PART I

Ontology: The Identity Theory and Functionalism

Introduction

Until nearly midway through the present century, the philosophy of mind was dominated by a "first-person" perspective. Throughout history (though with a few signal exceptions), most philosophers have accepted the idea, made fiercely explicit by Descartes, that the mind is both better known than the body and metaphysically in the body's driver's-seat. Some accepted Idealism, the view that only mind really exists and that matter is an illusion; some held that although matter does truly exist, it is somehow composed or constructed out of otherwise mental materials; some granted that matter exists even apart from mind but insisted that mind is wholly distinct from matter and partially in control of matter. Philosophers of this last sort we shall call "Cartesian Dualists."

Dualism and Behaviorism

All the aforementioned philosophers agreed that (a) mind is distinct from matter (if any), and that (b) there is at least a theoretical *problem* of how we human subjects can know that "external," everyday physical objects exist, even if there are tenable solutions to that problem. We subjects are immured within a movie theater of the mind, though we may have some defensible ways of inferring what goes on outside the theater.

All this changed very suddenly in the 1930s, with the accumulated impact of Logical Positivism and the verification theory of meaning. *Intersubjective verifiability* became the criterion both of scientific probity and of linguistic meaning itself. If the mind, in particular, was to be respected either scientifically or even as meaningfully describable in the first place, mental ascriptions would have to be pegged to publicly, physically testable verification-conditions. Science takes an intersubjective, "third-person" perspective on everything; the traditional first-person perspective had to be abandoned for scientific and serious metaphysical purposes.

The obvious verification-conditions for mental ascriptions are behavioral. How can the rest of us tell that you *are in pain* save by your wincing-and-groaning behavior in circumstances of presumable disorder, or that you *believe that broccoli will kill you* save by your verbal avowals and your nonverbal avoidance of broccoli? If the verification-conditions are behavioral, then the very meanings of the ascriptions, or at least the only facts genuinely described, are not inner and ineffable but behavioral. Thus Behaviorism as a theory of mind and a paradigm for psychology.

In psychology, Behaviorism took primarily a methodological form: Psychological Behaviorists claimed (i) that psychology itself is a science for the prediction and control of behavior, (ii) that the only proper data or observational input for psychology are behavioral, specifically patterns of physical responses to physical stimuli, and (iii) that *inner* states and events, neurophysiological or mental, are not proper objects of psychological investigation – neurophysiological states and events are the business of biologists, and mental

states and events, so far as they exist at all, are not to be mentioned unless operationalized nearly to death. Officially, the Psychological Behaviorists made no metaphysical claims; minds and mental entities might exist for all they knew, but this was not to be presumed in psychological experiment or theorizing. Psychological theorizing was to consist, *à la* Logical Positivism, of the subsuming of empirically established stimulus–response generalizations under broader stimulus–response generalizations.

In philosophy, Behaviorism did (naturally) take a metaphysical form: chiefly that of Analytical Behaviorism, the claim that mental ascriptions simply *mean* things about behavioral responses to environmental impingements. Thus, "Edmund is in pain" means, not anything about Edmund's putative inner life or any episode taking place within Edmund, but that Edmund either is actually behaving in a wincing-and-groaning way or is disposed so to behave (in that he would so behave were something not keeping him from so doing). "Edmund believes that broccoli will kill him" means just that if asked, Edmund will assent to that proposition, and if confronted by broccoli, Edmund will shun it, and so forth.

But it should be noted that a Behaviorist metaphysician need make no claim about the meanings of mental expressions. One might be a merely Reductive Behaviorist, and hold that although mental ascriptions do not *simply mean* things about behavioral responses to stimuli, they are ultimately (in reality) made true just by things about actual and counterfactual responses to stimuli. (On the difference between "analytic" reduction by linguistic meaning and "synthetic" reduction by *a posteriori* identification, see the next section of this introduction.) Or one might be an Eliminative Behaviorist, and hold that there are no mental states or events at all, but only behavioral responses to stimuli, mental ascriptions being uniformly false or meaningless.

Any Behaviorist will subscribe to what has come to be called the "Turing Test." In response to the perennially popular question "Can machines think?", Alan Turing (1964) replied that a better question is that of whether a sophisticated computer could ever pass a battery of (verbal) behavioral tests, to the extent of fooling a limited observer into thinking it is human and sentient; if a machine did pass such tests, then the putatively further question of whether the machine really *thought* would be idle at best, whatever metaphysical analysis one might attach to it. Barring Turing's tendentious limitation of the machine's behavior to verbal as opposed to nonverbal responses, any Behaviorist, psychological or philosophical, would agree that psychological differences cannot outrun behavioral test; organisms (including machines) whose actual and counterfactual behavior is just the same are psychologically just alike.

