of difficulty is reversed.

We must insist in other words that

(vii) Tom's being in pain = Tom's ψ -ing

does not entail that

(vi) being in pain = ψ -ing

on the grounds that, if it did, then the truth of

(viii) Fred's being in pain = Fred's $\{$ -ing

(where $\{$ -ing is different from ψ -ing) would entail that

(ix) being in pain = $\{$ -ing

which would conflict with (vi). This shows that conditions (i) and (ii) above are not necessary for the identity of A's ϕ -ing with B's ψ -ing.

An additional difficulty is that it makes many events which are presumably identical into distinct occurrences. Presumably Caesar's death is the same event as Caesar's assassination; but these can only be the same event, by Kim's standard, if and only if Caesar = Caesar and dying = being assassinated. But clearly the latter statement of identity is false.

Counter-examples can be multiplied with ease. On the assumption that Caesar died at noon, and given that Caesar died only once, it would be surprising if "Caesar died" and "Caesar died at noon" did not describe the same event. But this is only possible on Kim's analysis if, apart from it being true that Caesar = Caesar, it is also true that dying = dying at noon. Since this cannot be the case, either the two quoted sentences must describe different events (counter-intuitively) or else Kim's analysis must be wrong.

Evidently there are grave difficulties in devising identity-conditions for attribute-instances. The trouble is, I believe, that the whole terminology of attributes and attribute-instances is a mistake. It leads us to expect that attributes and attribute-instances are related in a certain way, and yet when we come to examine the matter we find that they are not. But there does fortunately exist a way in which this philosophy of attributes can be corrected.

Nagel's terminology of attributes and attribute-instances introduces an interpretation of the gerundive noun. Whereas those who have the conviction that events are ubiquitous, on the other hand, would be inclined to say that a phrase of the form "John's noticing that the ship was sinking" specified an instance of an event-attribute, or alternatively a particular event. Either of these interpretative views would be simply as good as

each other, or as any, in the absence of any positive clues as to how these nouns do in fact operate; but there are positive clues, which both Nagel and the event-theorists seem to have over-looked. These lead to a quite different view of the role of the gerund.

In the first place, we have the clue unearthed by Chomsky that gerundive nouns have an internal structure which is strikingly different from that of the possessive noun (although their superficial appearance is the same); and secondly, there is a certain amount of evidence that the correct interpretation of the gerund is factive. Gerundive phrases like

John's being in pain

John's believing that p

John's wanting Q

can, when in subject position, be freely interchanged with the corresponding phrases

that John is in pain

that John believes that p

that John wants Q

which in turn can be freely interchanged with the corresponding phrases

the fact that John is in pain

the fact that John believes that p

the fact that John wants Q²³

Footnote 23 printed overleaf

It is not an objection to this view that

(41) His singing was annoying

can be said to have a different meaning and a quite different set of truth-conditions, from

(42) The fact that he sang was annoying

For what we have as the subject term of (41) is a phrase which is in fact ambiguous: it can either be the gerund from "He sang", in which case (41) and (42) do share their meanings and truth-conditions, or else it can be the mixed nominal from "He sang", in which case it refers to something like a process or activity, so that (41) records something equivalent to

(43) His actual singing was annoying²⁹

Footnote 23 brought forward from page 141

But this interchange cannot take place where the phrases occur as grammatical objects, as in

He that p

and He the fact that p

Here there is non-equivalence when "...." is filled by explained, said, thought, etc.

But it is a distinctive and revealing mark of the gerundive nominal that its sentence-forming complement can always be prefixed to the sentence from which the nominal is derived, with the introduction of a "that". Thus:

- (i) John's riding his bicycle bothered her
- (ii) It bothered her that John rode his bicycle
- (iii) His working out the problem astonished us
- (iv) It astonished us that he worked out the problem

etc. Cf. Bruce Fraser: Some Remarks on the Action-Nominalisation in English (in Jacobs and Rosenbaum pp. 84ff)

29. Distinct of course from (i) His actually singing was annoying

It is only (41) in this second sense that makes it distinct from (42): but the ambiguity in phrases like "His ϕ -ing" is only to be expected where the main verb of the sentence has no grammatical object, for it is the relation of the grammatical object to the subject which standardly enables us to distinguish the mixed nominal from the gerund. In examples in which the gerundive and mixed forms do diverge, as in

(44) His singing the song

(45) His singing of the song

which both of course come from

(46) He sings the song

it is clear that factively understood adjectives do attach to the gerund, but not to the mixed nominal. Further examples of adjectives normally used in a factive sense are "pleasing", "annoying", "surprising", "predictable", "strange"; while further examples of adjectives normally understood in a non-factive sense are "rapid", "sudden", and "perceptive".

An additional morsel of evidence for the view that gerunds refer to facts is that in those cases in which the grammatical distinction between gerund, derived and mixed nominal is clear (i.e. where ambiguity doesn't arise), the sentence-forming complements attachable to nominals other than gerunds are quite clearly of a sort which, from their meaning, are true of entities other than facts. We cannot predicate of the two nouns

(47) John's refusal of the offer

(48) John's refusing to the offer

a complement having a factive meaning. To take two instances only, we naturally say things like

(49) John's refusal of the offer was the second of the day

(50) John's refusing of the offer was a tedious process

So it is not true simply that gerunds do take factive predicates; it is also true that nominals other than gerunds only take predicates of a non-factive kind, or which have a non-factive interpretation.³⁰

30. The arguments for the non-existence of mental events can be extended without difficulty to answer the question of whether mental entities of other categories exist. I shall now briefly indicate how this is done for the category of mental states. Sentences which are alleged to describe mental states are of the following sort, propositional and non-propositional, respectively:

- (i) John believes that spring is already here
- (ii) Tom wants his tea

for which the gerundives respectively are

- (iii) John's believing that spring is already here
- (iv) Tom's wanting his tea

Again, it seems right to interpret these nominals as fact-referring nominals, appropriate, as before, to complementation by such predicates as astounded us, surprised his brother, and the like, which can be construed as sentence-modifiers and prefixed accordingly. The mixed nominals for the sentences (i) and (ii) however, are clearly not acceptable (*John's believing of the fact that spring is already here, *Tom's wanting of his tea). And as for the derived nominals:

- (v) John's belief that spring is already here
- (vi) *Tom's want (of?) his tea

we have, in (v), a noun which specifies only the content of what John believed; and in (vi)* merely a nonsense-phrase.

Now it may well be the case that the un-noun-like structure of gerundive nouns is actually connected with their factive sense; for there is more than a hint of complementarity between the idea that gerunds are unlike nouns, on the one hand, and the idea that facts are unlike things, on the other. (A fact, after all, can be minimally described as what a sentence, when used to make a true statement, is used to state). But where does this leave the efforts of Nagel and the event-theorists to construct an identity theory of the mental with the physical?

It leaves them in a very peculiar position. If gerundive nouns can only be used to specify facts, and if facts are merely what true statements state, then by nominalising a sentence to obtain a gerundive noun in Nagel's style we achieve nothing more than a way of saying, at best, such things as the following. Every mental fact is identical with a physical fact; or: some mental facts are identical with physical facts; or: no mental facts are identical with physical facts, and so forth. A physicalist can say such things if he chooses to. But in doing so, he will only be saying, in a roundabout but none the less permissible way, that true mental sentences and true physical sentences, taken in their appropriate pairs, bear a certain kind of (presumably strong) relationship to each other. As Dennett recognised,³¹ the point of identity theory would then be substantially lost.

To summarise this argument: there is little grammatical evidence for supposing that there is such an ontological category as the category of mental events. The favourite ways of trying to specify events involve the use of the gerundive noun, but it seems that the only things which these can be used to specify are facts. Allowing that there are mental facts while suggesting that there are no mental events is not, it must be

31. Content and Consciousness, footnote to p. 17.

emphasised, merely a matter of returning with one hand what was taken away with the other; because facts cannot be events, and nor can particular events be facts (although it may of course be a fact that a particular event occurred). If the arguments here expounded are correct, then the conclusion must be that a physicalist who hankers after an identity theory has no option but to turn his attention to mental facts, or, if he chooses, to true mental sentences. These seem to be the only two coherent alternatives he has; although he will do well to remember that the completely ubiquitous connection between the sentence and the gerundive noun in English shows that, between them, there is nothing to choose.

4. ACTIONS AND CAUSES AGAIN: SOME REMARKS ON THEIR GRAMMAR

Finally I append some considerations of a similar sort about actions and causes. The question to which I shall address myself is whether there are actions, in any sense other than that which is implied by the existence of sentences of action. In fact my aim will be to express doubts as to whether there are. Some of the considerations relevant to this question have been aired in previous sections, so rather than involve myself in repetition, I shall proceed at once to discuss what seems to me to be one crucial question, which can be expressed like this: does it follow from what is called the causal analysis of actions either that actions are events, or that there are such things as actions? To probe this problem we naturally need to satisfy ourselves of the plausibility of the causal analysis of actions, and having done that, we need to see what its implications are.

What I mean by the causal analysis of actions can be best expressed by saying that action-verbs are causal verbs, or, roughly speaking, that verbs of action can be analysed grammatically as having the form cause S, where S

is a sentence or sentence-like element. A causal analysis of actions has been implicitly expressed in the philosophical literature (notably by Davidson)^{31a} as saying that for every action except a basic action, we can consider it as something which causes so-and-so to happen.

This doctrine goes hand in hand with a certain view of the identity of actions. In order to see this, consider the following case. Suppose that there is a physical event describable as the movement of my right arm in a downwards direction (at a certain time); and that in the situation there are various other sentences which could be given in answer to the question "what did Taylor do?". Here are just a few:

- D1. He pressed the plunger down
- D2. He caused the bridge to blow up
- D3. He intentionally caused the bridge to blow up
- D4. He blew up the bridge
- D5. He sprained his wrist
- D6. He damaged the plunger mechanism (by pressing too hard)
- D7. He killed the sentry who was on duty on the bridge

With the exception for the moment of D3, we can suppose that each of these sentences describes or reports something that happened. Davidson, I think, would assert that each sentence describes the same event; the event specified, perhaps, by "my right arm's moving downwards". He would also assert that what makes these descriptions descriptions of an event which is an action, is the presence in the list of the description D3., for according to him an event is an action when it has a description in terms of the agent's intention.

31a. D. Davidson: Agency

But it might appear that the way in which causality enters this action-situation in fact contradicts the view which says that all the D-sentences specify the same event. For instance, it might be said that D1. and D4. must apply to, or report different events, on the grounds that (i) pressing the plunger down occupies a different time-span from blowing up the bridge, and that (ii) pressing the plunger down causes the bridge to blow up, and must therefore be regarded as a different event from it.

But this conclusion is avoidable. We can avoid saying that D1. and D4. describe different events, while at the same time agreeing with facts (i) and (ii). The confusion which would tempt someone to conclude that they were different events lies in thinking that the bridge's blowing up is the same event as his blowing up the bridge. But the supporter of the causal analysis can reply that the bridge's blowing up is an event which has something he did (pressing the plunger) as its cause, while this is not true of his blowing up the bridge. His blowing up the bridge is an action of his, and is analysable by saying that something he did caused the bridge to blow up; but it is obviously false that his blowing up the bridge can be analysed in terms of something he did causing his blowing up the bridge - for this would involve one of his actions causing another. So when we say that he blew up the bridge, at least according to the theory, we mention something he did, but we do so in a way which has reference to an effect of some more basic action of his (in this case: his pressing the plunger); which is to say that in describing him as having blown up the bridge we describe his pressing the plunger down in terms of something it caused - the bridge's blowing up. Although his pressing the plunger down and the bridge's blowing up occupy different time spans, his pressing the plunger down and his blowing up the bridge do not. And, consistently with this analysis, although his pressing the plunger down causes the bridge's blowing up, it does not cause his blowing up the bridge. To describe an

action in terms of its effects is not, on this view, to describe other actions as well, but simply to describe the same event - the event which was his action - in different ways.

This is how the causal analysis of action can be expounded as part and parcel of a certain view about the identity of actions. In the context of this analysis, actions are named by phrases like "his blowing up the bridge", "his pressing the plunger down", and it is suggested that the relation between the things which these descriptions name and the things which descriptions like "the bridge's blowing up" name can be one of event-causality.

Of course, if we accept that his blowing up the bridge can be analysed in terms of his doing something which caused the bridge to blow up, then we are forced to consider, in the same way, whether the something he did can be analysed in terms of a more basic action still, which, when described in terms of its effect, is his doing that something. The possibility of a regress to ever more basic actions and ever more basic descriptions seems to offer itself here. But we need not conclude that the regress goes on for ever, for we can always point to something an agent did which is not such that he did it by doing something else, and this we can call the action which is most basic to the situation - perhaps it is an action of moving one's limbs in a certain way. (As Davidson's radical suggestion does:

".... our primitive actions, the ones we do not do by doing something else, 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" 32)

whether this view of basic actions is correct I do not know. For the moment, however, the point is this: that just because some actions consist in doing something more basic which has an intended effect, we

32. Davidson: Agency, p. 22.

need not conclude that basic actions are analysable in the same way. Whether these have causes, say in terms of mental antecedents, is a separate question.

So what I have been referring to as the causal theory of actions - the theory that all actions except basic actions are analysable in terms of deeds which caused something, does not suggest a proliferation of different actions for every single thing the agent does, but many different descriptions of the same action (descriptions which mention different effects of the single thing the agent did); not a different action for each description. It would also be denied that an event like my putting poison in his drink lasts a different length of time to an event like my killing him, in a situation in which I kill him by putting poison in his drink. For suppose the poison takes a long time to do it's work. Then when I've finished putting the poison in his drink, the theory would say that I've finished killing him - even though he does not actually die until later. The temptation to say that I haven't finished killing him even though the poison has been put in his drink arises, I suspect, because there can be no certainty in such a situation that the effect of the poison's having been put in his glass will actually come about as planned. Until he actually dies, the possibility remains that he will vomit the liquid up and survive. But such an example only shows that we cannot describe an action in terms of its effects until the effects actually come about - and not that there are two actions. So the theory would say.

The theory therefore is that an action-sentence is analysed as one which is equivalent to one which contains two verbs, and the word "cause". But it is important to bear in mind at this point that the causal analysis is not ontological so much as grammatical. The causal analysis by itself does not imply either that there are actions, or that there are none, since it only gives the grammatical form which sentences of action have. Such an

analysis is not wholly irrelevant to the ontology of action, however, as my subsequent remarks will show.

