PHILOSOPHY IN AMERICA

Essays

William P. Alston
Bruce Aune
S. F. Barker
Stanley Cavell
Thompson Clarke
Marshall Cohen
Joel Feinberg
Jerry A. Fodor
Charles D. Parsons
Alvin Plantinga
John Searle
Abner Shimony
Fred Sommers
Judith Jarvis Thomson

Edited by
MAX BLACK

CORNELL UNIVERSITY PRESS
ITHACA : NEW YORK

my arm?'[1] the notions of ascriptiveness and defeasibility can provide no answer. Here as elsewhere in philosophy, analytic techniques help to answer the penultimate questions, while the ultimate ones, being incapable of *answer*, must be come to terms with in some other way.

[1] *Philosophical Investigations* I, 621.

VIII

EXPLANATIONS IN PSYCHOLOGY[1]

by JERRY A. FODOR

Assistant Professor, Department of Humanities, Massachusetts Institute of Technology

In this paper I will try to say what a psychological explanation is. This project should be distinguished from others to which it is indirectly related. Thus, I shall not be trying to settle the mind-body problem, nor shall I examine the alleged incompatibility between freedom of choice and the existence of psychological laws. What I shall have to say will be relevant to those problems only in this respect: Philosophers who have argued that psychology could (or could not) account for consciousness or for choice have sometimes supported their arguments by reference to features psychological explanations are alleged to have: that they employ causal laws, that they are concerned only with motions, that they are concerned only with aberrant behaviour, that they consist solely in the delineation of stimulus-response connections, and so on. In so far as philosophical claims have been based upon such analyses of psychological explanation, what I have to say should be relevant to assessing those claims. Moreover, there is at least one philosophical issue to which this paper is directly relevant. It is sometimes said that the programme implicit in the doctrine of the unity of science cannot be carried through unless it is possible to reduce the concepts employed in psychology to neurological concepts. We shall see that, though such reduction is not possible in principle, this fact is nevertheless compatible with the unity of science.

In so far as psychology affords explanations of behaviour, saying what a psychological explanation is involves saying what it is to explain behaviour.[2] However, not all explanations of behaviour are

[1] This paper has been influenced by several discussions of psychological explanation, and not least by those with which it explicitly takes issue. I wish to acknowledge a particular indebtedness to J. A. Deutsch, *The Structural Basis of Behaviour*, Chicago, 1960, and Hilary Putnam, "Minds and Machines," in Sidney Hook, (ed.) *Dimensions of Mind*, New York 1960.

[2] Throughout this paper I shall follow the current psychological practice of using 'behaviour' in a much more general way than ordinary language would

psychological explanations. You bought the chocolate one and I want to know why. Well, because you prefer chocolate, because vanilla was more expensive, because chocolate keeps better, because you were asked to buy chocolate, because you felt like it. Any or all of these may do as explanations, for any or all of them may be what I want to know. None of them, however, is a psychological explanation. To say what a psychological explanation is involves distinguishing psychological explanations from such explanations as those.

'But surely what you propose to do would be a waste of effort? Psychological explanations are what psychology texts supply. If you want to know what a psychological explanation is, go and look.' Psychologists do not always agree about what sort of thing a psychological explanation is or about what sort of things are psychological explanations. Such disagreements are important because they affect the course of research and the constraints that psychological theories are required to meet. Lashley showed that the presence of conductive metal strips in the cortex of a chimpanzee did not materially interfere with shape recognition and hence that the 'fields' some gestalt theorists had supposed must function in visual perception could not involve macroscopic variation of the electrical potential of the chimpanzee's brain.[1] What, then, must we say about field theories of perception? That depends in part upon what we say about the status of theoretical constructs in psychological explanations, and, in particular, upon whether we hold such constructs admissible even when their identification with neurological states or processes seems unlikely or impossible. What is involved is a question about the constraints theories in psychology ought to be required to meet, hence a question about what a psychological explanation is.

appear to warrant. It is, perhaps, not an accident that ordinary language often fails to supply words sufficiently general to describe the subject-matter of a special science (Cf. the use of 'matter' and 'energy' in physics); among the insights a science may achieve is the discovery that phenomena that appear dissimilar to uninstructed intuition are susceptible of similar explanations and thus ought to fall within the domain of a single discipline. The fact that we must invent a term like 'matter' to say what physics is about is related to the fact that it is not *obvious* (for example) that the laws determining the trajectory of missiles also account for the orbit of the moon.