Philosophical Behaviorism adroitly avoided a number of nasty objections to Cartesian Dualism (see Carnap 1932/33; Ryle 1949; Place, this volume; Smart 1959; Armstrong 1968, ch. 5; Campbell 1984), even besides solving the methodological problem of intersubjective verification: it dispensed with immaterial Cartesian egos and ghostly nonphysical events, writing them off as ontological excrescences. It disposed of Descartes's admitted problem of mind–body interaction, since it posited no immaterial, nonspatial causes of behavior. It raised no scientific mysteries concerning the intervention of Cartesian substances in physics or biology, since it countenanced no such intervention.

Yet some theorists were uneasy; they felt that in its total repudiation of the inner, Behaviorism was leaving out something real and important. When they voiced this worry, the Behaviorists often replied with mockery, assimilating the doubters to old-fashioned Dualists who believed in ghosts, ectoplasm, and/or the Easter Bunny. Behaviorism was the only (even halfway sensible) game in town. Nonetheless, the doubters made several lasting points against it. First, anyone who is honest and not anaesthetized knows perfectly well that he/she experiences and can introspect actual inner mental episodes or occurrences, that are neither actually accompanied by characteristic behavior nor are merely static hypothetical facts of how he/she would behave if subjected to such-and-such a stimulation. Place (this volume) speaks of an "intractable residue" of conscious mental states that bear no clear relations to behavior of any particular sort; see also Armstrong (1968, ch. 5) and Campbell (1984). Second, contrary to the Turing Test, it seems perfectly possible for two people to differ psychologically despite total similarity of their actual and counterfactual behavior, as in a Lockean case of "inverted spectrum"; for that matter, a creature *might* exhibit all the appropriate stimulus–response relations and lack mentation entirely (Campbell 1984; Fodor and Block 1972; Block 1981; Kirk 1974). Third, the Analytical Behaviorist's behavioral analyses of mental

ascriptions seem adequate only so long as one makes substantive assumptions about the rest of the subject's *mentality* (Chisholm 1957, ch. 11; Geach 1957, p. 8; Block 1981), and so are either circular or radically incomplete as analyses of the mental generally.

So matters stood in stalemate between Dualists, Behaviorists and doubters, until the mid-1950s, when Place (this volume) and Smart (1959) proposed a middle way, an irenic solution.

The Identity Theory

According to Place and Smart, contrary to the Behaviorists, at least some mental states and events are genuinely inner and genuinely episodic after all. They are not to be identified with outward behavior or even with hypothetical dispositions to behave. But, contrary to the Dualists, the episodic mental items are not ghostly or nonphysical either. Rather, they are neurophysiological. They are identical with states and events occurring in their owners' central nervous systems; more precisely, every mental state or event is numerically identical with some such neurophysiological state or event. To be in pain is to have one's (for example) *c*-fibers, or possibly *a*-fibers, firing; to believe that broccoli will kill you is to have one's B_{bk}-fibers firing, and so on.

By making the mental entirely physical, this Identity Theory of the mind shared the Behaviorist advantage of avoiding the nasty objections to Dualism; but it also brilliantly accommodated the inner and the episodic as the Behaviorists did not. For according to the Identity Theory, mental states and events actually occur in their owners' central nervous systems; hence they are *inner* in an even more literal sense than could be granted by Descartes. The Identity Theory also thoroughly vindicated the idea that organisms could differ mentally despite total behavioral similarity, since clearly organisms can differ neurophysiologically in mediating their outward stimulus–response regularities. And of course the connection between a belief or a desire and the usually accompanying behavior is defeasible by other current mental states, since the connection between a *B*- or *D*-neural state and its normal behavioral effect is defeasible by other psychologically characterizable interacting neural states. The Identity Theory was the ideal resolution of the Dualist/Behaviorist impasse.