Support for the grammatical causal analysis of action-verbs can be found in Lakoff's The Nature of Syntactic Irregularity.³³ Lakoff's proposal was that the derivation of (51) could be represented as (53), via the intermediate form (52) - (here I leave out considerations of tense)

(51) John opened the door

(52) John caused (the door open)

(53) John caused (the door be open)

The sentence "the door be open" of (53) combines with an abstract element Inchoative to form the verb of change "open" which appears in (52); and which in its turn combines with an abstract element Causative to form the transitive verb of causation "open" which appears in the surface form (51). This analysis of transitive "verbs of causation" is allegedly supported by the ambiguity of sentences where a verb of causation is formed from an adjective capable of either comparative or positive degree, e.g.

(54) John hardened the metal

which, meaning either "John made the metal hard" or "John made the metal harder", is said to be derived from (56), via the intermediate form (55)

33. Lakoff: The Nature of Syntactic Irregularity. In Mathematical Linguistics and Automatic Translation. Report no. NSI-16, Computation Laboratory of Harvard University, 1965.

- (55) John caused (the metal harden)
 (56) John caused (the metal be hard(er))

where the ambiguity of (54) is adequately explained in terms of the underlying form (56) which uses the contained sentence to represent the two alternative readings.³⁴

A controversial feature of Lakoff's proposal³⁵ is the syntactical derivation of words from phrases, in contrast to finding their synonyms in the lexicon. The general form of this lexicanization transformation is, according to Lakoff's idea,

34. Support for Lakoff's analysis is also supposed to derive from the occurrence of "pro-forms" such as the "do-so" construction, or the ordinary grammatical pronouns, in sentences like

- (i) John hardened the metal but it surprised us that it would do so
 (ii) John opened the door but it took him a long time to bring it about

where the "do-so" phrase refers to the metal's becoming hard, and where the last pronoun of (ii) refers to the door's becoming open. My own attitude to such "pro-form" arguments is that they ought to be handled with great care (Cf. my previous section); it is not at all certain that all occurrences of a pro-form refer to material contained in the sentence in which they occur. If

- (iii) John married Mary although it surprised us that she went through with it
 (iv)? John married Mary though it surprised us that she did it.
 (v)? John resembled Mary but it surprised us that she did.

are all grammatical, which is doubtful, it seems more plausible to explain the reference of the final pronouns as a reference to an implied sentence; in these cases the symmetrical sentence "Mary married John". Sentences like (i) and (ii), in other words, may owe their apparent grammaticality to a reference to an implied sentence: "the metal became hard(er)" and "the door became open" respectively. A theory which actually represented all the implications of a sentence in its deep analysis would make the argument from pro-forms more acceptable as a general strategy - but so far as I know most of the current generative theories of syntax fail to do this.

35. Criticised e.g. by Fodor: Three Reasons for not Deriving "Kill" from "Cause to Die" Linguistic Inquiry, I, 1970.

"Cause to V_{Intr} " \rightarrow " V_{Tr} "

Perhaps Lakoff's analysis can be brought more directly into line with the Davidsonian analysis of action-sentences if we represent (51) as derived from (53) via (57):

(57) John did something which caused (the door open)

(53) John did something which caused (the door be open)

and if we represent lexicalization as permitting not

"Cause to V_{Intr} " \rightarrow " V_{Tr} "

but

"Do something which causes to V_{Intr} " \rightarrow " V_{Tr} "

I do not intend to go into the merits and de-merits of this revised wording of the lexicalization transformation. Of greater interest is to notice a certain difficulty about verb-modification. For whether we phrase the analysis of D.4. He blew up the bridge as

(59a) He caused the bridge to blow up

or

(59b) He did something which caused the bridge to blow up

a difficult problem will arise as to how to transfer any modification

which might attach to the single verb in D.4. Which verb does it attach to in the underlying sentence? Or, to pose the converse question: when the two verbs of the underlying sentence are both modified, one by one modifier and one by another, how are they both accommodated by the single verb in the analysandum sentence? To make the point specific, consider

(60) Peter blew up the bridge on Sunday

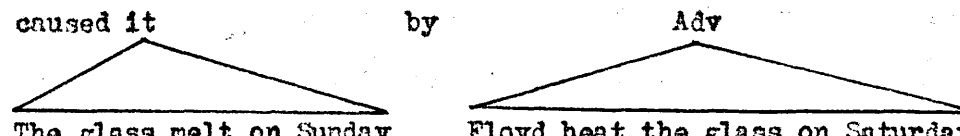
Does this have the structure

(61) Peter did something on Sunday which caused the bridge to blow up
(on Monday?)

or the structure

(62) Peter did something (on Saturday?) which caused the bridge to blow up on Sunday

A somewhat similar argument, put forward by Fodor,³⁶ is to the effect that the Davidson-Lakoff proposal would result in some underlying forms having nonsensical surface forms. Fodor cites the case of

(63) Floyd caused it by Adv

 The glass melt on Sunday Floyd heat the glass on Saturday

or in other words

(64) Floyd caused the (glass to melt on Sunday) by (heating it on Saturday)

which would, on application of the "Cause to V_{Intr} " \rightarrow " V_{Tr} " rule, produce

36. Fodor, op. cit., p. 433.

(65)* Floyd melted the glass on Sunday by heating it on Saturday

These are difficult problems. The only conclusion to draw is that they must be given solutions before the causal analysis can expect to find acceptance.

Having expounded the main aspects of the causal analysis of actions in the grammatical sense, we must now go on to deal with the ontological question of whether there are actions in some sense other than that implied by the existence of action-sentences themselves. Let me quickly summarise where we have got to so far. The grammatical analysis says that an action-sentence typically contains a causal verb; it can be grammatically analysed roughly according to this schema:

Schema 1. $A\ V-s\ O \rightarrow A[+Cause][\subscript S\ O\ V]\subscript S$

thus John grows tomatoes \rightarrow John $[+Cause][\subscript S\ Tomatoes\ grow]\subscript S$

The grammatical analysis is designed to reveal the fact that action-verbs are typically (but not always) of the causative type. This only suggests that one action-sentence may be construed, from its context, as a sentence about an action which caused an event, and this again, perhaps, for the same reasons, as a sentence about a more basic action which caused an event which caused it. On one theory, basic actions are movements of the body (as Davidson puts it: "All I ever do is move my body; the rest is up to nature"). But this causal analysis of action-sentences in terms of the events which more basic actions cause does not remove the need for a logical analysis of action-sentences generally. Whether we are treating of "John blew up the bridge" or "John did something which caused the bridge to blow up", we still have a sentence about John's actions, and one which stands just as much in need of ontological analysis as the original.

By analogy with what I said in the last section, it seems to me in this case, that if there are actions, in a sense over and above that implied by saying that there are sentences of action, it must be possible to find nominals to name or refer to them in speech. There are only actions in the strict sense if we can frame statements which assert that one action is different from another, or that one action is the same as another; and either of these types of statement, it seems, require noun-phrases of a type suitable to pick actions out. The noun-phrases standardly used to nominate actions are of grammatically gerundive type, i.e. phrases like "his blowing up the bridge", "his pressing the plunger", etc. But if this is the only way to nominate actions directly, we shall again have to question, in view of (a) the close relation between the gerund and the sentence, (b) the peculiar internal structure of the gerund and its factive sense (see last section), whether the gerundive construction really does provide a way of referring to actions, as the causal analysis seems to pre-suppose.

What other types of nominal might do the trick? We have at our disposal, to repeat, at least three grammatical types of nominal expression: the gerund, the derived, and the mixed. With a sentence like

John refused the offer

these are respectively

John's refusing the offer

John's refusal of the offer

John's refusing of the offer

Now Chomsky has made an observation concerning the derivation of derived

nominals which is extremely relevant to the ontological analysis of event and action-sentences. His observation is³⁷ that where a verb is a causal verb, there is no derived nominal. The sentence

(66) John grows tomatoes

contains a causal verb, at least according to the view which gives its grammatical analysis as

John [+Cause][_s the tomatoes grow]_s

but we find that only the gerundive and mixed, but not the derived nominal, can be formed:

(67) John's growing of tomatoes (gerundive)

(68*) John's growth (of?) tomatoes (derived)

(69) John's *growing* of tomatoes (mixed)

It is of interest to see intuitively how generally this failure occurs. It occurs for the sentence

(70) John blew up the bridge

where the verb is also allegedly causal, for the three nominals in this case would be:

37. See Remarks on Nominalisation, pp. 192-3.

(71) John's blowing up the bridge (gerundive)

(72*) John's blow up of the bridge (derived)

(73) John's blowing up of the bridge (mixed)

And it occurs, perhaps more significantly, for the explicitly causal sentence

(74) John caused the death of Herbert

for here we have

(75) John's causing the death of Herbert (gerundive)

(76*) John's cause of the death of Herbert (derived)

(77?) John's causing of the death of Herbert (mixed)

Without fathoming the complexities of the derivation of these sentences, the hypothesis seems worth entertaining that where we do have a causal verb, we cannot form a derived nominal.

Carlota Smith³⁸ invites us to reconsider this hypothesis, for she thinks that counter-examples can be found. Her suggestion is that many causal verbs (e.g. "convert", "accelerate", "expand", "conclude" "alter", "rotate", "terminate", "submerge", "assassinate") have derived nominals just like "refuse" does. To take two examples; for the case of

38. Carlota Smith: On Causative Verbs and Derived Nominals in English. *Linguistic Inquiry*, I, April, 1972, pp. 136-138.

(73) John accelerated the car

we have

(79) John's accelerating the car (gerundive)

(80?) John's acceleration of the car (derived)

(81) John's accelerating of the car (mixed)

and for

(82) John expanded the metal

we have

(83) John's expanding the metal (gerundive)

(84?) John's expansion of the metal (derived)

(85) John's expanding of the metal (mixed)

and so on. Smith's hypothesis is that causative verbs which take a nominalizing suffix of Latin origin ("-tion", "-al", "-ment") do have a derived nominal, while other verbs (e.g. "change", "turn", "stop", "kill", "raise", "end") do not. Phrases such as (80) and (84) are acceptable in her view, while phrases like

(86*) John's stop of the car (derived)

(87*) John's end of the meeting (derived)

are not. But this is barely convincing. I seriously doubt whether many people would unhesitatingly accept (80?) or (84?); and quite apart from this intuitive matter, it is hard to explain why the presence of Latin-based nominalising suffixes should make any difference. I think therefore that we can accept Chomsky's suggestion that derived nominals cannot be formed from causative verbs.

Now mixed nominals for causative verbs, although they form in a quite regular way, almost certainly name processes, and (again from an intuitive point of view) not actions at all. This leaves us with the gerund. But if all we are left with by way of a referring phrase for actions is the gerund, then conclusions similar to those derived for the case of mental events have to be accepted. The nature of action, that is to say, lies in the nature of facts. There are no actions in a strictly ontological sense, but only fact-reporting nominals and the sentences which are transformationally related to them.

Chapter IV: Prospects for an Identity Theory

1. Introduction
2. The Truth-Value of Physicalism
3. The Mental and the Physical

Appendix: Identification and Explication

1. INTRODUCTION

In this last chapter I chiefly want to accomplish two things; the first of which is to finally assess whether or not there is a version of physicalism which is likely to be true, and the second of which is to discharge the obligation I described near the beginning of Chapter I to accurately describe how the mental is distinguished from the physical. The results of these projects turn out to be inter-dependent, in the sense that what I shall suggest to be the most accurate description of the mental-physical distinction goes part of the way towards explaining why some forms of physicalism are necessarily false while others are not. These two projects occupy sections 2 and 3 of the present chapter respectively.

I shall argue that physicalism does have a certain formulation which, in all likelihood, makes it true. I shall approach my discussion of this topic via a brief summary of the main arguments and conclusions of the essay so far. In Chapter I (sections 4 and 5) it was explained that physicalistic hypotheses taken as a group could be divided into reductive hypotheses on the one side, and non-reductive hypotheses on the other. The difference between these was a difference of generality contained in their statement; identifying a mental property with a physical property or identifying a person's having a mental property with a person's having a physical property are reductive identifications which require bridge laws to establish them, while identifying mental "particulars", taken one at a time, with physical "particulars" taken one at a time, would be a non-reductive hypothesis, and ipso facto one which requires no bridge laws and which has no scientific confirmation or falsification. Since there is no generality at this non-reductive level, there is no possibility that counter-examples to such a hypothesis could be discovered by science.

Now as far as the meanings of these hypotheses are concerned, the

most important contention advanced in the intervening discussion is that phrases having the forms "A's ϕ -ing" or "A's ϕ -ing at T", which find their place in identity theories at the Time-Specific Level and the Person-Specific Level respectively, are most probably only suitable for the specification of facts. The argumentation which led to this conclusion was as follows.

We saw that in order to avoid the difficulties surrounding the spatial location of things like thoughts, beliefs, pains, after-images and sensations, Nagel proposed that things like a particular person's having a thought, a particular person's having a sensation or being in a state of belief etc., should be what any plausible identity theory of the mental should analyse in physical terms; and he called things of this sort "instances of attributes". I explained in the last chapter how it was possible to regard "instances of attributes" (like John's believing that p, Fred's desiring that q, and so on, as instances of mental states, - for nothing more substantial is involved here than a change of terminology. And I also explained how things like John's noticing that p, Fred's remembering that q, and so on, might accordingly be regarded as particular mental events - or, to use the "attribute" terminology, instances of mental event-attributes. I suggested that the general idea behind Nagel's proposal was one which we should accept. But having accepted it, we had to face the problem of spelling out the conditions under which attribute-instances are the same and when distinct. When we tried to answer this question, we found that the most obvious answers fail; and I tried to explain how some of the puzzlement which surrounds this problem can be traced to the peculiar internal structure of the gerundive nouns which, again to use Nagel's terminology, we get from filling the variable-place of an attribute-specification and nominalising.

The question which we had to ask ourselves at that stage, was this:

if a philosopher were to propose an identity theory of the mental, on either the Time-Specific Level or the Person-Specific Level (described in Chapter I), then what would his identity-statements be saying, if they consisted of gerundive nouns flanking the '=' sign of identity? Or in other words, what do nouns of this sort typically designate? Or in other words again: what general category of objects is there, which contain the items which the nouns of these identity statements designate?

The fact that standards of identity for the so-called attribute-instances could not be found, and the manner in which these failures presented themselves, lead us to question whether there were any such things as attribute-instances at all. I concluded that there were not, and that gerundive nouns should, by contrast, be interpreted factively. Not only did this conclusion seem to be suggested by the grammatical evidence, but it also provided an explanation of why the "attribute-instance" interpretation of those phrases lead to so much difficulty in the first place.