[1] Cf. Lashley, K. S., K. L. Chow, and J. Semmes, 'An Examination of the Electrical Field Theory of Cerebral Integration,' *Psychological Review*, Vol. 58, 1951, pp. 123–136.

'Psychological explanations are what psychology texts supply. If you want to know what a psychological explanation is, go and look.' An account of psychological explanation on which *no* psychological theory turned out to be an explanation would be *ipso facto* unacceptable. One must start by assuming some clear cases if one is to start at all. But we need not suppose even the clear cases immaculate. It will be no surprise if it turns out that the best available psychological theories could be improved by simplification, by integration with theories in related disciplines, and so on. One reason for wanting to characterize psychological explanation is that an acceptable account would afford a basis for the criticism and improvement of theories psychologists propose.

We want an account of psychological explanation that shows what makes the clear cases clear. One might say we are trying to discover the criteria psychologists use to assess the adequacy of psychological explanations, except that this formulation is misleading in two ways. First, it fails to do justice to the extent to which an account of psychological explanation may require reconstruction: the criteria psychologists use may, on some occasions or to some extent, be inconsistent, or unreasonable, or too weak, or too strong, and in such cases we would wish to substitute criteria that are consistent, and reasonable and just strong enough. Secondly, we must not confuse the task of saying what a psychological explanation is with that of saying how psychologists use the verb 'explain'. The former investigation is not linguistic in any of the usual senses of that term, nor do I suppose that the account of psychological explanation I will propose is analytically true by virtue of the meaning of 'explain'. That is, I reject the view that the metatheory of a science must consist solely of analytic statements. On the contrary, it may well be characteristic of psychological explanations that they presuppose the truth of some such empirical assumptions as: that all behaviour is directed towards drive reduction, or that it is under the control of the central nervous system, or that it tends towards the achievement of a state of equilibrium, or whatever. If this is the case, then such assumptions will be built into our characterization of psychological explanation: to explain behaviour will involve showing how it affects reduction of drive, how it is controlled by the central nervous system, or how it tends towards the establishment of an equilibrium.

Philosophers have often remarked that consonance with very general propositions about the world sometimes achieves the status of a necessary condition upon explanations in the sciences. But the conclusion they have drawn is only that such propositions serve as implicit definitions of key terms and are thus effectively analytic despite their empirical appearance. If, however, this entails that we could never have grounds for abandoning such propositions, it would appear to be false.

I want to claim that not only psychological theories, but also the metatheory of psychology may, in the relevant sense, be subject to empirical disconfirmation. To show that learning can occur without reward is to show both that some behaviour is *not* directed towards the reduction of drive and that an account of psychological explanation according to which explaining learned behaviour invariably consists in showing how it affects drive reduction is an inadequate account.

This view may seem simply paradoxical. 'If consonance with the proposition P is a necessary condition placed upon the acceptability of psychological theories by some metatheory, then surely no disconfirmation of P is possible since, *ex hypothesis*, no theory incompatible with P is acceptable.' What that argument overlooks is that major revolutions in scientific thought often affect not only our beliefs as to what explanations are true, but also our notions about what constitutes an explanation. Thus, it may be true both that our notion of a satisfactory explanation includes consonance with some very general empirical assumptions and that such assumptions could be abandoned in face of overwhelmingly persuasive counter-explanations of a previously unanticipated type.

I want to say what a psychological explanation of behaviour is, for I hold that behaviour is susceptible of psychological explanation. Some philosophers deny this. They maintain either that psychological explanation is concerned solely with *aberrant* behaviour or that psychological explanation is not concerned with behaviour at all, but only with motions. We shall have to examine these views. An account of psychological theories is required to say what psychological theories are about.

In the *Concept of Mind*, Gilbert Ryle writes:[1]

[1] Ryle, G., *The Concept of Mind*, New York 1949, p. 326.

> The classification and diagnosis of exhibitions of our mental impotences require specialized research methods. The explanation of the exhibitions of our mental competences often requires nothing but ordinary good sense, or it may require the specialized methods of economists, scholars, strategists and examiners. But their explanations are not cheques drawn on the accounts of some yet more fundamental diagnoses. So not all, or even most, causal explanations of human actions and reactions are to be ranked as psychological.

It is clear that Ryle has been careful not to burn his bridges. He says only that explaining mental competences *often* requires nothing but good sense. This might equally be said of 'impotences' and lapses, for 'his attention wandered', 'it slipped his mind', 'he was tired', 'he didn't think what he was doing', etc. may all be satisfactory explanations. If, however, Ryle holds that psychological explanations can be given only in cases of failure to perform, or in cases where the performance somehow runs contrary to expectations, then Ryle is simply wrong.