Moreover, there was a direct deductive argument for the Identity Theory, hit upon independently by David Lewis (1972) and D. M. Armstrong (1968, this volume). Lewis and Armstrong maintained that mental terms were *defined* causally, in terms of mental items' typical causes and effects. For example, "pain" *means* a state that is typically brought about by physical damage and that typically causes withdrawal, favoring, complaint, desire for cessation, and so on. (Armstrong claimed to establish this by straightforward "conceptual analysis"; Lewis held that mental terms are the theoretical terms of a commonsensical "folk theory" (see Part V below), and with the Positivists that all theoretical terms are implicitly defined by the theories in which they occur.) Now if by definition, pain is whatever state occupies a certain causal niche, and if, as is overwhelmingly likely, scientific research reveals that that particular niche is in fact occupied by such-and-such a neurophysiological state, it follows by the transitivity of identity that pain is that neurophysiological state; QED. Pain retains its conceptual connection to behavior, but also undergoes an empirical identification with an inner state of its owner. (An advanced if convolute elaboration of this already hybrid view is developed by Lewis (1980); for meticulous criticism, see Block (1978), Shoemaker (1981), and Tye (1983).)

Notice that although Armstrong and Lewis began their arguments with a claim about the meanings of mental terms, their Common-Sense Causal version of the Identity Theory itself was no such thing, any more than was the original Identity Theory of Place and Smart. Rather, all four philosophers relied on the idea that things or properties can sometimes be identified with "other" things or properties even when there is no synonymy of terms; there is such a thing as synthetic and *a posteriori* identity that is nonetheless genuine identity. While the identity of triangles with trilaterals holds simply in virtue of the meanings of the two terms and can be established by reason alone, without empirical investigation, the following identities are standard examples of the synthetic *a posteriori*, and were discovered empirically: clouds with masses of water droplets; water with H_2O; lightning with electrical discharge; the Morning Star with Venus; Mendelian genes with segments of DNA molecules; temperature (of a gas) with mean molecular kinetic energy. The Identity Theory was offered similarly, in a spirit

of scientific speculation; one could not properly object that mental expressions do not mean anything about brains or neural firings.

So the Dualists were wrong in thinking that mental items are nonphysical but right in thinking them inner and episodic; the Behaviorists were right in their physicalism but wrong to repudiate inner mental episodes. Alas, this happy synthesis was too good to be true.

Machine Functionalism

In the mid-1960s Putnam (1960, this volume) and Fodor (1968) pointed out a presumptuous implication of the Identity Theory understood as a theory of "types" or *kinds* of mental items: that a mental state such as pain has *always and everywhere* the neurophysiological characterization initially assigned to it. For example, if the Identity Theorist identified pain itself with the firings of *c*-fibers, it followed that a creature of any species (earthly or science-fiction) could be in pain only if that creature *had* *c*-fibers and they were firing. But such a constraint on the biology of any being capable of feeling pain is both gratuitous and indefensible; why should we suppose that any organism must be made of the same chemical materials as we in order to have what can be accurately recognized as pain? The Identity Theorist had overreacted to the Behaviorists' difficulties and focused too narrowly on the specifics of biological humans' actual inner states, and in so doing they had fallen into species chauvinism.

Fodor and Putnam advocated the obvious correction: What was important was not its being *c*-fibers (*per se*) that were firing, but what the *c*-fiber firings were doing, what their firing contributed to the operation of the organism as a whole. The *role* of the *c*-fibers could have been performed by any mechanically suitable component; so long as that role was performed, the psychology of the containing organism would have been unaffected. Thus, to be in pain is not *per se* to have *c*-fibers that are firing, but merely to be in some state or other, of whatever biochemical description, that plays the same causal role as did the firings of *c*-fibers in the human beings we have investigated. We may continue to maintain that pain "tokens," individual instances of pain occurring in particular subjects at particular times, are strictly identical with particular neurophysiological states of those subjects at those times, viz., with the states that happen to be playing the appropriate roles; this is the thesis of "token identity" or "token physicalism." But pain itself, the kind, universal, or "type," can be identified only with something more abstract: the causal or functional role that *c*-fiber firings share with their potential replacements or surrogates. Mental state-types are identified not with neurophysiological types but with more abstract functional roles, as specified by state-tokens' causal relations to the organism's sensory inputs, motor outputs, and other psychological states.

Putnam compared mental states to the functional or "logical" states of a computer: just as a computer program can be realized or instantiated by any of a number of physically different hardware configurations, so a psychological "program" can be realized by different organisms of various physiochemical composition, and that is why different physiological states of organisms of different species can realize one and the same mental state-type. Where an Identity Theorist's type-identification would take the form, "To be in mental state of type *M* is to be in the neurophysiological state of type *N*," Putnam's Machine Functionalism (as I shall call it) has it that to be in *M* is to be merely in some physiological state or other that plays role *R* in the relevant computer program (that is, the program that at a suitable level of abstraction mediates the creature's total outputs given total inputs and so serves as the creature's global psychology). The physiological state "plays role *R*" in that it stands in a set of relations to physical inputs, outputs, and other inner states that matches one-to-one the abstract input/output/logical-state relations codified in the computer program.