The corresponding problem about the identity-conditions for facts is, as is usual with such questions, not without it's difficulties. But gerunds are tied in a certain strict way to sentences: every English sentence converts by a standard transformation to a gerundive noun, and every gerundive noun converts, by the same transformation in reverse, to a sentence.¹ It would be counter-intuitive, I think, to suggest that gerund A

1. This theory that gerunds can be put in one-to-one correspondence with sentences actually requires a small - but not serious - amendment. This is due to the fact that gerunds, unlike sentences, are without tense; the sentences

P was about to go

P is about to go

for instance, both transform to give a single gerund:

P's being about to go

But we can always extract the tense of a sentence and nominalise the

(Footnote continued overleaf)

names the same fact as gerund B if and only if the true sentence to which gerund A is transformationally related is the same true sentence as that to which gerund B is transformationally related. Perhaps a better suggestion is this: gerund A names the same fact as gerund B if and only if the true sentence to which gerund A is transformationally related can be used to make the same statement as the true sentence to which gerund B is transformationally related.

It is difficult to decide, however, whether this is the best answer. But whatever the answer, the hypothesis that gerunds must be factively interpreted remains. All in all then, and to summarise the interpretations put on physicalism at each of the three levels of generality, we have the following types of item to analyse in physical terms. On the Property Level we have mental properties like being in pain, noticing that p, etc.; on the Person-Specific Level we have mental facts like John's being in pain, Joe's noticing that p, etc.; and on the Time-Specific Level we have mental facts like John's noticing that p at T, Joe's being in pain at time T', and so forth. We must now turn for the last time to the question of whether the doctrines of physicalism, at each of these levels, and understood in these ways, might be true.

Footnote 1 brought forward from page 164

result as usual, so that the sentences

It was the case that P is about to go

It is the case that P is about to go

correspond respectively to the distinct gerunds:

Its having been the case that P is about to go

Its being the case that P is about to go

where underlined "is" is tenseless. With this amendment gerunds and sentences can be matched one-to-one without exception.

2. THE TRUTH-VALUE OF PHYSICALISM

As far as the truth-value of these doctrines is concerned, the conclusions we have arrived at so far can be summarised as follows:

First, that a reduction at the Property Level of mental properties to behavioural properties is ruled out on empirical grounds (see Chapter I, section 3). Second, that a reduction of the mental to the behavioural at the Person-Specific Level is also ruled out on empirical grounds (see Chapter I, section 3); and third, that a reduction at the Property-Level of mental properties to cerebral properties is, as Putnam observed, ruled out on empirical grounds provided that the cerebral properties themselves are given a detailed enough kind of physico-chemical description.

This leaves various possibilities to be considered. The first is whether a reduction of the mental to the cerebral at the Property Level is possible under certain less detailed descriptions of the cerebral side of the equation; or alternatively, whether there exists some type of cerebral description which would make a reduction at the Person-Specific Level possible. It will be remembered that Putnam rejected both of these possibilities, on the empirical grounds that two psychologically identical organisms could have brains composed of different types of material, and also that the same organism could be in a psychologically similar condition on two occasions in his life history although the material "hardware" of his brain might, between these occasions, have changed. In the section of Chapter II in which I first discussed this matter, I expressed agreement with Putnam on these facts, while suggesting that further argumentation might reveal a level of physical description which would null such differences and provide a sense in which the psychologically identical could be the physically identical, either at the Property Level or at the Person-Specific Level.

I shall embark on this argumentation shortly. Before doing so, however, I must say something of another possibility that remains to be considered. This is the possibility of a non-reductive theory at the Time-Specific Level. When explaining the details of this level in Chapter I, I emphasised that no bridge law or nomological statement could be advanced in direct support of such a theory, that it was open to falsification in the way an ordinary scientific hypothesis was, and that for these reasons it was best regarded as a philosophical or metaphysical view. Indeed this is just what is meant by saying that a theory on the Time-Specific Level is non-reductive.

It has been customary in recent years to suppose that a physicalistic theory on this Time-Specific Level would concern either the connection between mental events and physical events, or the connection between mental states and physical states. My suggestion has been that mental facts and physical facts form the subject-matter of such a theory; the questions now therefore are two-fold: the first concerns the general conditions under which a mental fact like John's noticing that p at T would be identical with a physical fact like John's ϕ -ing at T, where " ϕ -ing" is some cerebral term; and the second concerns whether these conditions are such as to make this identity impossible.

Now this is not an easy question; nor is it one which has received the benefits of philosophical exploration. My tentative remarks (in the last section) as to the first question were to the effect that gerund A names the same fact as gerund B if and only if the true sentence to which gerund A is transformationally related can be used to make the same statement as the true sentence to which gerund B is transformationally related.²

2. An alternative suggestion, which I reject, is that Fact A is identical with Fact B if and only if they have the same explananda and the same explanantia. Or in symbols:

Applied to the question of the mental and the physical, the problem can now be illustrated by the following example. Suppose John does notice that p at T, and suppose that John does ϕ at T: are the two sentences

John notices that p at T

John ϕ 's at T

such that they can be used to make the same statement?

This is a question typical of those upon whose answers the truth of Physicalism at the Time-Specific Level depends. There are two circumstances, as I see it, in which the answer to this exemplifying question would be affirmative. The first, briefly, is this: if the speaker of the first sentence intended the mental verb "notices" to mean what the physical verb " ϕ 's" means, and if his audience, on that occasion of speech, understood him to be intending this, then presumably we can say that the first sentence is being used to make a statement of the same fact as that which the second sentence could be used to make. Or the situation could be reversed. The speaker of the second sentence could intend (etc., etc.,) that his physical word " ϕ 's" be taken by his audience in the same sense as that in which the mental word "notices" is normally taken. But this would be a trivial way of getting an affirmative answer.

A second circumstance in which the answer would be affirmative, and a

Footnote 2 brought forward from page 167

$$F_A = F_B \leftrightarrow (f)(f \text{ expl } F_A \leftrightarrow f \text{ expl } F_B \ \& \ F_A \text{ expl } f \leftrightarrow F_B \text{ expl } f)$$

where F_A is fact A, F_B is Fact B, where "expl" reads "explains", and where the quantifier ranges over facts. I reject this suggestion on the grounds that whether or not a certain specified fact does explain F_A as well as F_B is something which itself is liable to depend upon whether F_A is identical with F_B . The suggestion is, to that extent, circular.

less trivial one, is where there does exist some truth at a higher (and reductive) level of generality, either to the effect that the mental property of noticing that p is identical with the physical property of β -ing; or else to the effect that John's noticing that p is identical with John's β -ing. Such truths have to be established by science and not by philosophy. But were they to have been established, at the time T at which John both notices that p and β 's, then we would have a situation in which a reductive theory would support a non-reductive theory: truths established at a more general would support a truth at the least general, Time-Specific Level.

There is not space in this essay to elaborate any further on these brief remarks. I hope to have suggested the direction in which the truth-value of physicalism at the Time-Specific Level might be sought; we must now turn to inspect the matter at the other levels.

We want ultimately to consider whether for each mental property there might be a physical property identical with it. But let us approach this question by considering something weaker, viz. whether for each mental property-word there might be a cerebral property-word which is co-extensive with it.

In the case of belief, this amounts to considering whether every particular belief p might be such that

A. $(x)(p)(\exists s)(x \text{ believes that } p \equiv x \text{ is in cerebral condition } s)$

If such a thing were true, it would mean, to take one instance, that the property-word "Believing that swans fly" would be co-extensive with the property-word "being in cerebral condition S"; but not that believing that swans fly is being in cerebral condition S - for which a proper statement of law would be required.

Can we make intelligible to ourselves the situation in which proposition

A would be true? The first thing to observe in this connection is that for the case of a belief which is so obvious or fundamental that every minimally conscious person has it, there is little doubt that there does exist a cerebral condition which obtains in such a person when but only when he has the belief - perhaps the state simply described by saying that the cortex is active. Then it would be true, for this belief, that whoever "P" named, the statements "P believes that p" and "P's cortex is active" are equivalent in truth-value. This is a situation which not only can we imagine to obtain, but which, in all likelihood, does obtain for some beliefs of this fundamental or obvious kind. However it would require a sociological or cross-cultural survey to make sure that everybody in a state of minimal consciousness did in fact have such a fundamental belief - the belief that day follows night might be a candidate.

But being able to imagine situations of this sort does not settle our questions regarding proposition A in general. This is because, in general, beliefs are neither fundamental nor obvious; and so some more refined physical condition would have to be found for these, such that any organism was in that refined condition when and only when (extensionally speaking) he had one of these non-obvious beliefs. In the case of the belief that time is cyclical, for instance, our task is to guess as to whether there is some cerebral condition S, such that anyone happens to have that belief when and only when he is in that special cerebral condition. In this situation it is no longer sufficient to pick on a condition as loose as having an active cortex, since the belief, not being fundamental or obvious, is not automatically held; many persons with fully active cortices no doubt believe the opposite, and many more have no opinions on the subject whatsoever. It seems clear that the physical condition found to be associated with this belief would have to be specific in a high degree; and moreover, that a similar thing would have to apply to beliefs in general.

so that (with one small exception) each different belief-state would have to be associated with a different cerebral state. That this must be so is easily seen from the following simple argument: if the state of belief that p was found to be associated with the presence of cerebral condition \underline{c} , and if the state of belief that q was also found to be associated with the presence of cerebral condition \underline{c} , then it would be impossible for anyone to believe p and not believe q , since not only would the truth of

so-and-so believes p

have to be accompanied by the truth of

so-and-so is in c

but the falsity of

so-and-so believes q

would have to be accompanied by the falsehood of

so and so is in c

(this is what material equivalence means). But this last fact would of course conflict with the fact that the person in question believed that p . So in other words, it seems that if proposition \underline{A} is to be true, then it must be true that each belief-state happens to be present when and only when a certain cerebral state happens to be present, and in such a way that variations in belief are closely attended by variation in cerebral condition. The reservation I mentioned concerns pairs of beliefs, if any,

which either because of their logical equivalence or for some other reason, must be held or abandoned together. In these cases we might wish to reserve a distinction between the beliefs (we count them as two) while only needing a single cerebral state to accompany them both; so that in such cases a difference between the beliefs would not be mirrored by a difference between two cerebral conditions - as is true in the general case.

To say all this is to minimally describe the situation which would have to obtain in order for proposal A to be true. Most importantly, perhaps, is the fact that there need be no conflict between this kind of situation and Putnam's observation that psychologically identical organisms can have brains made of different types of stuff. It is of vital importance to see how such a conflict can be made to disappear. The essential point is that the cerebral conditions (or configurations, we might call them) could be described, presumably, without any reference to the actual kind of material in which they appeared, and yet without the use of mental or psychological language. There are many ways of giving a physical description of a mechanism: one is in terms of the molecular composition of the material out of which its parts are constructed; one is in terms of the sizes and strengths of the component mechanical parts of the mechanism; another is in terms of how these parts function in relation to one another; perhaps another is in terms of the flow of the "input" through the mechanism, the manner in which it is processed by the mechanism before issuing in "output" (Freud's early hydrological model of the psychical processes is perhaps an example). Given these different modes of physical description, it is not difficult to appreciate that two mechanisms which are dissimilar when described in molecular terms might easily be similar when described in mechanical terms; that two mechanisms which are dissimilar when described in mechanical terms might be similar under a description in functional terms; and that two mechanisms which

are dissimilar under a functional description might be similar, conceivably, under a "processing" description. Only under the extremely dubious assumption that the only way of providing a physical description of a mechanism is in terms of the actual type of material of which it is composed does Putnam's observation entail that mental property-words could not be co-extensive with physical property-words.

We can surely envisage the discovery of a pattern of cerebral organisation which was such that the matter of the brain happened to become differently configured with the arrival or exit of a new belief or desire. The hypothesis requires that for each mental change there is a physical change - but not vice-versa: it also requires, as I have suggested, that any particular parcel of cerebral matter (any brain) is capable of assuming as many configurations as there are sentences which can report what that particular subject believes, fears, etc. These need only be finite in number, in spite of the fact that the number of propositional-attitude sentences in the language is infinite, since any organism obviously has no more time in his life than for a finite number of attitudes. We are dealing with his performance at this point, we might say, rather than with his competence.

In order to explain in a more detailed and unambiguous way what this system of cerebral configurations would be like, it is instructive to consider a more general hypothesis. The idea that there is a system of cerebral configurations such that a different one obtains when and only when the agent is in a different particular mental state seems, on the face of it, the same idea as that which says that the details of a person's successive mental states are written in the material of his brain in such a way that we could find out what a person's beliefs or memories were by inspecting his cerebral matter directly. Now there is indeed one sense in which this equivalence obtains. But the idea that mental states have

an internal representation in the person's brain or nervous system is one which must be treated with the utmost care, for in fact there are several different forms which it can take, each differing considerably in credibility from the others. It is for this reason that the least plausible formulations of the "brain-writing" hypothesis must be carefully stated and rejected. Doing this will also serve to sharpen the details of the particular version of the brain-writing hypothesis which I want to suggest is probably correct.

One thoroughly implausible version of the notion that a person's mental states have an internal representation has been advanced by J. Zeman, and exposed to justified criticism at the hands of D. C. Dennett.³ According to Zeman, the brain contains a system of "universal writing" which "says" or exhibits exactly what the mental condition of the subject is, and which will eventually be laid open to inspection and made "readable" by the relevant sciences of the brain. But as Dennett points out, the postulation of such a system of brain-writing involves a clear committal of what is known as the homunculus fallacy. For in order for a system of brain-writing to be a system of writing in the ordinary sense of the word, a writer must also be assumed; and to assume such a writer is to assume an agent who means something by producing the various pieces of brain writing he does produce. To write is to act in a certain linguistic way; and to act in this linguistic way is, among other things, to have certain intentions towards an audience. So postulating a system of brain-writing cannot explain the mental states or capacities of the agent whose brain it is, because the postulation involves postulating a separate internal person,

3. J. Zeman: Information and the Brain, in N. Wieser and J. P. Schele (eds.) Nerve, Brain and Memory Models. N.Y. 1963. See Dennett, Content and Consciousness p. 87 for the reference to Zeman and for Dennett's criticisms.

whose mental states and capacities clearly stand in need of the same kind of explanation. The threat of regress is obvious.

The notion that the brain contains an independent system of writing, which inscribes the relevant mental information about the person concerned in his cerebral matter, is therefore the wrong notion to entertain. An apparently more plausible idea is that true information about a persons mental life is somehow represented or stored in his cerebral matter; so that if a person desires or believes something, the desire or the belief is somehow contained in his brain. This is not to suggest that the information is written there; but that it is so to speak lodged in the matter of his brain in much the same way that the information in a book is contained in a library when the book is given a place on one of its shelves.