That normal functioning often needs to be accounted for is clear enough in cases other than behaviour. To explain how an internal combustion engine works is to account for its normal performance; the account will not include an explanation of backfires, misfires, and overheating. Backfires and misfires can be explained, but explaining them is not part of explaining how an internal combustion engine works. And backfires and misfires are certainly not *all* that can be explained. Engineering schools offer courses in the theory of the internal combustion engine, not in the theory of backfires and misfires.

If the situation is less obvious in the case of behaviour, that is because, of the variety of types of explanation we can give to account for what someone did, the one we want for practical purposes is rarely couched in terms of underlying psychological mechanisms. Analogously, if the insurance agent wants an explanation of the fire, we do not offer him physics. Yet presumably a physical explanation could be given and would be appropriate on certain occasions. Roughly: the appropriateness of an explanation is determined not by the phenomena it seeks to account for but by the question it seeks to answer.

It is clear from myriad examples that psychological explanations

of normal behaviour can be (and often are) given and accepted. Thus, consider:

1. Freud explained that the occurrence of dreams is a mechanism for dealing with stimuli which would otherwise interrupt sleep.[1]

2. An explanation of the perceptual constancies accounts for our ability to see true colour even under adverse lighting conditions.[2]

3. Broadbent explained our ability to follow two conversations at once by reference to a hypothetical system of filters and stores.[3]

4. Skinner explained learned perceptual distinctions by reference to histories of reinforcement.[4]

It is irrelevant whether the explanations instanced in 1–4 are in fact correct accounts of the phenomena with which they are concerned. I am interested only in the point that what each purports to explain is either a 'competence' or a bit of perfectly normal human behaviour. It is a sufficient argument against Ryle's account of psychological explanation that it renders such explanations as 1–4 logically inappropriate. If a certain view of explanation entails that most of psychology will have to be abandoned without hope of replacement, that shows something is wrong with the view, not that something is wrong with psychology.

It appears that neither an appeal to explanations of phenomena other than behaviour nor an appeal to the received practices of psychologists uncovers support for the claim that psychological explanations must be limited to accounting for aberrations. On the contrary, the psychologist's concern with aberrant phenomena is often motivated primarily by the belief that they represent the automatic consequence of the application to atypical situations of the principles governing normal behaviour. What Teuber has said about the motivation for studying illusions applies, *mutatis mutandis*, to areas of psychology other than perception: '. . . to speak of illusions as special cases—curiosa of perception, as it were—is tendentious . . . the explanation for perceptual illusions will

[1] Cf. e.g. Freud, S., *General Introduction to Psychoanalysis*, New York 1920.
[2] Cf. Teuber, H.-L., 'Perception,' *Handbook of Physiology*, Vol. 3, Washington 1960.
[3] Cf. Broadbent, D. E. *Perception and Communication*, New York 1958.
[4] Cf. Skinner, B. F. *The Behaviour of Organisms*, New York 1938.

be sought among the general laws of perception. Once these laws are known, the illusions themselves will be understood'.[1]

In so far as psychology is concerned to explain behaviour at all, it is concerned to explain normal behaviour *inter alia*. But philosophers have sometimes argued that psychological explanations are not (that is, cannot be) explanations of behaviour. In an article entitled 'Behaviour', Hamlyn writes:[2]

No mechanism of any sort can do more than account for movements, reactions, and the like. It may, of course, be the case that a particular movement or series of movements may exemplify a kind of behaviour; it may be classifiable as such, and capable of such an interpretation. It is this possibility which permits us on any particular occasion to describe both the movements and the behaviour, though to do these things will by no means be to do the same thing. Thus, no mechanism can be given which will account for behaviour *per se* however much we may feel that the behaviour will have been accounted for incidentally in providing a mechanism for the movements which constitute behaviour on a particular occasion. At other times, however, the movements involved may be different, though we may still describe the behaviour in the same way.

Unlike Ryle's, this view of psychological explanation is found among psychologists. Thus, to choose an example at random, Tinbergen[3] characterizes the domain of the behaviour sciences as '. . . the total *movements* made by the intact animal'.