The Functionalist, then, mobilizes three distinct levels of description but applies them all to the same fundamental reality. A physical state-token in someone's brain at a particular time has a neurophysiological description, but may also have a functional description relative to a machine program that the brain happens to be realizing, and it may further have a mental description if some mental state is correctly type-identified with the functional category it exemplifies. And so there is after all a sense in which "the mental" is distinct from "the physical": though there are no nonphysical substances or stuffs, and every mental token is itself entirely physical, mental characterization is not physical characterization, and the property of being a pain is not simply the property of being such-and-such a neural firing.

Cognitive Psychology

In a not accidentally similar vein, Psychological Behaviorism has almost entirely given way to "Cognitivism" in psychology. Cognitivism is roughly the view that (i) psychologists may and must advert to inner states and episodes in explaining behavior, so long as the states and episodes are construed throughout as physical, and (ii) human beings and other psychological organisms are best viewed as in some sense *information-processing* systems. As cognitive psychology sets the agenda, its questions take the form, "How does this organism receive information through its sense-organs, process the information, store it, and then mobilize it in such a way as to result in intelligent behavior?" During the 1960s, the cognitive psychologists' initially vague notion of "information processing" (inspired in large part by the popularity of "Information Theory" in regard to physical systems of communication) became the idea that organisms employ internal representations and perform computational operations on those representations; *cognition* became a matter of the rule-governed manipulation of representations much as it occurs in actual digital computers.

The working language of cognitive psychology is of course highly congenial to the Functionalist, for Cognitivism thinks of human beings as systems of interconnected functional components, interacting with each other in an efficient and productive way.

Artificial Intelligence and the Computer Model of the Mind

Meanwhile, researchers in computer science have pursued fruitful research programs based on the idea of intelligent behavior as the output of skillful information-processing given input. Artifical Intelligence (AI) is, roughly, the project of getting computing machines to perform tasks that would usually be taken to demand human intelligence and judgment. Computers have achieved some modest success in proving theorems, guiding missiles, sorting mail, driving assembly-line robots, diagnosing illnesses, predicting weather and economic events, and the like. A computer *just is* a machine that receives, interprets, processes, stores, manipulates and uses information, and AI researchers think of it in just that way as they try to program intelligent behavior; an AI problem takes the form, "Given that the machine sees this as input, what must it already know and what must it accordingly do with that input in order to be able to . . . [recognize, identify, sort, put together, predict, tell us, etc.] . . . ? And how, then, can we start it off knowing that and get it to do those things?" So we may reasonably attribute such success as AI has had to self-conscious reliance on the information-processing paradigm.

This encourages the aforementioned idea that *human* intelligence and cognition generally are matters of computational information-processing. Indeed, that idea has already filtered well down into the everyday speech of ordinary people, among whom computer jargon is fairly common. This tentative and crude coalescing of the notions *cognition*, *computation*, *information*, and *intelligence* raises two general questions, one in each of two directions. First, to what extent might computers approximate minds? Second, to what extent do minds approximate computers?

The first question breaks down into three, which differ sharply and importantly from each other. (i) What intelligent tasks will any computer ever be able to perform? (ii) Given that a computer performs interesting tasks X, Y, and Z, does it do so *in the same way* that human beings do? (iii) Given that a computer performs X, Y, and Z and that it does so in the same way humans do, does that show that it has psychological and mental properties, such as (real) intelligence, thought, consciousness, feeling, sensation, emotion, and the like? Subquestion (i) is one of engineering, (ii) is one of cognitive psychology, and (iii) is philosophical; theorists' answers will depend accordingly on their commitments in these respective areas. But for the record let us distinguish three different senses or grades of "AI": AI in the weakest sense is cautiously optimistic as regards (i); it says these engineering efforts are promising and should be funded. AI in a stronger sense says that the engineering efforts can well serve as modelings of human cognition, and that their successes can be taken as pointers toward the truth about human functional organization. AI in the strongest sense favors an affirmative answer to (iii) and some qualified respect for the Turing Test: it says that if a machine performs intelligently *and* does so on the basis of a sufficiently human-like information-processing etiology, then

there is little reason to doubt that the machine has the relevant human qualities of mind and sensation. (AI in the strongest sense is fairly strong, but notice carefully that it does not presuppose affirmative answers to either (i) or (ii).)