This is a much more complicated suggestion than Zeman's hypothesis, and its critical assessment depends to a large extent upon what kind of representation or store is envisaged. If it is envisaged that mental information is represented or stored in such a way as to make it retrievable by other mechanisms of behavioural control and eventually used as one of the determinants of behaviour, then two things become clear. The first is that the system of representations must, in all likelihood, be generative in the way that a generative grammar is, for the reason that, if this were not so, there could not exist a retrieval mechanism which could select the appropriate information at the appropriate time, and extend its retrieval capacities to new information in future cases. Moreover if the representational system was non-generative it would be unlearnable by a human observer. That is to say, it would only be possible to learn and understand the representational system if one knew how a finite stock of sub-representations combined to make up the representation of a complete piece of information. It may be baffling to imagine what such a generative representational system would look like physically, but it is likely that unless it did

contain generative features of the sort currently attributed to a language-system, the retrieval mechanism would not have sufficient material to work on successfully. (It would be unable to distinguish the agent's belief that John is easy to please and the agent's belief that John is eager to please as being grammatically different kinds of beliefs).

Whether or not the fact that the representational system must be generative involves this brain-writing hypothesis in the homunculus fallacy all over again depends upon whether the postulation of a generative representational system requires the additional postulation of an internal agent having linguistic intentions of its own. I am not completely sure whether this is so, although the content of my first chapter perhaps suggests that it is. But in any case, it is clear that the kind of brain-writing system now under consideration involves a committal of the homunculus fallacy for a different reason. For in order for the stored or represented information to be used by the organism (via the operation of a retrieval mechanism) in the determination of its behaviour, something very much like another internal agent has to be postulated as that mechanism whose operations are effective in both putting the belief-information to work, and in assessing its compatibility with newly-acquired beliefs. Not only would we have to postulate a mechanism which knows which situations are appropriate for retrieving a piece of belief or memory information and putting it to work, but the mechanism must play the role of rational assessor, refusing to store two blatantly inconsistent pieces of belief-information and cataloguing the connections between one piece of belief-information and another. And here the regress sets in again, for the rational assessor will need a store or library of his own in which to lodge his own procedural principles.

These or similar considerations appear to affect any brain-writing hypothesis whose point is to explain how belief and memory information can

be stored and used by analogy with a library and a retrieval device. (It is worth adding that, as such, they affect the methodology of a long tradition of scientific research into the location of the memory store in the brain). But the version of the brain-writing hypothesis which I began to explain, and which I want to defend, is not affected by considerations of this kind. This version is arrived at by dropping the idea of an information store altogether, and giving a different interpretation to the proposition that mental information can be represented in the brain. It says simply that for each mental state or condition of the subject, a certain cerebral configuration might be located such that the subject was in that mental state or condition when and only when the cerebral configuration obtains. This "when and only when" clause is extensional, and so understood, the hypothesis says no more than that the relevant mental state and the relevant cerebral configuration accidentally co-exist.

Now I think there is an important observation to be made about this version of the brain-writing hypothesis, and this is that there seems to be an acute psychological difficulty in imagining that a cerebral configuration and a mental condition may always occur together without at the same time imagining the co-occurrence to be law-like, or nomological, in kind. Were we to discover that a person believed that p when and only when (extensionally speaking) his brain was in a particular configuration C, then it would be hard not to suppose that various subjunctive statements, of the sort which law-statements license, were also true. It would be hard not to suppose that were the person to believe that p, his brain would be in configuration C, or that if the person's brain were not in configuration C, he would not believe that p, and so on. It is for this reason that it seems pointless from a practical point of view to distinguish the hypothesis that cerebral configurations happen to co-occur with specific mental conditions from the hypothesis that they do so in

accordance with a law. In other words, the hypothesis that is being defended is that there is no a priori reason for supposing psycho-cerebral laws to be impossible.⁴

It is interesting to notice what could be said about the meaning of cerebral configurations, if there actually were any law-statements connecting the psychological and the cerebral. The usual method of attacking the argument that psycho-cerebral laws are possible consists in pointing out that there is no way of framing such laws which does not re-introduce psychological or intensional concepts on the physical side of the equation. This line of attack is admissable and powerful, and indeed I have been using it myself to show how the more ambitious forms of the brain-hypothesis are ruled out (Zeman's method required an internal agent or homunculus with mental capacities and conditions of his own, and the representational or storage method was eventually seen to require something like the same thing). On my hypothesis, according to which a cerebral configuration could stand in a law-like relation to a mental condition, these objections are avoided, but a different kind of meaning enters the situation. For although each one of these configurations is without sentential meaning in itself, anyone who know which mental state was nomologically associated with it could, by ascertaining that the configuration obtained, thereby ascertain which mental state obtained; and this allows us to say, I think, that the obtaining of each particular configuration has a kind of non-sentential significance for the observer.

In his original paper on Meaning,⁵ Grice introduced a terminology and

-
4. In spite of Putnam's observation (see Chapter 2) that different organisms can share mental attitudes although their cerebral matter may be of a different sort. The kind of difference which Putnam speaks of is on the level of the type of matter of which their brains are composed; but I have already suggested that the configuration system which I have supposed to exist can be described independently of the sort of matter of which any actual brains are made.
 5. Philosophical Review, 1957, pp. 377-88.

and a typography which exactly suits the job of explaining this point.

The sense of the word "means" which occurs in statements like

Those spots mean measles

The thunderclouds mean rain

he identified as the natural sense; while the sense in which the word occurs in statements like

P's statement "A" means that p

he identified as the non-natural sense - the two senses to be distinguished from each other by the use of the words "means_N" and "means_{NN}" respectively. This is just the contrast which I myself want to employ: for instead of saying that a particular cerebral configuration has a non-sentential significance for the observer, I could equally have said, in Grice's typography, that the configuration means_N that the subject has such-and-such a mental state.

It is worthwhile to notice that whenever we have a law-like relation between two types of phenomena, we can always speak of meaning_N - at least, whenever the law in question is a natural law, a law of nature. It is because the relation between spots and measles is law-like, and because the relation between thunderclouds and rain is law-like, that the presence of the first-mentioned phenomenon means_N that the second-mentioned phenomenon is (or will be) present. In the mental-physical case too, of course, it is because the relation between cerebral configuration and mental states is law-like that we can say, having observed the cerebral configuration, that it means_N that the subject is in the related mental state. There need

be nothing essentially linguistic about configuration systems, and no question of the systems being a code or anything else about which questions of meaning and translation would arise. Just as you cannot ask what a sound pattern means or what language it is in, so there is no room for questions about what a cerebral configuration means (in the non-natural sense), or about what language it is in.

The hypothesis that there can be psycho-cerebral laws is not yet completely free from philosophical difficulties, however, for there is a different line of argument which attempts to show that there can be no laws of any kind which connect the mental and the physical. This line of argument proceeds by suggesting that there can be no law-like connections between mental phenomena and behaviour - with the implicit suggestion that the same considerations apply to attempts to construct law-like connections between mental phenomena and cerebral phenomena. However, I think it can be shown that none of the factors which make it impossible to obtain nomological connections between mental sentences and sentences about behavioural motions apply to the "centralist" hypothesis now under consideration. It can be shown, that is to say, that the mental is not totally irreducible, but only irreducible when the physical phenomena selected are behavioural motions. Let us appreciate why.

A typical explanation of why there can be no psycho-physical laws of a kind which connect the mental with the physical motions of behaviour comes from Davidson:

"Any efforts at increasing the accuracy and power of a theory of behaviour forces us to bring more and more of the whole system of the agent's beliefs and motives into account. But in inferring this system from the evidence, we necessarily impose conditions of coherence, rationality and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena." 6

6. Psychology as Philosophy, p. 4.

Now I think there are really two parts to this explanation. One consists of the fact that we cannot point to a single state of desire for a certain object (to take a specific example) and find a specific kind of physical change in the world which the agent initiates as a result, because the agent who has that desire will initiate different changes - or none - depending upon what his other beliefs, intentions and fears are. To say this is part of what is involved in saying, as we did in Chapter 1, that what the agent who has that desire will do depends not upon what his environment is like, but upon what he believes or judges his environment is like. The second feature of this explanation consists, in Davidson's own words, of "emphasising the holistic character of the cognitive field"; and what is specifically meant by this, is that our theory of the total mental state of any particular subject is underdetermined by the evidence. We can, while saving the phenomena, adjust and revise our theory in some places, if we make compensatory adjustments and revisions in other places. We can re-interpret what a man means by uttering certain sentences, to take one clear case, providing that we make reasonable compensatory adjustments in our theory of what he intends or believes. These two features of the explanation are, I think, distinct from each other. The first emphasises that a man's behaviour depends upon more than a single isolated facet of his total mental state; and the second emphasises how underdetermined by the evidence our third-person view of another person's mental life is.

I doubt whether the second feature actually pulls much weight in the explanation. Many theories in the sciences are, if we believe Quine, underdetermined by the evidence, so that two theories can account for all the evidence and yet be formally incompatible with each other. The degree of underdeterminedness in the physical sciences may be smaller; but it is there. While the constraints on a theory of behaviour are coherence,

rationality and consistency, as these apply to the beliefs (etc.) which are supposed to be held by the subject, the constraints on a theory of physical phenomena are logical coherence, and formal consistency. I should like to add that even if we can mutually adjust our judgements as to a persons mental states, this does not entail that the subject has anything but a fixed view of them. It merely indicates that, pending the opportunity to inspect a man's brain to find out what he believes, people other than the subject are relatively badly off when it comes to knowing what his (the subject's) mental states are. What the subject means or intends by uttering a certain sentence is perfectly available and determinate to him, that is to say, since he does not learn about his mental life in the way we observers do, by inspecting what he does or listening to what he says.

The first feature of the explanation, therefore, does most of the work in effectively explaining why psycho-physical laws are impossible in a theory of behaviour. But this feature has no application to the problem of psycho-physical laws where the physical phenomena are cerebral. For in this case the laws are not causal, as they are in a theory of behaviour; they do not try to assert what will happen as a result of a subjects possessing a certain mental state. The problem of representing the subjects procedure of decision in the light of his own (possibly conflicting) beliefs and fears has no counterpart in the case of psycho-cerebral laws. These therefore are not in the same boat as psycho-physical laws which purport to connect a subjects mental state with his behaviour.

In this section I have tried to defend the view that psycho-cerebral laws are possible, which is to say that the mental is not totally irreducible to the physical, as is generally supposed. Whether there are such laws, and what they are like if there are, are of course separate questions; although the progress of research into the workings of the brain does seem to suggest that such laws will eventually be uncovered.⁷ However if the

Footnote 7 printed overleaf

considerations advanced in this section are correct, then it follows that the neurophysiologist's investigations are likely to be piecemeal, of necessity, in the sense that psycho-cerebral laws would only be discoverable singly. For only if the system of cerebral representations was generative, in the sense I explained earlier, would he be able to predict the details of some psycho-cerebral laws on the basis of his knowledge of a handful of basic ones, in the way a generative grammarian can make predictions about the structure of some sentences on the basis of his investigations into the structure of a quite small chosen number of others. The fact that the most intelligible system of cerebral representations is not generative means that the neurophysiologist will have to proceed step by step. The practical reduction of the mental to the cerebral may be difficult; but it is not, if I am right, an enterprise which is by any means philosophically unintelligible.

3. THE MENTAL AND THE PHYSICAL

It would be unsatisfactory to end an essay on physicalism without some remarks on the meanings of the words "mental" and "physical". The question of what distinguishes these meanings, as I remarked in the section of Chapter I entitled The Need for Mental Terms, is a question which lies at the centre of the subject; but although philosophers since Descartes have been convinced that there is a clear difference, its exact nature has been a matter of considerable dispute.

Footnote 7 brought forward from page 162

There are already plenty of negative law-like connections between the mental and the cerebral, of the form: "if the brain is in such-and-such a state, the subject will not feel (hear, see, etc.) anything". (See also the Guardian newspaper, 28th June, 1972, p. 8, which contains an interesting report to the effect that "specific aspects of behaviour, such as fear of the dark, can be permanently induced by chemicals".)

Two methods of tackling this question must be distinguished; one method is ontological, in as much as it concerns the phenomena themselves, while the other is linguistic, in as much as it concerns the difference between mental and physical language. This difference corresponds, of course, to two of the senses in which the words "mental" and "physical" are in fact employed: for sometimes philosophers speak of the differences between mental and physical phenomena, and sometimes they speak of the difference between mental and physical terms or sentences. I shall briefly mention the ontological approach, then I shall mention a method which is partly ontological and partly linguistic, and then I shall embark on an explanation of the difference of my own, which I believe is almost totally linguistic.

Those who adopt the ontological approach to the distinction often argue as follows: they say that the distinctive features of mental phenomena - those features which distinguish them from physical phenomena - provide us with an explanation of why it is that mental phenomena and physical phenomena cannot (logically cannot) be connected in a statement of natural law, or else identified as the same thing. Indeed, once something has been found which adequately distinguishes mental phenomena from physical phenomena, little more needs to be said to justify their non-identity and their non-reducibility.

Descartes, for instance, thought that mental phenomena could be distinguished from physical phenomena on the grounds of the non-spatiality of the former.⁸ If this is correct, then it follows at once that mental phenomena cannot be physical phenomena, at least if Leibniz's Law is

8. Descartes: Meditations

constitutive of identity. Or again, some modern philosophers⁹ believe that mental phenomena possess the property of Intentionality, while physical phenomena do not. Again, the conclusion of non-identity must follow if these philosophers are correct.

Now in point of fact, this ontological approach to the question of the meanings of the words "mental" and "physical" is altogether wrong. In the first place, philosophers who argue in this way do so without taking any notice of the difficult semantical question of whether there are any mental phenomena, or whether there are any physical phenomena, in the strictest sense of these existence-asserting phrases. Given a sentence like

(1) Harry believed that the moon was made of cheese

it is not a matter of arbitrary choice to divide the referring terms from the predicates, as my previous chapters will, I hope, have made clear. And yet to say that what the sentence is about is Harry's belief, and to go on to speak of its non-spatiality, as Descartes would have done, carries with it a theory of the semantics of the sentence which must be properly argued for, and not just assumed to be true without argument. The Cartesian can of course point to sentences in which the phrase "Harry's belief" actually occurs as a subject, or even to sentences having the phrase "Harry's believing" as subject. But if they assert that these sentences contain references to entities of a sort which can be spatial or non-spatial, then they will again have made semantical assumptions, or assumptions about interpretation, which

9. E.g. H. H. Price Some Objections to Behaviourism in Hook (ed.) Dimensions of Mind (pp. 79-84). Price's view is an extension of Brentano's theory that "...intentional inexistence is exclusively characteristic of mental phenomena. No physical phenomenon manifests anything similar. Consequently, we can define mental phenomena by saying that they are such phenomena as include an object intentionally within themselves" (From The Distinction between Mental and Physical Phenomena.)

stand in need of defense. Indeed the reason for my own disagreement with the contentions of Descartes and those who agree with him on this subject, is that the semantical assumptions implicit in the doctrine of non-spatiality are wrong. My arguments in the last chapter were designed to show that what such a sentence as (1) is about, in the strictest sense, is not Harry's belief, or even Harry's believing, but Harry. Moreover any sentence in which the phrase "Harry's belief" actually occurs as logical subject will only be about what Harry thought; and any sentence in which the phrase "Harry's believing" occurs as logical subject will be about the fact of his believing. None of them are about any well-defined entity or particular such as a particular event or state.