There are two sorts of reasons for holding that psychology is concerned to explain movements in the sense in which movements are contrasted with behaviour. First, one may be impressed, as Hamlyn is, with the fact that why-questions about behaviour are appropriately answered by citing reasons rather than causes. Hence, it is argued, if psychology is a causal science, its explanations cannot be explanations of behaviour. Second, one may be impressed, as psychologists often are, by the need to eliminate from the 'observation base' of the science (i.e. from the vocabulary

[1] Teuber, *op. cit.*, p. 1601.
[2] Hamlyn, D. W. 'Behaviour,' reprinted in Chappell, V. C. (ed.) *The Philosophy of Mind*, Englewood Cliffs, N.J., 1962, p. 65.
[3] Tinbergen, N. *The Study of Instinct*, Oxford 1951, p. 2. Emphasis mine.

in which its predictions are couched) any term whose application requires interpretation of the phenomena. It must be possible to determine by purely observational procedures whether a prediction of the theory has been verified, since to use theoretical constructs in describing the phenomena upon which the confirmation of the theory depends is held to be circular. 'To describe behaviour requires interpretation of movements according to certain standards . . .'[1] Hence, it is only by limiting the theory to accounting for motions that we can assure ourselves that its explanations and predictions will be susceptible of purely objective verification.

It is notable that this position is open to a *reductio ad absurdum* argument similar to that to which Ryle's succumbed. That is, if we were literally to proscribe the psychological explanation of behaviour, it would turn out that not even learning theory is properly part of psychology, since *not even so basic a psychological notion as that of a response can be characterized in terms of movements alone.*[2] In laboratory situations, an organism is said to have mastered a response when it regularly produces any of an indefinite number of types of functionally equivalent motions under the appropriate stimulus conditions. That some reasonable notion of functional equivalence can be specified is essential, since we cannot in general require that two motions manifesting the same response be identical either in their observable properties or in their physiological basis. Thus, a rat has 'got' the bar pressing response if and only if it habitually presses the bar upon food deprivation. Whether it presses with its left or right front paw or with three or six grams of pressure is, or may be, irrelevant. Training is to some previously determined criterion of homogeneity of performance, which is to say that we permit variation among the *motions* belonging to a response so long as each of the variants is functionally equivalent to each of the others: *viz.* so long as each of the motions is correctly related to the bar, to the general stimulus situation, and to the history of the organism.

Not only does the requirement that psychology concern itself with motions alone prohibit the employment of such basic notions as 'response', it also prohibits the construction of a reasonable criterion of identity for motions themselves. An otherwise indis-

[1] Hamlyn, *op. cit.*, pp. 63–64.

[2] Cf. Chomsky, N. 'Review of Skinner's *Verbal Behaviour*', *Language*, Vol. 35 No. 1, 1959.

tinguishable pair of motions may be produced by quite different physiological mechanisms and hence be the outcome of quite different psychological processes. In order to take account of this fact, it may very often be necessary to determine identity and difference of motions by identity and difference of the muscular contractions that produce them[1] and, in case the same muscular contractions are sometimes under the control of different central processes, we may finally have to determine identity and difference of motions by identity and difference of hypothetical underlying causal mechanisms at the neural level.[2]

In short, the requirement that we characterize the events upon which the confirmation of a theory depends *only* in terms of their immediately observable properties may render the systematic explanation of those events impossible. Among the goals of theory construction is that of providing a conceptual framework for the coherent description of the phenomena with which the theory is concerned. That is, it is one of the achievements of a satisfactory theory that it provides a way of determining identity and difference of the confirming events such that, *on that determination*, the occurrence of those events is rendered susceptible of explanation. The view that such determinations can in principle be made on the basis of purely observable features of behaviour is so far from being obviously true as to make its adoption as a methodological rule extremely ill-advised. In the present case, it is by no means clear that a science of the motions of organisms is possible: that is, it is unclear that anything systematic could be said about the motions of an organism unless we permitted ourselves to identify motions not solely on the basis of their immediately observable properties, but also by their relation to such hypothetical states as drives, needs, goals, muscle contractions, neurological firings, and so on. To put it somewhat differently, among the facts which drive us to theory construction in psychology is the existence in non-verbal behaviour of the counterparts of ambiguity and synonymy. Just as, in linguistics, not every utterance of the phonemic sequence 'bank' is an utterance of the same word, so in psychology, not every occurrence of a given movement or muscular contraction is an instance of the same behaviour. Conversely, in linguistics two

[1] This is, in fact, what Tinbergen does in the volume cited above.