Yet another reason for disagreeing with Descartes' notion of non-spatiality as a criterion of the mental is simply this: Harry's belief, like any other persons thoughts or memories, would, were it to exist as an entity or some kind, be spatial or non-spatial to no greater extent than anything else of a logically similar (but clearly non-mental) kind, like Harry's stumble, Harry's haircut, Harry's height or weight, Harry's burial, and so forth. And if Harry's believing could be shown to be a state of affairs, then its location in space is easily identifiable as the location in space which Harry himself occupies.

As for H. H. Price's idea that mental phenomena are distinguished from physical phenomena by having the property of Intentionality, much the same arguments can be made to apply. Suppose it was said that Harry's belief has the property of Intentionality, in the sense of being "about" some non-existent state of affairs - his being irresistible to women, say - then we need only observe that a sun-flowers need can be intentional in just the same sense, when the object of its need was the substance water in a world which for some cosmic reason had de-hydrated. Surely no-one would assert that the sun-flower's need for water was anything but a purely physical affair? A quite different approach to the problem of distinguishing the

meanings of the words "mental" and "physical" is to concentrate attention upon the language, specifically sentences, by which the phenomena are described, rather than on the properties of the phenomena themselves.

The classical answer to the question about the distinction between the mental and the physical, that of Brentano and Chisholm, is in point of fact partly linguistic, in so far that mental phenomena are specified as those which must be "described" by the use of intensional language. Adding the doctrine that intensional language is, in a sense, irreducible is then said to secure the overall view that mental phenomena themselves are not reducible to non-mental phenomena. Before exposing what I take to be the shortcomings of this programme, let us be quite clear of its aim. Its aim is eventually to characterise mental phenomena, but to do so in a way which is partly linguistic. Chisholm's plan, as I said, is to accomplish two things. First, to characterise intensional sentences; and then to demonstrate that in order to describe a mental phenomenon you must use such a sentence. So this, the classical method, incorporates both the ontological approach to the mental/physical distinction as well as the linguistic approach. It not only attempts to characterise the difference between mental and physical phenomena; but its method of doing so is to elucidate differences in the language with which these phenomena must be described.

Chisholm's original opinion (which has subsequently undergone some modifications) was that a sentence is intensional if it has any of the following logical properties: that its truth is not dependent upon the truth of a contained embedded sentence; that its truth is not dependent upon whether its contained nouns all refer; and (thirdly), that substitutivity of identity in an embedded sentence fails to preserve the truth-value of the whole. It is not to my purpose at this particular moment to emphasise how murky the concept of an intensional sentence is. The point to be observed is that an intensional sentence is defined by Chisholm as

one which has any one of several features. Let us next observe how he takes sentences of this sort to be linked to the mental:

"We do not need to use intensional sentences when we describe non-psychological phenomena But when we wish to describe perceiving, assuming, believing and other attitudes, then either (a) we must use sentences which are intensional, or (b) we must use terms we do not need to use when we describe non-psychological phenomena" 10

We need to understand what Chisholm means by saying that we need to use intensional sentences in order to describe mental phenomena. (My own preference would be to understand by this that we need to use intensional sentences in order to predicate a psychological phrase of a person). It could mean either that no intensional sentence has a non-intensional logical equivalent, or that no intensional sentence has a non-intensional nomological equivalent. The first disjunct, if true, would prevent a non-intensional analysis of an intensional sentence (assuming that the aim of analysis is to provide logical equivalents); and the second disjunct, if true, would prevent a non-intensional nomological equivalent being found for an intensional sentence. But notice now that to assert the first disjunct is only to deny the doctrine of logical behaviourism, the doctrine that sentences containing mental verbs have logical equivalents which do not; but the falsehood of logical behaviourism must be obvious anyway, since sentences containing mental verbs cannot mean the same as sentences not containing mental verbs. As to the second disjunct, it can either mean that there are no psycho-behavioural laws, or else it can mean that there are no psycho-cerebral laws. From the examples in Chisholm's text,¹¹ we must assume that he means the former. But we saw the evident plausibility of this doctrine (essentially Brentano's) in Chapter I; whereas we saw more recently how implausible it was to assume for similar reasons that there are no psycho-

10. Chisholm: Perceiving, p. 172. My underlining.

11. See Perceiving, Chapter 11.

cerebral laws. One of my complaints with Chisholm's statement (as expressed and intended by him) is that it amounts to no more than a denial of two positions which are, respectively, evidently false and evidently implausible: the position of logical behaviourism and the position of reductive, or nomological behaviourism. It does not affect the more plausible version of materialism which asserts a nomological link between the psychological and the cerebral.

A second, seemingly less central feature of Chisholm's proposal is his use of the phrase "describe psychological phenomena". I have already said that I should prefer to phrase things differently; but the point is not simply terminological, because it affects the coherence of the idea that descriptibility in intensional terms is a distinguishing feature of mental phenomena. I have been stressing throughout the latter part of this essay how important it is to obtain a correct (or defensible) view of what psychological phenomena there actually are. To obtain such a view involves answering many complex questions about the logical form of sentences containing mental verbs (defined by enumeration (see below)); or, in Dennett's words,¹² it involves developing a theory as to which of the terms of these sentences are referential. The conclusion to my own investigations into this question was that, at least for event-sentences like "John noticed that the ship was sinking", no event whatever is in fact "referred to". This is one instance in which a lack of a semantical theory might lead one to say that some mental phenomenon had been described, whereas a critical view of the logic of such a sentence would suggest the opposite. Chisholm might better have said that we need to use intentional sentences when we wish to predicate a psychological phrase of a person.

A formally decisive reason for saying that mental sentences do not

12. See Content and Consciousness, Chapter 1.

"describe psychological phenomena" derives from the fact that they are often intensional. Anyone who supposed that a phrase like "Galileo's noticing that the earth moves" describes an individual would automatically invite the response that substituting for "the earth" a different noun for the same object might turn the complete description into one which is true of nothing. This would happen if Galileo did notice that the earth moves but did not notice that the planet formerly thought to be at the centre of the universe moves. That is, the formula

$$E = \{x \mid x \text{ is Galileo's noticing that the earth moves}\}$$

does not determine a well-defined entity.

I can now bring into prominence a doubt which I earlier laid aside. Even if Chisholm was strictly correct in supposing that a sentence containing a psychological verb "described a mental phenomenon", and even if he was also right in supposing that sentences containing psychological verbs are intensional; and if he used these two facts to infer the conclusion that intensional sentences described psychological phenomena, then several gaps in his theory would still remain. In the first place, it would remain totally obscure as to why intensional sentences should be so peculiarly appropriate for describing mental phenomena. The point here is that intensional sentences are classified as those which fit into various patterns of inference. And yet why should sentential features of this logico-grammatical or syntactical kind bear any relevance to the kinds of thing which those sentences described?

(A connected puzzle (and this is the last thing I shall say about Chisholm's proposal) concerns the concept of intensionality itself. Intensionality, as predicated of sentences, is not in fact a unified phenomenon at all, in as much as its definition consists of disjoining

three seemingly unrelated inference-features. Admittedly, intensionality is in some sense the opposite (i.e. the failure) of extensionality; but to define intensionality like this is no help at all, for in order to get a fully adequate theory of extensionality, some theory must be evolved as to what substitutions do fall under the general heading of "the substitution of identically referring expressions". Do phrases other than proper names refer? In what sense do predicates refer? In what circumstances can one predicate be substituted for another? And so on. Intensionality and extensionality are not quite the clearly exclusive concepts that philosophers like to imagine. Until a theory is produced which incorporates a sense in which they are exclusive, and clearly exclusive, our understanding of what it is for a sentence to be intensional, and everything that depends on that understanding, must remain obscure.)

I now want to move on to make some tentative suggestions of my own about the mental-physical distinction. I cannot claim finality for my suggestions on this immensely difficult subject, and indeed I shall spend a good deal of time in considering objections; but I do claim for what I shall say that it has the virtue of bringing into prominence a feature of the mental which is not sufficiently stressed. For what I wish to suggest is that there is an essential connection between the mental and one type of self-consciousness, which is of such a type as to be sufficient to provide us with an adequate way of making the distinction between the mental and the physical. Explaining this suggestion in a completely unambiguous and clear way is no easy matter, however, and there are several difficulties in doing so which I cannot pretend to have seen clearly how to overcome satisfactorily.

One of the major difficulties is to give an adequate description of the kind of self-consciousness which, as I shall argue, any person must have towards himself if a mental sentence is to be true of him. (For

what I want to suggest is that, as a matter of fact, mental sentences are all and only those which cannot be true unless there is a certain reflexive attitude present on the part of the subject of that sentence; and that mental sentences are distinguishable from physical sentences in just this respect. To be a little more exact: my suggestion is that this reflexive attitude, this kind of consciousness, must be directed by the person concerned towards himself as the subject of the sentence in question. That is to say, the subject of a true mental sentence must be conscious of himself as subject of the sentence concerned.

Now, of course, any fact at all concerning a person can be one which the person in question can in principle be conscious of; but my suggestion is that the mental facts are just those which cannot be facts unless the person in question is conscious of them as facts. For instance I can be conscious of myself as having an ingrown toenail or a bruised forehead, but it can be true that I have an ingrown toenail or a bruised forehead even if I am not conscious of myself under those descriptions. Those facts - the fact that I have a bruised forehead and the fact that I have an ingrown toenail - are physical facts; the sentences which express them can be true without any consciousness on the part of the subject that they are true.

Examples of this kind perhaps lend a certain initial plausibility to my method of drawing the mental-physical distinction; but other matters must be discussed before the account can be considered complete. For instance there are so many different senses to such words as "conscious", "aware", etc., and so many difficult and subtle distinctions to be drawn between one sense and another, that it is essential that a careful description be given to the particular attitude I have in mind. This descriptive task is the first to which I shall address myself. I shall then try and explain the importance of this reflexive attitude to the

conditions under which mental sentences in general are true. And then finally I shall make some suggestions about the relevance of this method of drawing the mental-physical distinction to the theories of physicalism.

In one sense of the word "conscious", a person is conscious of something if he is fully aware of it, in such a way that, when the person is a human being with a command of a language, he could express the fact that he is conscious of it, either to some other person or to himself. It is only essential to this rather full-blooded concept of consciousness that its owner could thus express himself linguistically, if he chose to - not that he actually does. There is however a much less full-blooded concept of consciousness - and a rather more ubiquitous one - according to which a person can be conscious of a thing even though he is not conscious of it in the former, full-blooded, sense. If a person is looking at a photograph, say, and attending to or focusing his attention upon one particular part of it, then he will very likely be conscious of other details in the photograph in this second, weaker sense. If the photograph were to be snatched away from him suddenly, and if he were asked about the photograph as a whole, then very likely he would not be able to say what the peripheral details were. This fact, if it were a fact in any situation, would show that the person had no consciousness of the peripheral details in the first sense of the word "conscious"; and yet more likely than not he would have been conscious of those peripheral details in the second sense of the word. The evidence for this, and what I think is the kind of evidence which is essential to this concept of consciousness, is that he could be reminded what those peripheral details were like; and that if they were shown to him, they would seem familiar. I shall argue that it is consciousness of the sort exemplified in this example which, in a reflexive form, is the kind which is an essential concomitant of the mental. In other words, I shall argue that a mental fact cannot exist unless it

itself is the object of this type of consciousness on the part of the person whose mental fact it is.

Up to this point in my explanation I have been mainly concerned to make it seem plausible that there is a kind of consciousness of things and/or facts which is distinct from the kind of consciousness which is often referred to as "full awareness", and which I earlier called "full-blooded consciousness". The next step is to advance some positive reasons for supposing that this weaker type of consciousness is what distinguishes the mental from the physical, and that it does so not simply by accompanying the mental fact, as a vague intransitive background attitude, but by being transitively directed onto the mental fact itself. I shall persist in referring to the stronger, full-blooded consciousness, with the description "full-blooded consciousness", while reserving the unqualified term "consciousness" for the weaker variety of the attitude of that name. The terms "fully self-conscious" and "self-conscious" are used accordingly.

The next task before us is to roughly delineate which sentences are intuitively accepted as the mental sentences, for not until this is done can we begin to assess any theory as to what characterises them. Again, this is no easy matter, for the reason that intuitions about the content and the structure of the category of the mental are likely to differ from philosopher to philosopher, and it is hard to see exactly how such differences might be reconciled. Arguments might conceivably arise, for instance, as to whether such words and phrases as "stupid", "clumsy", "ignorant", "depressed", "is in pain", "had a sensation" are part of the mental vocabulary, and, if they are, as to whether they are more or less fundamental than the concepts of belief or thought or imagination. However, I shall assume that the verbs "think", "believe", "desire", "seem to see", "fear" and "expect", at the very least, occupy a place at the centre of the concept of the mental, and that verbs like "know", "learn", "see", "notice",

"remember" do also; and I shall say that the verbs of the first group are Category A mental verbs, and that verbs of the second group are Category B mental verbs. These assumptions are made on intuitive grounds: no one is likely to dispute the membership of the listed verbs to the mental vocabulary, however disputable the membership of words like "stupid", "depressed", "clumsy" and so forth might be. Perhaps another justification - though this is not a formidable piece of argument - is that where the route through an argument is unclear, it is generally a prudent policy to choose one direction rather than none, and see what results.

Having very roughly identified some central mental verbs, I next define a mental sentence as one which either has the following structure, or is clearly convertible without loss of meaning into one which does:

Person name + Mental verb + Noun-phrase

where the noun-phrase place can be occupied by a noun-phrase of any kind whatever. Notice that according to this specification, some mental sentences will be intensional ("Galileo believed that the earth moved"), while some will not ("Galileo saw the sun"). Intensional or not, it is mental sentences thus specified that I suggest cannot be true unless they are at the same time objects of a true sentence beginning "so-and-so is (weakly) conscious that", where "so-and-so" is the name of the subject of the contained mental sentence. This, then, is the theory I wish to defend.

My method of defending this theory will be to answer the several objections that will very likely be brought against it. Some of the objections, it must be admitted, do look powerful; but I think that all of them can either be answered directly or else defused, and in such a way as to provide indirect support for my contention.¹³ The objections are

Footnote 13 printed overleaf

not related to each other in any very clear way, so rather than try and provide a spurious continuity to the argument, I shall simply list them, together with their answers, one by one.