[2] For an interesting example, Cf. Luria, A. R., *Speech and the Development of Mental Processes in the Child*, London 1959.

phonemically distinct utterances ('bachelor', 'unmarried man') may be equivalent in significant respects. So, in psychology, two quite different patterns of motions (swimming to the right and running to the right in a T-maze) may be instances of the same behaviour: a fact we notice when we discover that an organism trained to produce one will, under appropriate circumstances, produce the other without further training. The consequences of such facts are identical in both sciences. If we are to capture the relevant generalizations, identity and difference of the events with which the science is concerned must often be determined by reference to properties other than those that are directly observable. In particular, in both sciences we attempt to construct theories containing levels sufficiently abstract to enable us to mark the respects in which events whose observable properties are identical may nevertheless be functionally distinct and the respects in which events whose observable properties are distinct may nevertheless be functionally identical.[1]

But it may still be said that the explanation of behaviour requires reasons while causal explanations provide not reasons but causes. There is, I think, something to this argument: explanations of behaviour are very often given by appealing to motives, utilities, strategies, goals, needs, desires, and so on.[2] It seems clear that such explanations will not be forthcoming from a causal science where this is understood to be a science which affords explanations *only* by appealing to causal chains and causal laws.[3] I shall argue that psychology is not a causal science in *that* sense. At any event, at the present stage there is no need to suppose that, because some explanations of behaviour are not causal, psychology must be limited to saying '. . . that in certain circumstances people behave

[1] Cf. Chomsky, N., *Syntactic Structures*, The Hague 1957; J. J. Katz and J. A. Fodor, 'The Structure of a Semantic Theory', *Language*, Vol. 39, No. 2, June 1963.

[2] Which need not blind us to the fact that causal explanations of behaviour are sometimes precisely what the situation requires. 'It was the liquor he drank that made him behave so badly.'

[3] The notion of a causal explanation is not itself so clear that it is evident precisely what is being asserted or denied when it is claimed that psychology is or is not a causal science. I shall follow Hamlyn in adopting the most restricted interpretation of this notion. In particular, I shall use 'causal explanation' and 'mechanistic explanation' as roughly interchangeable. To deny that psychological explanations are causal in this sense is not, of course, to deny that they may be causal in some broader sense.

in certain ways . . .'[1] or that we should ' . . . content ourselves with the programme of accounting for behaviour in terms of the capacities or dispositions from which it is derivable',[2] an undertaking which, as Hamlyn rightly remarks, 'is not a scientific programme, but one which may be carried out by anyone with sufficient experience of human affairs'.[3] Rather, the argument shows that we need to understand how a science can afford explanations and predictions of events in terms which do not refer solely to the causes of those events.

Psychology is the systematic attempt to explain and predict the behaviour of organisms. It is assumed that at any instant behaviour is the joint product of two sorts of factors:

1. Stimuli currently impinging upon the sensory receptors of the organism.
2. Internal states of the organism.

The relative contribution of each of these factors to the determination of behaviour probably varies greatly for behaviour of different kinds. While knowledge of local stimulus conditions contributes greatly to accurate prediction of certain kinds of instinctive behaviour and certain kinds of conditioned behaviour, in the case of verbal behaviour knowledge of the stimulus situation often affords very little grounds for predicting what the organism will do.

I shall argue that psychological explanation is essentially a two-phase process, the first phase of which is the development of a theory of the internal states of the organism such that (*a*) the terms of the theory which do not refer to behaviour are functionally characterized and (*b*) the theory is capable of adequately predicting the behaviour of the organism given knowledge of the current stimulus situation. Each of these conditions must be discussed at length.

Quite aside from any physiological considerations, it is possible to say a number of things about the kinds of internal states organisms must be supposed to have if characteristic features of their behaviour are to be accounted for. For example: the behaviour of an organism in a specified stimulus situation is very often partly determined by the previous stimulations it has encountered. Much of the most careful work in recent psychology

[1] Hamlyn, *op. cit.* p. 66. [2] *Ibid.* [3] *Ibid.*

has been devoted to exhibiting the differences between naive and sophisticated behaviour and to determining which patterns of stimulation are conducive to the development of sophistication. But though it is obvious that organisms of identical genetic endowment often differ profoundly in their response to novel stimulations depending on features of their individual life histories, it is not obvious how this fact should be accounted for. The problem becomes apparent when we notice that the degree to which, and the conditions under which, prior stimulations determine current behaviour differ markedly from species to species: discriminations difficult for the octopus to learn are easy for the rat, imprinting is known in birds but not in monkeys, operant conditioning is easier with fish than with planaria, verbal learning occurs only in man. The susceptibility of behaviour to alteration by experience would thus appear to vary from species to species.[1]

If we are to account for the alteration of behaviour as a result of prior stimulation, we must assume that some at least of the internal states that determine the way an organism responds to current stimulation are themselves the product of its previous experiences. Since the laws governing the formation of such states may be supposed to differ from species to species, it becomes understandable that the same history of stimulation produces very different behaviour in organisms of sufficiently different biological types. Conversely, if genetically identical organisms have such internal states in common only in case their life histories have been similar in relevant respects, then we expect relevantly dissimilar life histories to produce differences in behaviour. Finally, the assumption that some such experientially induced states are inherently unstable and tend to decay in a lawful fashion provides for the possibility of explaining such characteristic features of long term memory as stereotyping, elimination of detail, tendency towards 'good form', etc.[2]

It goes without saying that the laws which presumably determine the careers of such internal states (and, in particular, the laws which determine under what stimulus conditions they arise and how they contribute to the production of behaviour) are arrived at indirectly. The internal states of the organism are assumed to have

[1] Cf. Thorpe, W. H. *Learning and Instinct in Animals*, London 1956.

[2] Cf. e.g. Bartlett, F. C. *Remembering*, New York 1932, for a discussion of characteristic features of the decay of memories.

those properties required to account for the observed features of its behaviour. This is a sort of reasoning that is perfectly ordinary in sciences other than psychology. Radio telescopy shows the star to be very active, light telescopy shows it to be very dim. Perhaps we are dealing with a bright star very far away. The function of the theory is, *inter alia*, to save the appearances.

The sense in which terms referring to internal states are functionally characterized in theories developed in the first phase of psychological explanation may now be made clear. Phase one psychological theories characterize the internal states of organisms only in respect of the way they function in the production of behaviour. In effect, the organism is thought of as a device for producing certain behaviour given certain sensory stimulations. A phase one psychological explanation attempts to determine the internal states through which such a device must pass if it is to produce the behaviour the organism produces on the occasions when the organism produces it. Since, at this stage, 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, it follows that what a phase one theory tells us about such states is what role they play in the production of behaviour. It follows too that the evidence to be adduced in favour of the claim that such states exist is just that assuming they do is the simplest way of accounting for the behavioural capacities the organism is known to have.

It should be noticed that explanations afforded by phase one theories are *not* causal explanations, although a fully elaborated phase one theory claims to be able to predict behaviour given sufficient information about current sensory stimulations. Phase one explanations purport to account for behaviour in terms of internal states, but they give no information whatever about the mechanisms underlying these states. That is, theory construction proceeds in terms of such functionally characterized 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. Now, 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 (including, perhaps, behaviour which consists in his telling me why he left). Moreover, it is an explanation

which, *ceteris paribus*, might have been adequate for the prediction of behaviour since I might have known that *if* he had been reminded of his engagement he would certainly have left. Yet, it is not a causal explanation in the sense in which that term is usually used. That is, it is not at all like a reflex-arc explanation of a knee-jerk response or an explanation of the trajectory of a billiard ball; no causal laws are invoked, nor is any notion of a causal chain at issue.

We thus arrive at the following view of phase one psychological explanations. Organisms are observed to produce certain types of behaviour either spontaneously or as the consequence of certain types of stimulation. A phase one psychological theory attempts to account for these observations by reference to hypothetical internal states which, together with the relevant stimulation, are supposed to produce the observed behaviour. The regularity of the observed behaviour is thus explained and rules provided which enable us to predict what the organism will do in any of indefinitely many novel situations. Phase one explanations are arrived at indirectly in that we attribute to the organism whatever internal states are required to account for its behavioural repertoire. The characterization of these states is thus purely functional since we know about them only what role they play in the production of behaviour.

A characteristic feature of phase one explanations is that they are compatible with indefinitely many hypotheses about the physiology of the organism. We have seen that phase one explanations are *not* causal explanations precisely because they make no claims about the mechanisms underlying internal states. In a phase one explanation, we picture the organism as proceeding through a series of internal states that terminate in the production of observable behaviour. But we make no attempt to say what these states are states of: what internal mechanisms correspond to the functionally defined states we have invoked. Now, the set of mechanisms capable of realizing a series of such functionally defined states is indefinitely large. Only our ingenuity limits the number of mechanisms we could devise which, upon exposure to the relevant stimulations, would go through a sequence of internal states each functionally equivalent to a corresponding state of an organism and would then produce behaviour indistinguishable in relevant respects from the behaviour of the organism.

We may say that each mechanism capable of realizing the series

of states a phase one theory attributes to an organism is a *model* of the theory. And we may now see why phase one explanations are inadequate accounts of behaviour. For, in the first phase of psychological explanation, we say no more of an organism than that it is one of an indefinitely large number of possible models of a theory. Which such model the organism is is something a phase one explanation does not determine.