The first objection concerns mental sentences containing Category B verbs. While conceding that the theory looks acceptable for sentences containing Category A verbs, the objection says that a sentence which contains a Category B verb can be true although its subject is quite ignorant of its truth, and is, hence, not conscious in any sense that the sentence "fits" him. Sentences containing Category B verbs are those which require, for their truth, the truth of the contained sentence: P cannot know that p unless p, P cannot learn or notice or see that p unless p, and so on. The objection says that it is this fact, the fact that a person's knowledge is not wholly a condition of the person but also a condition of the world, which allows the mental sentence to be true even in cases where the subject of the sentence believes it to be false. The following "situation" illustrates this argument more fully. There is a fly on the window pane;

Footnote 13 brought forward from page 195

Quinton has raised a problem which does not fall into this category. Considering the theory that "everything mental is an object of consciousness" (Mind and Matter, p. 226), he says that the theory precipitates an infinite regression, since it entails that "every act of consciousness is an object of consciousness (but) If x is an act of consciousness then it must be the object of a further act of consciousness y which itself is the object of z and so on" (pp. 226-7; my emphasis). His point here - in my terms - is that since "is conscious that" is itself a mental verb, it cannot form a true sentence unless it is itself embedded in a larger "is conscious that" proposition, and so on ad infinitum.

But in point of fact it is only grammar that produces the illusion that every act of consciousness needs a further act of consciousness, for an act of consciousness by a person can be its own object. The situation is similar to looking at oneself. With the aid of a mirror, P can look at himself looking at himself, and if he does so, then we can expand the story and say that P is looking at himself looking at himself looking at himself looking at himself etc. But here it is only the narrative which is potentially infinite, and not the number of lookings. Grammatical objects are embedded in grammatical objects, but that is all.

P, glimpsing something out of the corner of his eye, comes to believe that there is a fly on the window pane, in such a way that it is the fly's being on the window pane which produces this belief. Here P's belief is true; it is not "accidentally" produced, as in the example Grice gives in The Causal Theory of Perception¹⁴ of the clock on the shelf, and yet P would deny that he knew that there was a fly on the window pane, on the grounds that he thinks it possible that what he saw out of the corner of his eye might have been a smudge of dirt and not a fly at all. The situation is one in which allegedly - P has knowledge that there is a fly on the window pane (conceived as non-accidental true belief that there is a fly on the window pane), and yet P himself is disposed to deny that the knowledge is his. The mental sentence is true, but the subject disavows it. (The example can be re-formulated so as to apply to noticing, learning, hearing, remembering, etc.)

But this objection, and the others it can be made to generate, is not, it seems to me, at all conclusive. It is not conclusive because it does not seem at all plausible to ascribe knowledge to a person in a situation of the kind described. Even if knowing that p does require non-accidentally coming to believe truly that p, it is doubtful whether this is everything it requires, and this is a fact which (as I see it), the described situation has the merit of showing. Our concept of knowledge is more adequately filled out by saying, in addition, that if one person correctly ascribes another person with knowledge that such-and-such, then some implication is contained that the ascriber can and does assess himself as the possessor of that piece of knowledge. (This is clearest of all when a person correctly ascribes himself with knowledge that such-and-such, for in this case he obviously cannot do so unless he is in a position to assess himself as the

14 Grice, PASS 35, 1961. In Warnock, ed., The Philosophy of Perception (Oxford 1967). The relevant example appears on pp. 103-4.

possessor of that knowledge. The assessment and the ascription go hand-in-hand). Rather than deflating our theory, the objection merely has the effect of bringing to light one respect in which the theory that is knowledge is non-accidental true belief is incomplete.

It is clear that any fact or situation offered as a counter-example to the theory under consideration must, to succeed as a counter-example, at least be more credible than the provisions of the theory itself. The counter-example offered in the most recent objection failed to fulfill this requirement. On the face of it, the next two objections to be considered seem better placed in this respect. They belong together. The first concerns psycho-analysis, and the second concerns "subliminal perception".

Remember that the thesis to be defended is that in order for a mental sentence to be true, the person whose name occurs as its subject must think of himself, or be conscious of himself in the relevant sense, as a person of whom that mental sentence is true. Now someone might object that this is tantamount to a denial of most of the things said by Freud. Central to any psycho-analytical account of people's behaviour is the idea that many - indeed perhaps most - of the facts about a person's mental life are completely hidden from the person himself, so that, to revert to philosophical terminology, a mental sentence about a certain person can be true even though the person in question has no consciousness of its truth. Indeed, according to some schools of analysis, it is precisely because of the fact that people are "unconscious" of such truths about themselves that the phenomena which those truths report can be so painfully effective in structuring their behaviour.

Although it may seem conclusive, this objection is, in fact, difficult to assess. In the first place it is a matter of extreme and urgent controversy whether Freud's account of the mind is to be taken as a work of science, to be validated or invalidated by the usual procedures of test and

confirmation, or whether it should be seen in some other way. And secondly (a point which is more relevant at this juncture) there is the problem of deciding exactly what sense to attach to Freud's word "unconscious". Presumably it signifies a lack of consciousness, in one of the sense of "conscious" which I elucidated. But which? It is difficult to be sure that we have the correct answer to this question, but it seems rather plausible to think that the word "unconscious" in Freud's sense signifies only a lack of consciousness of the full-blooded sort, and not a lack of consciousness of the weak sort; so that my suggestion that desires, intentions and thoughts, etc., are necessarily the objects of consciousness in a weak sense is not at all incompatible with the Freudian idea that many desires and thoughts etc., are held unconsciously, if this word is understood as signifying a lack of consciousness of the strong, "full-blooded" sort.

The reason why it is plausible to say this is that one of the features of objects of weak consciousness is that they can be brought into full focus either by reminding the person in question, by pointing it out to him, or by enabling him to discover it for himself. I said that if a person is conscious of a fact about himself in the weaker sense, then that fact can be elevated into an object of full-blooded consciousness by one or another process of this latter kind. And the point here is that unconscious thoughts, desires, beliefs, etc., in the psycho-analysts' sense, are supposed to have precisely this feature. It would be fair to say that the entire institution of psycho-analytical practice presupposes that just such a process is available; that mental states and attitudes at one time labelled "unconscious" are of such a kind that they can be lifted into the full-blooded consciousness of their owner. Their range of efficacy then changes, of course, but this fact need not make any difference to the fundamental idea that mental states and attitudes which are unconscious in

the psycho-analytical sense are "present" to the mind at some level, before the analysis begins to lift them from it.¹⁵

So the Freudian objection, when understood, appears to have the surprising effect of endorsing my own suggestions about consciousness. The next objection is similar in this respect. It concerns cases of habitual or learned behaviour. In these cases, the objection says that it must presumably be true that we notice, see and hear things as part and parcel of performing the behaviour which we have learned or become habituated to, without having any kind of awareness that we do notice, see and hear the things we do. Indeed this seems to be part of what it means to act with accomplishment or skill, when the exercise of the skill is beyond the learning phase. In riding a bicycle or driving a car, for instance, we do not have to continually adjust our behaviour in the light of what we notice or perceive in the way we did while learning; and yet without our actually noticing or perceiving the relevant obstacles or obstructions we could hardly ride or drive successfully, i.e. without accident. So here it seems is another case in which mental sentences become true of people without their having any conception of themselves as being the subjects of those sentences.

In considering this objection we must of course leave out of account, if we can, the fact that, largely due to the nature of the tasks in question, we do not have to store the information gathered in perceiving and noticing things for a very long period of time. Although this is true we must leave it out of account, because the objection to be considered is not that we

15. In his essay on the theory of the emotions, J-P. Sartre remarked that "unreflective conduct is not unconscious conduct". (Sketch for a Theory of the Emotions, Methuen, 1962, p.61). Nor, obviously, is it reflective conduct. In analagous terminology, my suggestion as to the distinguishing feature of the mental might be supported in large measure by saying that mental activity is that activity a condition of whose occurrence is that the subject undertakes or undergoes it unreflectively (though not unconsciously).

"forget" or cease to have any long-term use for the knowledge we acquire in this type of mental activity, but that the acquisition of this knowledge takes place, whenever it does take place, without any awareness on the part of the subject that it is taking place. We are not being asked to consider the objection that the subject cannot testify as to his noticings and perceivings some time after they have taken place, but to the objection that however soon after they took place the subject simply has no conception or reflexive awareness that they are taking place "to" or "in" him.

Perhaps the paradigm case of habitual or learned behaviour where there are mental occurrences (to put it uncritically) of which the subject has no conception whatever is the case of speaking. I do not mean uttering or babbling. When we speak, at least according to Grice and the "communication theorists" of meaning, the speaker must be accredited with a complex and interconnected set of linguistic intentions, such as the intention to produce a belief of a certain kind in his audience, or the intention to get his audience to recognise a certain intention of his (the speaker's); and yet any adult or reasonably skilled user of the language has no conception of himself, at the time at which he speaks, as being the owner of any of them. Confronted with these cases, how can the thesis that the relevant kind of self-consciousness by the subject is a condition of his mental sentences' truth possibly be defended?

Again, this seems on the face of it a powerful objection; and yet I think that there are various defenses available. Some are better than others. The first, which I merely mention without using, says that subliminal perceptions and unrecognised intentions can only exist because non-subliminal perceptions and recognised intentions exist; so that although there are cases of perceiving and intending where the subject has no awareness of himself as someone who is perceiving or intending, these are strictly dependent upon the cases where the subject does have that

kind of self-conception. (Acting with skill is dependant upon having learnt the skill). So, subliminally perceiving and intending without recognising are cases developed from the central cases by extension. But even if this were true, it would not provide the defense which my proposition needs, since what that proposition says is that reflexive conscious is present in all cases in which a mental sentence is true of him, and not just in the central cases.

The second defense is designed to support that exceptionless proposition. It is similar to the one deployed against the psycho-analytical objection.

In order to make it credible that there is a sense in which a person who subliminally perceives things is conscious of himself as the subject of those perceptions, it is only necessary, I think, to bring into prominence the differences between a person subliminally perceiving something and a person in a state of complete coma. Consider a very familiar and mundane situation in which a series of subliminal perceptions might occur; the activity of bicycle riding. Someone who has learnt to ride a bicycle, and who is competent in doing so, will normally deploy the various skills needed without being fully conscious of the fact that he is deploying them. The individual motions of his body which are necessary to balance the machine successfully will not have to be initiated by him in quite the self-conscious way in which they had to be while he was learning to ride. Indeed it is well known that too much attention to the various necessary motions is likely to cause the skilled cyclist to lose his balance. It has to be admitted that in this sense the skilled bicyclist sublimates his perceptions of his own balancing movements; and doubtless too he sublimates many of the perceptions he makes of the various obstacles in his path. But on the other hand his cognitive attitude towards these perceptions is not at all like the attitude of one who is bicycling in a

coma, who is perching on the saddle and avoiding crashing just by good luck. It is not the case, in other words, that he has no cognitive attitude towards them.

It is upon this relatively uncontroversial point that the evidence for his having some kind of weak consciousness of his perceptions is based. The argumentation is the same for the case of the supposedly unrecognised intentions which are said to attend the ordinary communicative use of language. Probably no one except the most expert philosophers of language have any very clear idea of the different and interconnected intentions which attend the various types of speech behaviour. And yet someone who uses language effectively is not at all like the person who utters sentences in his sleep or under an anaesthetic. As part of their attempt to describe the difference between the two types of case, philosophers have suggested that an account of the various intentions which attend the ordinary speech situation gives "an order which is there".¹⁶ Those (like myself) for whom this remark fails to provide a very satisfying description of the situation will, I suspect, prefer to say that such intentions, though not present to the full consciousness of their owner, are none the less objects of that weaker attitude of consciousness for whose existence I have been trying to expound the evidence.

I finally mention a more general objection to my suggestion about mental sentences. It runs as follows. It cannot be the case that, in order for a mental sentence about P to be true, P must see himself as being the subject of that truth, because he cannot see himself in that way unless the mental sentence is in fact true. Rather than it being a condition of the mental sentence's truth that the subject recognises its truth, the situation is quite the reverse, in as much as its truth is a condition of

16. Anscombe: Intention (Oxford, 1957), p. 80

his being able to recognise it as true. What this objection has the effect of making clear, however, is not so much a defect in my suggestion as an interesting and (I suspect) important similarity between mental sentences and sentences containing "performative" verbs. In both classes of cases, one condition for a sentence's truth is that the subject of the sentence thinks of himself, or is conscious of himself, as having that sentence true of him. If someone says "Jones promises that p", then in order for his statement to be true, i.e. in order for it to be true that Jones does promise that p, Jones must be conscious of himself as the bearer of the predicate ".....does promise that p". It is especially clear that this is the case when the speaker is Jones himself. If someone says "Jones believes that p", likewise, then in order for his statement to be true, i.e. in order for it to be true that Jones does believe that p, Jones must be conscious of himself as the bearer of the predicate "... does believe that p". Again, this is most clearly true in the case where Jones himself is the speaker. In both examples, Jones must be conscious of himself under the description in question in order for the description to fit him.

The discussion of this objection brings my remarks on the distinction between mental and the physical sentences to an end. Were I to have defended my proposal simply by trying to abolish the most obvious counter-examples, I should probably have failed to make out a convincing case for it. What I have attempted to do, by contrast, is to discuss the various counter-examples in such a way as to gradually fill out the description of that kind of consciousness which, as I have proposed, the subject of any true mental sentence has towards himself as the subject of such a sentence. In this proposal, although the terminology is different, there is more than an echo of that passage in Locke's Essay which says:

"....Such are perception, thinking, doubting, believing, reasoning, knowing, willing, and all the different actings of our own minds; - which we being conscious of, and observing in ourselves, do from these receive into our

understanding as distinct ideas as we do from bodies affecting our senses. This source of ideas every man has wholly in himself; and though it be not sense, as having nothing to do with external objects, yet it is very like it, and might properly enough be called internal sense". 17

The proposal about the distinction between the mental and the physical which I have just defended is of relevance to two other theses which I have defended in this essay. The first is that the mental is irreducible to the behavioural (where behaviour is conceived as mere motion); and the second is that the mental is in all likelihood reducible to the cerebral, either at the Property-Level or at the slightly less general Person-Specific Level. The demonstration of this relevance will bring the essay to a close.

To say that the mental and the behavioural cannot be connected in a statement of natural law is to say that it is not possible for a mental sentence and a behavioural sentence to match in truth-value in every physically possible world. And to say on the other hand that the mental and the cerebral can be connected in a statement of natural law is to say that it is possible that a mental sentence and a cerebral sentence are so connected - i.e. that it is possible for a mental sentence and a cerebral sentence to match in truth-value in every physically possible world. Up to this point little has been said to explain why this should be so. But we are now in a position to do just this. For the characterisation of the mental which it has been the purpose of this section to recommend would, assuming it to be correct, have the natural consequence not only that the behaviour which attends any specific mental condition varies from one occasion of the mental condition to another, but also that the cerebral condition which attends any specific mental condition is not likely to vary

17. John Locke, Essay Concerning Human Understanding, ed. A. C. Fraser, (Oxford 1394), Bk. II, Ch. 1, Sec. 4.

in the same way at all. Because any given mental condition, while there will be a range of possible behaviour available, there will be no variability at all in the subject's cerebral condition.

In what sense are these consequences natural? I advanced the idea that any mental sentence always requires, as one condition of its truth, an additional truth to the effect that a certain kind of self-consciousness is present in the subject, whereas physical sentences are those which never require this. I am presently contending that the nature of this reflexive consciousness explains why a behavioural sentence and a mental sentence will never match in truth-value, and that it does so in the following way. While to be conscious (in my sense) of a mental condition of one's own is not necessarily to have it "in focus", to be fully and explicitly aware of it, or to be able to express it linguistically - these are the features of what I called "full-blooded" consciousness - it is to be in such a state that one can bring it into focus and full awareness, or have other people do it for you under the right kind of manipulation. And the importance, in turn, of the ability to selectively focus upon one's own mental states is that it is just this ability which underlies people's capacity to behave as people; that is, to adjust and decide their actions in the light of the whole range of their present desires, thoughts, beliefs, and other mental conditions.

Indeed a strong case can be made out for saying that we are persons to just the degree to which we do selectively focus upon our thoughts and desires, etc., mutually adjust them when they conflict, alter them in the face of reasonable persuasion, open ourselves to rational criticism and moral evaluation, and, finally, cite them in explanation of the very different actions which we initiate. All this is to say that consciousness of one's own mental condition, in the sense of my exposition, is part and parcel of being able to behave as a person.

This, very roughly, is the way in which the presence of the relevant type of consciousness of one's own mental condition explains the variation which is found in the behaviour, if any, which those same mental conditions can be used to explain - the variation which, as we saw in Chapter I, directly spells the failure of any mental-to-behavioural reduction. And on the other hand, it seems clear that this type of self-consciousness of one's own mental conditions need have no analagous effect upon the cerebral conditions which accompany them. By contrast, if a person were an automatic mechanism of a sort such that any specific mental condition directly caused a certain specific type of limb-movement, then no doubt a behavioural reduction would stand the same chance of success as a cerebral reduction now stands. But this is manifestly not what a person is.

Appendix

IDENTIFICATION AND EXPLICATION

Introduction

In the section of the Chapter I entitled "Alternatives to a Behavioural Reduction" I mentioned an analytical device employed by Quine (and by those who follow him) which, though neither reductive nor non-reductive in the senses there distinguished, merits serious consideration on the grounds that it supposedly provides a way of analysing the mental in terms of the physical. Applied to a range or class of individuals, the device is normally called "explication"; while applied in particular instances it is more ordinarily called "identification" or "identification with ...".

Examples of explication and an explanation of the method in general.

If in a chess game I replace the black king with a button, for instance, then we can say that for the duration of the game the button is the black king - or, to use a different terminology, that the black king has been identified with the button. This device of "identification with" has nothing to do with strict identity or Leibniz's law, as Quine's example of the ordered pair $\langle x, y \rangle$ shows; in this case either $\{\{x\}, \{y, \wedge\}\}$ or $\{\{x\}, \{x, y\}\}$ or $2^x 3^y$ or $3^x 2^y$ can be identified with the ordered pair $\langle x, y \rangle$, while they are all clearly distinct from each other.¹ There is even a difference in ordinary speech corresponding to the difference between this sort of identification and cases of "strict" identity. In the former case we speak of identification with, and the latter case we speak of identification as: the small boy who identifies himself with Dan Dare is just being playful, but if he identifies himself as Dan Dare then he needs

1. Word and Object, paras. 53, 54.

psychiatric counsel.

In Chapter I we saw some more systematic cases of "identification with", which involved the introduction of classes. Classes very often play the key role: for Bloomfield, for example, phonemes were to be identified with classes of phones, morphemes with classes of phonemes, and sentences with classes of morphemes. And for Fodor, as we saw, mental states were to be identified with classes of neurological states. Quine's own legacy from the Bloomfieldian tradition in linguistics is considerable; we find him saying in *Word and Object* that phonemes

"... are sometimes construed as the classes of their approximations. In representing them rather as segments of norms I stress the qualitative clustering about statistical norms, and minimise the suggestion of an enclosing boundary. But we can still think of each norm as the class of the events that are occurrences of it" 2

and later in the book the method is extended to those abstract object sentences, in a way which Bloomfield himself might have done:

"A sentence is not an utterance event but a linguistic form that may be uttered often, once, or never; and its existence is not compromised by failure of utterance. But we must not accept this answer without considering more precisely what these linguistic forms are. If a sentence were taken as the class of its utterances, then all unuttered sentences would reduce to one, viz. the null class;

Nor should I like to take a sentence as an attribute of utterances But there is another way of taking sentences and other linguistic forms that leaves their existence and distinctness uncompromised by failure of utterance. We can take each linguistic form as the sequence, in a mathematical sense, of its successive characters or phonemes. A sequence a_1, a_2, \dots, a_n can be explained as the class of n pairs $\langle a_1, 1 \rangle, \langle a_2, 2 \rangle, \dots, \langle a_n, n \rangle$. We can still take each component character a_i as a class of utterance events, there being here no risk of non-utterance" 3

As a matter of fact this procedure fails to provide for the distinctness of different sentences, even when elaborated to this extent; for example,

2. Word and Object, pp. 89-90.

3. Word and Object, pp. 194-5.

(1) Let the meat cook

(2) Let them eat cook

would by Quine's standard be the same sentence. However my purpose in quoting these examples of identification with is only illustrative; the essential component of this method, and the one we must concentrate on, is to hit on an effective function which associates with each member of a certain category of objects some other construction - in practice usually set-theoretic - such that the arguments are mapped by the function onto its values in such a way as to preserve, among the value-objects, all the important differences which obtain among the argument-objects.

It may be difficult to grasp the exact purport of this abstract statement; but the procedure can be explained more fully as follows - and I continue to take the case of sentences (which are abstract objects) as an example. You start from a nominalistic premiss, to the effect that some objects are "concrete" and "simple" while some are abstract. You then explain or in a sense define an abstract object as being a class whose members are specified by an open sentence, according to this schema:

Abstract object 0 is the class whose members are all the simple or concrete things x such thatx.....

So for instance the abstract object redness might be introduced like this:

Redness is the class whose members are all the things x such that x is red.

and in the case of sentences, the general formula would be:

Sentence S_n is the class of things $\langle x_i, y_i \rangle$ such that x_i is the i th character of S_n and such that y_i is the number i .

where in addition, we explain a character and a number in set-theoretical terms so as to expand the definition to:

Sentence S_n is the class of things $\langle x_i, y_i \rangle$ such that x_i is the i th character of S_n and such that y_i is the number i , where the i th character x_i of S_n is the class of things w such that w is an utterance-event of x_i and where the number i is the class of things z such that z is an ordered i -tuple of objects.

Such effective functions as this - assuming that this is an effective function - are called by Quine proxy functions. Before we try to fathom the efficacy of this method in general, let us see how it would specifically be applied to the mental. Roughly, the idea would be to consider sentences like

(3) John believes that swans fly

(4) The cat wants to be on the roof etc

as relating John or the cat to a class of possible worlds. The propositional attitude sentence (3) would, if true, relate John to the class of possible worlds in which swans fly; while the egocentric attitude sentence (4), if true, would relate the cat to the class of possible worlds in which it is on the roof. A possible world would be a kind of four-dimensional version of a possible world state - where a possible world state is a class of ordered triples of natural numbers, each triple of which identifies a matter-occupied point.⁴ So in this case we start with points

Footnote 4 printed overleaf

of matter-occupied space as the basic individuals, and with the ordered triples of natural numbers which identify them. Then once having elaborated the notion of a possible world, we say that a class of possible worlds is to be identified with the object of an organisms want: the state of affairs which his having a certain attitude relates him to. The theory would finally be elaborated, perhaps, (and here I take leave of Quine) in the following way. Penultimate step: complex objects like wanting to be on the roof or trying to get onto the roof, believing that swans fly, etc., would each be identified with an ordered pair of classes consisting of the class of organisms who have that attitude to that state of affairs, and the relevant class of possible worlds. Even more complex objects, finally (this is the ultimate step), like the cats wanting to get onto the roof, or John's believing that swans fly, would perhaps be identified with an ordered pair having the cat or John as first member, and the previous ordered pair as second member.

This is the kind of elaboration which Quine's method (suggests) for explicating mental entities. Fodor's method of identifying mental states with classes of neurological states is more straightforward, in that it does not go as far as to invoke the complex apparatus of possible worlds. I now list three defects in ascending order of generality. They all pertain to Quine's possible-worlds analysis; the third, which is the most general, should trouble Fodor's proposal as well.

Footnote 4 brought forward from page

In discussing the example and explaining this system in Propositional Objects (Ontological Relativity pp. 139-160), Quine gives this preliminary notion the following refinements. In order to accommodate rotation of the number-axes, a possible world state is first explained to be the class of classes of ordered natural-number triples which identify congruent matter-occupied regions. In order to accommodate different measuring systems, a possible world state is then explained as: a class of classes of ordered natural-number triples which identify geometrically similar matter-occupied regions. See Quine's essay for other refinements.

Defects:

1. For the propositional attitudes - but not for the egocentric attitudes - the method fails to distinguish (say) believing that swans fly and believing that members of the sort *Cygnus Anstidae* fly; for the possible worlds in which these situations obtain are the same, although not every organism who has the first belief has the second. At least so we are prone to say. Quine acknowledges this difficulty: it is peculiar to the possible-worlds explication of attitudes which are propositional.

2. For egocentric and propositional attitudes alike, the method would fail to reveal the difference between (say) wanting a certain thing and fearing it, in the case where all the organisms who want that thing just happen to fear it as well. It is quite possible that they might - and yet the method of explication would (at the penultimate step) decree them to be identical. A fortiori the cats wanting to get onto the roof and the cats fearing to get onto the roof would (at the ultimate step) be the same. This difficulty is not peculiar to the mental: it is the familiar difficulty which arises with any set-theoretical or ordered pair analysis of relational expressions. Objection 2 is completely general.

3. When the method of identification - or explication - is used, the object to be explicated (the explicandum) is usually not a class, whereas the explicating object (the explicans) almost always is. And since alternative classes are available as the explicans, we cannot say that the explicans and the explicandum expressions are in any sense co-extensive. To take Frege's explication of numbers as a simple example: it just makes no sense to say that the expression "3" is co-extensive with the expression " $\{x \mid x \text{ is a triple}\}$ ". Explication is, on the other hand, "elimination" (as Quine puts it, Word and Object, p. 264); or in other words, as we could say, an explication which is systematic is an ontological reduction. I strongly doubt whether the word "reduction" is appropriate here: neither

it nor its complement seems to fit well. But explication is said to effect an elimination, or an ontological reduction, in the following way. Quine says that a theory Θ is reduced ontologically to a theory Θ' if we

"specify a function, not necessarily in the notation of Θ or Θ' , which admits as arguments all objects in the universe of Θ and takes values in the universe of Θ' . This is the proxy function. Then to each n-place primitive predicate of Θ , for each n, we effectively associate an open sentence of Θ' in n free variables, in such a way that the predicate is fulfilled by an n-tuple of arguments of the proxy function always and only when the open sentence is fulfilled by the corresponding n-tuple of values" 5

Effectively and truth-preservingly mapping the closed sentences of Θ onto the closed sentences of Θ' is no good by itself, Quine reminds us, since any sentence S of Θ can be associated with a sentence "xT" with x as the Gödel number of S and T the truth-predicate of Θ , thus trivially reducing any theory whatever to a theory of natural numbers. And an effective association of predicates of Θ with predicates of some theory Θ' is no good by itself either - since the Löwenheim-Skolem theorem says that any theory whatever can be modelled in the natural numbers. It was these trivialising results which prompted Quine to say that an extra condition has to be met - this being the specification of a proxy function, which effectively assigns each of the objects of the reduced theory Θ to some particular object of the reducing theory Θ' . The Löwenheim-Skolem theorem, by contrast, contains no proxy function; it does not determine which numbers are to be associated with the objects of the theory to be reduced.

Now my previous examples of identification, or explication, are examples in which proxy functions are supplied. My third criticism of the entire programme - to come to it at last - is that proxy functions are bound to be either non-effective or else circular. So if I am right, the apparatus of explication is useless as a device whereby to analyse the

5. Quine, Ontological Reduction and the World of Numbers, p. 205.

mental in terms of the physical. And the criticism entails that in the general case there is no such thing as ontological reduction in Quine's sense of the phrase.

I shall now explain these criticisms. The first thing to notice is that it is always possible to state any proxy function in a circular way; for instance you can say that

$F'(number\ N) = \text{the class of all } N\text{-membered classes}$

or that

$F''(\text{sentence } S_n) = \text{the class of things } \langle x_i, y_i \rangle \text{ such that } x_i \text{ is the } i\text{th character of } S_n \dots \text{ etc. (See above)}$

or, to cite Quine's own example in "Ontological Reduction and the World of Numbers", that

$F(n^\circ C) = n$

For the case where " $H(x, n^\circ C)$ " says "the temperature of x is n degrees centigrade"; where " $H_c(x, n)$ " says - equivalently - "the temperature in degrees centigrade of x is n "; and where the problem is to try and eliminate impure numbers (e.g. five degrees centigrade) in favour of pure numbers (e.g. five) only. But the important question is whether a non-circular statement exists in each case, which still preserves the effectiveness of the function. In the number case, according to Russell, at least, such a non-circular statement does exist;⁶ but the other cases are, I maintain, more doubtful. Consider the case of sentences. Not every sequence of characters is a sentence; indeed it seems to me that the only way of isolating the relevant (i.e. sentential) sequences of characters is

6. Russell, Introduction to Mathematical Philosophy, Ch. 2

by beginning at the outset with a knowledge of what sentences there are. And it is clear that unless there is an independent way of sorting the sentential sequences from the non-sentential sequences, there is no way of specifying an adequate proxy function which does not directly rely, in the process of assigning values to arguments, upon the very notion to be analysed, (the term mentioned in the value-position being contained in the term mentioned in the argument-position).

This is even more obviously the case with the function $F(n^{\circ}\text{C}) = n$, of the temperature example. And in the case of a proxy function for mental entities, the situation is the same. We have to say that complex objects like having a certain attitude to a state of affairs takes as values of the function an ordered pair of classes consisting of the class of organisms who have that attitude to that state of affairs. Any effective assignment along these lines will have to proceed, I think, by invoking, in the process of specifying the assigned concept, the concept to which it is being assigned. (The problem is just the same with Fodor's proposal that mental states are to be identified with classes of functionally identical neurological states. For here, in the case of any specific mental state, the shared function which the relevant neurological states have would be specified by reference to the mental state in question, in just the same way, if we are to take Fodor's "camshaft analogy" seriously (see Chapter II), that the shared function of those mechanisms which are to be identified with a valve-lifter is specified as that of lifting valves).