Many psychologists would claim that this last question is not properly within the domain of their science. J. A. Deutsch, for example, has argued persuasively that the production of adequate phase one theories exhausts the psychologist's professional responsibilities.

> For instance, to attempt to guess at the particular change which occurs in the central nervous system during learning in the framework of a theory purporting to explain behaviour is not only unnecessary but also purely speculative. That some type of change occurs may be inferred from the behaviour of an animal. What this type of change is cannot be arrived at, nor is it very important for the psychologist to know. This can be shown by taking the example of an insightful learning machine. . . . To be told that the semipermanent change in the machine which occurs when it learns is due to a uniselector arm coming to rest does not help us to understand the behavioural properties of the machine. Nor can it be checked by performing experiments on the behaviour of the machine. For the change could equally well be due to a self-holding relay, a dekatron selector, or any type of gadget known to technology capable of being turned from one steady state into another. In the same way, to speculate about terminal end boutons in the way that Hebb does or about changes of synaptic resistance seems to be trying to answer a question irrelevant, strictly speaking, to the psychological theorist. What behaviour would one of these assumptions explain which the others would not?[1]

Border disputes tend to be philosophical in the sense of that term in which it is synonymous with 'uninteresting'. But more is at issue here than whether the determination of the physical representation of a phase one theory in the nervous system of an organism is the duty of the psychologist or the neurologist or both.

It must be remembered that the talk of a first and second phase

[1] Deutsch, *op. cit.* p. 12.

of psychological explanation cannot be understood as expressing a chronological relation between types of psychological theories. It is offered as a reconstruction of psychological explanation, not as a history of the development of psychology. In historical fact, what happens is that research directed towards a functional account of behaviour is simultaneous with research directed towards determining the nature of the mechanisms whose functional characteristics phase one theories specify. This fact has two fairly important consequences. First, information about the mechanisms underlying behaviour may sometimes lead to hypotheses that are most naturally stated in functional terms and tested in terms of behaviour. The history of psychological research on memory is filled with experiments originally inspired by speculations about the neurology of the memory trace, just as the history of perception theory is filled with experiments inspired by speculations about the character of the neural events triggered by a stimulus array. Secondly, and more important, it seems reasonable to maintain that any phase one theory that is incompatible with known facts of neurology must be, *ipso facto*, unacceptable. To put it slightly differently, it is sufficient to disconfirm a functional account of the behaviour of an organism to show that its nervous system is incapable of assuming states manifesting the functional characteristics that account requires. To accept this principle is, of course, to build into our characterization of psychological explanation a blatantly empirical assumption about the causation of behaviour: namely that the nervous system does, in fact, constitute a model of some phase one theory. This may be an incorrect view of the relation between neural and molar events (we had anticipated the possibility that the metatheory of psychology might itself prove susceptible of empirical disconfirmation). But, if it is correct, it provides an extremely important constraint upon phase one theories. Moreover, it provides motive for precisely the sort of neurological speculations about which Deutsch professes suspicion. If consonance with neurological fact is a condition upon the adequacy of phase one theories, it is clearly good strategy for the psychologist to construct such theories in awareness of the best estimates of what the neurological facts are likely to be.

It should be noticed that the view of the relation between psychological and neurological theories espoused here is to be distinguished from all varieties of reductionism. On this view,

neurological structures are models of certain functionally characterized relations. A neurological theory thus provides an account of the mechanics of systems whose functional characteristics are given by phase one theories. But to attempt to reduce a functional account to a mechanistic account would be patently absurd; the relation between functional analysis and mechanistic analysis is not at all like the relation between macroanalysis and microanalysis, though the two have sometimes been confused.

In microanalysis one asks: 'What does X consist of?' and the answer has the form of a specification of the microstructure of Xs. Thus: 'What does water consist of?' 'Two molecules of hydrogen linked with one molecule of oxygen.' 'What does lightning consist of?' 'A stream of electrons.' And so on. In functional analysis, one asks about a part of a mechanism what role it plays in the activities characteristic of the mechanism as a whole: 'What does the camshaft do?' 'It opens the valves, permitting the entry into the cylinder of fuel which will be detonated to drive the piston.' Successful microanalysis is thus often contingent upon the development of more powerful instruments of observation or more precise methods of disection. Successful functional analysis, on the other hand, requires an appreciation of the sorts of activities characteristic of a mechanism and of the contribution of the functioning of each part of the economy of the whole.