Indeed, I think it is rather likely that the problem runs even deeper than this. For in order for values to be assigned to arguments, there must be both the objects which are arguments as well as the objects which are values. But then in what sense is anything eliminated? (Explication is elimination, according to the slogan). The temperature example makes

this general paradox clear. Either there are impure numbers or there are not.⁷ If there are not, then " $H(x, n^{\circ}C)$ " fails to represent the logical form of "the temperature of x is n degrees centigrade", and there is no function (therefore) which can begin " $F(n^{\circ}C)=...$ "; But if there are impure numbers, then the proxy function $F(n^{\circ}C)=n$ only succeeds in effectively assigning impure numbers to numbers by assuming impure numbers to exist, and by associating pure numbers with them on that very basis. Curiously, Quine admits this defect, when he says

"I must admit that my formulation suffers from a conspicuous element of make-believe..... I had to talk as if there were such things as $K^{\circ}C$ "⁸

But I cannot myself see how the defect can be admitted seriously, without at the same time admitting that the analysis deserves to be renounced. This is indeed an instance of the paradox of analysis: if objects like sentences, impure numbers, or having an attitude to a state of affairs are arguments of a proxy function, then their existence and the clarity of their distinctness conditions must be assumed in advance. And in the light of this consideration the idea of eliminating certain objects in favour of certain others begins to look somewhat absurd.

Final Remarks

The analysis by explication of such things as numbers, sentences, impure numbers and mental entities is a questionable procedure, because, as I have argued, the goal of analysis itself becomes opaque to understanding when it is realised that assumptions have to be made about the existence and identity of the analysandum objects in order for the analysis itself to get under way.

7. I assume at this point that not both " $H(x, n^{\circ}C)$ " and " $H_c(x, n)$ " represent the correct logical form of the sentence in question.

8. Quine. Ontological Reduction. The World of Numbers, p. 206.

Quine himself, of course, sees it differently. He considers that an analysis by explication is motivated in the first place by the fact that the analysandum objects do not have clear existence-conditions:

"We have, to begin with, an expression or form of expression that is somehow troublesome. It behaves partly like a term but not enough so, or it is vague in ways that bother us, or it puts kinks in a theory or encourages one or another confusion. But also it serves certain purposes that are not to be abandoned. Then we find a way of accomplishing those same purposes through other channels, using other and less troublesome forms of expression. The old perplexities are resolved" 9

Philosophers interested in nominalistic analyses have a partially similar view of the process of analysis: some sentences can be easily formalised in predicate logic, and these describe concrete individuals and perhaps single events; while others resist the formalism and describe (or purport to describe) abstract objects or objects whose existence is less fully intelligible. The problem for them, again, is one of effectively devising a substitute, from the easily formalisable part of the language, for the recalcitrant terms or expressions. Now even if it were possible, devising such a substitute for a recalcitrant term - either to "eliminate" or to "introduce" it - would not, in any case, add to the predictive or explanatory power of the easily formalisable theory, the theory whose terms describe simple or concrete objects; since as Putnam once suggested, Gödel's proof of the completeness of predicate logic ensures that all implications expressible in the nominalistic language can be proved in that language however many new set-theoretical constructions are introduced via proxy functions - or however few.

Perhaps the set-theoretical constructions out of basic individuals enable us to give the meaning of the non-basic language which resists the

9. Word and Object, p. 260.

formalism of predicate logic.¹⁰ In any event the methodological standpoint is the same: you have clear and formalisable terms and easily distinguishable objects on the one hand, and unclear and unformalisable terms and shadowy objects on the other; and you then set yourself the task of analysing the latter by devising a substitute in terms of the former.

My reaction to this procedure of Quine's and the nominalists - to sum up now - has been to say that in the general case you cannot devise a function which effectively takes shadowy objects into clear ones without appealing to your shadowy notion in the specification of the correct clear notion which is to be substituted for it; for notions thought to be shadowy are the very ones upon which we so often rely in grouping and categorising things which are thought to be basic and, relatively speaking, discrete. (E.g., the case of sentences and sequences of characters; the case of mental attitudes and classes of possible worlds). And even the idea of a shadowy or logically recalcitrant notion has in practice to be abandoned; for to make a proxy function properly effective, you have to incorporate assumptions about the existence and identity of the objects whose analysis is being undertaken. A proxy function effectively relates objects in the universe of one theory to objects in the universe of another;¹¹ but if the former set of objects exist anyway, then what could the function possibly succeed in achieving?

10. See Putnam: Mathematics and the Existence of Abstract Objects (Phil. Studies 1956). His suggestion about the meaning of the abstract or recalcitrant terms was that they "express all the implications that they entail" (p. 35). If a non-basic sentence S (i.e. a sentence with a term for an "abstract" object) conjoined with a basic sentence e entails that e', then one of the implications that S expresses is that $e \supset e'$; although, by Gödel's result, if $e \supset e'$ is valid, it can be proved without considering S.

11. See the quotation in my text which is identified by footnote 5, above.

Bibliography

- Achinstein, P. and Barker, S. (eds.) The Legacy of Logical Positivism. Baltimore, 1969.
- Anscombe, G. E. M., Intention. Oxford, 1957.
- Armstrong, D. M., A Materialist Theory of the Mind. London, 1968.
- Austin, J. L., "Ifs and Cans", pp. 205-232 In Urmson and Warnock (eds.).
- Ayer, A. J., (ed.). Logical Positivism. New York: The Free Press, 1959.
- Black, M., (ed.). Philosophy in America. London, 1965.
- Borst, C. V., (ed.). The Mind/Brain Identity Theory. London, 1970.
- Brentano, F., "The Distinction between Mental and Physical Phenomena".
In Psychologie Von Empirischen Standpunkt, 1874.
Also in Chisholm, (ed.).
- Bromberger, S., "An Approach to Explanation". In Butler, R. J., (ed.).
- Butler, R. J., (ed.). Analytical Philosophy (2nd Series). Oxford, 1968.
- Capitan, W. H., and Merrill, P., (eds.). Art, Mind and Religion. Proceedings of the 1965 Oberlin Colloquium in Philosophy. Pittsburgh University Press, 1967.
- Carnap, R., Der Logische Aufbau der Welt. Berlin-Schlachtensee Weltkreisverlag, 1923.
- Carnap, R., "Psychologie in Physikalischer Sprache". Erkenntnis, 1933.
Reprinted in Ayer, (ed.): Logical Positivism as "Psychology in Physical Language".
- Carnap, R., "Testability and Meaning". Philosophy of Science 3, 1936;
Philosophy of Science 4, 1937.
- Carnap, R., "The Methodological Character of Theoretical Concepts". In H. Feigl and M. Scriven (eds.), pp. 33-76.
- Chisholm, R., Perceiving: A Philosophical Study. New York, 1957.
- Chisholm, R., (ed.). Realism and the Background of Phenomenology. Allen and Unwin, 1967.
- Chomsky, N., Syntactic Structures. The Hague, Mouton, 1957.
- Chomsky, N., "Perception and Language", pp. 199-205, in Wartofsky, (ed.).
- Chomsky, N., Aspects of the Theory of Syntax. M.I.T. Press, 1965.
- Chomsky, N., "Current Issues in Linguistic Theory", in Fodor and Katz (eds.) pp. 50-113.

- Chomsky, N., "Remarks on Nominalisation", in Jacobs and Rosenbaum (eds.).
- Clark, R., "Concerning the Logic of Predicate Modifiers", Mous, 1970.
- Danto, A., and Morgenbesser, S., (eds.). Philosophy of Science, New York, 1960.
- Davidson, D., "Actions Reasons and Causes". Journal of Philosophy, 1963.
- Davidson, D., "On Events and Event-Descriptions", in Margolis (ed.).
- Davidson, D., "Mental Events". In Swanson and Foster, eds.
- Davidson, D., "The Logical Form of Action Sentences". In Rescher (ed.).
- Davidson, D., "Agency" [Unpublished manuscript]
- Davidson, D., "Psychology as Philosophy" [Unpublished manuscript]
- Dennett, D. C., Content and Consciousness. London and New York, 1969.
- Descartes, R., Meditations.
- Feigl, H., and Scriven, M., (eds.). Minnesota Studies in the Philosophy of Science, Vol. I. University of Minnesota Press, 1956.
- Fodor, J. A., "Explanations in Psychology". Pp. 161-179 in Black (ed.).
- Fodor, J. A., Psychological Explanation. New York, 1963.
- Fodor, J. A., "Three Reasons for not Deriving "Kill" from "Cause To Die"". Linguistic Inquiry I, 1970.
- Fodor, J. A., and Block, N. J., "What Psychological States are Not". Philosophical Review, April, 1972.
- Fodor, J. A., and Katz, J. J., (eds.). The Structure of Language. Englewood Cliffs, 1965.
- Fraser, B., "Some Remarks on the Action-Nominalisation in English". In Jacobs and Rosenbaum (eds.).
- Geach, P. T., "Some Problems about Time". In Strawson (ed.).
- Grice, P., "Meaning". Philosophical Review 66, 1957.
- Grice, P., "The Causal Theory of Perception". Proceedings of The Aristotelian Society (Supplementary Vol.) 35, 1961. Also in Warnock (ed.).
- Grice, P., "Utterer's Meaning, Sentence-Meaning and Word-Meaning" Foundations of Language 4, 1963.
- Hampshire, S., Feeling and Expression. London, 1961.
- Harman, G., "Knowledge, Reasons and Causes". Journal of Philosophy, 1970.

- Harris, Z., "Discontinuous Morphemes", Language 21, 1945.
- Harris, Z., "Morpheme to Utterance". Language 22, 1946.
- Harris, Z., "Distributional Structure". Word 10, 1954.
- Harris, Z., "Phoneme to Morpheme". Language 31, 1955.
- Hempel, C. G., "Logical Positivism and the Social Sciences". In Achinstein and Barker (eds.).
- Hook, S., (ed.). Dimensions of Mind. Collier Books, New York, 1961.
- Jacobs, R. A., and Rosenbaum, P. R., Readings in English Transformational Grammar. Waltham, Mass., 1970.
- Katz, J. J., "Mentalism in Linguistics". Language 40, 1964.
- Kim, J., "On the Psycho-Physical Identity Theory". American Philosophical Quarterly, 1966.
- Lakoff, G., "The Nature of Syntactic Irregularity". In Mathematical Linguistics and Automatic Translation. Report no. NSF-16, Computation Laboratory of Harvard University, 1965.
- Locke, J., Essay Concerning Human Understanding. Edition: A. C. Fraser, Oxford, 1894.
- Margolis, J., (Ed.). Fact and Existence, Proceedings of the University of Western Ontario Colloquium 1966 (Oxford 1969)
- Melden, A. I., Free Action. London, 1961.
- Nagel, T., "The Meaning of Reduction in the Natural Sciences". In Danto, A., and Morgenbesser, S., (eds.).
- Nagel, T., "Physicalism". Philosophical Review, 1965. Also in Borst (ed.).
- Neurath, O., Einheitwissenschaft und Psychologie. Vienna: Gerold and Co., 1933.
- Neurath, O., Foundations of the Social Sciences. University of Chicago Press, 1944.
- O'Connor, J., (ed.). Modern Materialism: Readings on Mind-Body Identity. New York, 1966.
- Place, U. T., "Is Consciousness a Brain Process?" British Journal of Psychology, 1956.
- Price, H. H., "Some Objections to Behaviourism". Pp. 79-84 In Hook, (ed.).
- Putnam, H., "Mathematics and the Existence of Abstract Objects". Philosophical Studies, 1956.
- Putnam, H., "Minds and Machines". In S. Hook, (ed.).

- Putnam, H., "The Mental Life of Some Machines". In J. O'Connor (ed.).
- Putnam, H., "Psychological Predicates". In Capitan and Merrill, (eds.).
- Putnam, H., "Is Semantics Possible?" Metaphilosophy I, 1970.
- Quine, W. V., From a Logical Point of View. New York: Harper Torch book, 1963.
- Quine, W. V., Word and Object. Cambridge, Mass. M.I.T. Press, 1960.
- Quine, W. V., "Ontological Reduction and the World of Numbers". In The Ways of Paradox, pp. 199-207.
- Quine, W. V., The Ways of Paradox and Other Essays. Random House, New York, 1966.
- Quine, W. V., "Propositional Objects". In Ontological Relativity, pp. 139-160.
- Quine, W. V., Ontological Relativity and other Essays. Columbia University Press, 1969.
- Quinton, A., "Kind and Matter". In J. P. Smythies, (ed.), pp. 201-33.
- Rescher, N., (ed.). The Logic of Decision and Action. Pittsburgh University Press, 1967.
- Russell, B., Introduction to Mathematical Philosophy. London, 1970.
- Sartre, J.-P., Sketch for a Theory of the Emotions. Methuen, 1962.
- Searle, J., "Chomsky's Revolution in Linguistics". New York Review of Books, June 29th, 1972.
- Skinner, B. F., The Behaviour of Organisms. New York: Appleton-Century-Crofts, 1938.
- Smart, J. J. C., "Sensations and Brain Processes". Philosophical Review, 1959.
- Smart, J. J. C., "Materialism". Journal of Philosophy, 1963.
- Smith, C., "On Causative Verbs and Derived Nominals in English". Linguistic Inquiry I, April, 1972.
- Smythies, J. P., (ed.). Brain and Mind: Modern Concepts of the Nature of Mind. New York Humanities Press, 1965.
- Strawson, P. F., Individuals, London, 1959.
- Strawson, P. F., (ed.). Studies in the Philosophy of Thought and Action. Oxford University Press, 1963.
- Strawson, P. F., Meaning and Truth. Oxford, Clarendon Press, 1970.
- Swanson, J. W., and Foster, L., (eds.). Experience and Theory. University of Massachusetts Press, 1970.

- Taylor, B., "Mental Events: Are There Any?" Australasian Journal of Philosophy, December, 1973.
- Taylor, C., The Explanation of Behaviour. London, 1964.
- Urmson, J. O., and Warnock, G. J., (eds.). J. L. Austin: Philosophical Papers. Oxford University Press, 1970.
- Warnock, G. J., (ed.). The Philosophy of Perception. Oxford, 1967.
- Wartofsky, (ed.). Boston Studies in the Philosophy of Science, 1961/2.
1963.
- White, A., Philosophy of Mind. Random House, New York, 1967.
- Wieser, N., and Schele, J. P., (eds.). Nerve, Brain and Memory Models.
New York, 1963.
- Zeman, J., "Information and the Brain". In N. Wieser and J. P. Schele,
(eds.).