Explanation in psychology consists of a functional analysis and a mechanistic analysis: a phase one theory and a determination of which model of the theory the nervous system of the organism represents. Neither aspect of the explanation is dispensible. In particular, a neurological account without the corresponding phase one account would amount to no more than a description of a series of biochemical and electrical interactions. It would fail to describe the role of these interactions in the production of behaviour.[1] To put it succinctly, a complete psychological explanation requires more than an account of what the neurological circuitry is; it requires also an account of what such circuitry does. This second sort of account is given in terms of the familiar constructs of psychology: drives, motives, strategies, and so forth.

[1] I want to make it clear that I do *not* deny that accounts of functional relations may play an important role within neurology. There is, of course, nothing wrong with saying that the firing of a certain neuron inhibits the firing of some other. My point is rather that, *vis-à-vis* explanations of behaviour, neurological theories specify mechanisms and psychological theories do not.

Notice that explanations outside psychology often have this same double aspect: functional analysis plus mechanistic analysis. We say 'The camshaft functions to lift the valves at the proper time by displacing the tappets.' That is, we say what the camshaft does and we say how it does it. Neither account is adequate without the other.

Psychologists and philosophers who have complained that it is possible to trace an input from afferent to central to efferent neurological systems without once encountering motives, strategies, drives, needs, hopes, and so forth have thus been right in one sense but wrong in another, just as one would be if one argued that a complete causal account of the operation of an internal combustion engine never encounters such a thing as a valve lifter. In each case, the confusion occurs when a term properly figuring in functional accounts of mechanisms is confounded with terms that properly appear in causal accounts. From a functional point of view, a camshaft is a valve lifter. But a mechanistic account of the operations of internal combustion engines does not seek to replace the concept of a valve lifter with the concept of a camshaft, nor does it seek to reduce the former to the latter. What it does do is explain how the valves get lifted, what mechanical transactions are involved when the camshaft lifts the valves.

There is no sense to the question 'What does a valve opener consist of?' where this is understood as a request for a micro-analysis. Functions do not have parts; valve openers are not made of rods, springs and atoms in the sense that camshafts are.[1] There is a sensible question: 'How are the valves opened in this (sort of) engine?' This question invites a mechanistic account, and in such an account the term 'camshaft' may appear. Analogously, there is a sensible question: 'What is the mechanism of drive reduction in this (sort of) organism?' This question invites a neurological account, and in such accounts the term 'circuit' may appear.

Drives, motives, strategies, etc. are internal states postulated in attempts to account for behaviour in phase one theories. In

[1] To add to the confusion, however, it may be observed that some *mechanisms* are designated by their function. This is why in one sense it does and in another sense it does not make sense to ask: 'What is a can opener made of?' Again, it is because 'mousetrap' is ambiguous between function and mechanism that it makes sense to talk of building a better one. Analogously, it is customary to designate *neurological* structures in terms of their supposed *psychological* functions: hence, the 'speech centre', the 'association cortex', etc.

completed psychological explanations they serve to characterize the functional aspects of neurological mechanisms. That is, they function in accounts of the relation between the operation of such mechanisms and the molar behaviour of organisms. But drives, motives and strategies are not themselves neurological mechanisms nor do they have a microanalysis in terms of neurological mechanisms. The remark 'A drive is not a neurological state' has the same logical status as the remark 'A valve lifter is not a camshaft.' That is, it expresses a necessary truth.

If the position just presented is correct, it would appear that much of the discussion of theoretical identification[1] that has arisen in attempting to determine the relation between neurological and psychological concepts must in fact be irrelevant to that problem. It need not be denied that, in general, no *a priori* determination can be made of the cases in which considerations of economy or elegance may require scientists to identify states or events previously held to be distinct. Nor need it be denied that, far from being arbitrary decisions, such identifications often have the status of major scientific discoveries. Above all, there is no reason to suppose an adequate view of language would require us to hold that such identifications invariably involve changes of meaning. But, important though these insights are for a proper understanding of scientific explanation, on the present view they do not apply to the relation between neurological and psychological theories; since psychological terms are understood to be names for functions, psychological states are not available for microanalysis and theoretical revision could identify them only with other functions, not with mechanisms.

[1] Cf. Place, U. T. 'Is Consciousness a Brain Process', reprinted in Chappell, *op. cit.*, Smart, J. J. C. 'Sensations and Brain Processes', reprinted in Chappell, *op. cit.*; and Putnam, *op. cit.*