The opposite issue, that of assimilating minds to computers, is very close to the philosophical matter of Functionalism. But here too there are importantly distinct subquestions, this time two: (i) Do human minds work in very like the way computers do as computers are currently designed and construed; for example, using flipflops grouped into banks and registers, with an assembly language collecting individual machine-code operations into subroutines and these subroutines being called by higher-level manipulations of real-world information according to programmed rules? (ii) Regardless of architecture, can human psychological capacities be entirely captured by a third-person, hardware-realizable design of *some* sort that could in principle be built in a laboratory? Subquestion (i) is of great interest (see Parts III and IV below), but is not particularly philosophical. Subquestion (ii) is tantamount to the fate of Functionalism.

Anomalous Monism

Donald Davidson (this volume, 1973, 1974) took a more radical view of the split between the token identity thesis (for mental and neurophysiological states or events) and the Identity Theorists' type thesis. He gave a novel and ingenious argument for token identity, based on his "Principle of the Anomalism of the Mental": "There are no strict and deterministic laws on the basis of which mental events can be predicted and explained" (this volume). The argument is roughly that since mental events interact causally with physical events and causality requires strict laws, the mental events must have physical descriptions under which they are related to other physical events by strict laws.

But then Davidson used the same principle to argue that in the matter of type identification, mental events are even worse off than the Machine Functionalist had suggested: Since in fact, mental types are individuated by considerations that are nonscientific, distinctively humanistic, and in part normative, they will not coincide with any types that are designated in scientific terms, let alone neurophysiological types. Thus, there will be no interesting type-identification of mental states or events with anything found in any science. The latter conclusion is not entirely explicit in Davidson, for he leaves some slack in the "strictness" a law must have in order to count as a scientific law. But he has refused to grant that the generalizations afforded by a Functionalist psychology, in particular, would count as sufficiently strict, for they are infested by *ceteris paribus* qualifications that can never be discharged.

Critics have replied that Davidson's case against the Functionalist type-identification is unproven, for that identification is entirely consistent with his premises (Van Gulick 1980; Lycan 1981; Antony 1989). Moreover, either the *ceteris paribus* qualification eventually could be discharged by a completed Functionalist psychology, or if not, then there is no reason to doubt that the same is true of other special sciences, such as biology.

Other commentators have criticized Davidson's notion of "supervenient" causation (Johnston 1985; Kim 1985; and see Part V below.)

Homuncular Functionalism and Teleology

Machine Functionalism supposed that human brains may be described at each of three levels, the first two scientific and the third familiar and commonsensical. (1) Biologists would map out human neuroanatomy and provide neurophysiological descriptions of brain states. (2) Psychologists would (eventually) work out the machine program that was being realized by the lower-level neuroanatomy and would describe the same brain states in more abstract, computational terms. (3) Psychologists would also explain behavior, characterized in everyday terms, by reference to stimuli and to intervening mental states such as beliefs and desires, type-identifying the mental states with functional or computational states as they went. Such explanations would themselves presuppose nothing about neuroanatomy, since the relevant psychological/computational generalizations would hold regardless of what particular biochemistry might happen to be realizing the abstract program in question.

Machine Functionalism as described has more recently been challenged on each of a number of points, that together motivate a specifically teleological notion of "function" (Sober (this volume)

speaks aptly of "putting the function back into Functionalism"):

(i) The Machine Functionalist still conceived psychological *explanation* in the Positivists' terms of subsumption of data under wider and wider universal generalizations. But Fodor (this volume), Cummins (1983), and Dennett (1978) have defended a competing picture of psychological explanation, according to which behavioral data are to be seen as manifestations of subjects' psychological capacities, and those capacities are to be explained by understanding the subjects as systems of interconnected components. Each component is a "homunculus," in that it is identified by reference to the function it performs, and the various homuncular components cooperate with each other in such a way as to produce overall behavioral responses to stimuli. The "homunculi" are themselves broken down into subcomponents whose functions and interactions are similarly used to explain the capacities of the subsystems they compose, and so again and again until the sub-sub-. . . components are seen to be neuroanatomical structures. (An automobile works – locomotes – by having a fuel reservoir, a fuel line, a carburetor, a combustion chamber, an ignition system, a transmission, and wheels that turn. If one wants to know how the carburetor works, one will be told what its parts are and how they work together to infuse oxygen into fuel; and so on.) Thus biologic and mechanical systems alike are hierarchically organized, on the principle of what computer scientists call "hierarchical control."

(ii) The Machine Functionalist treated functional "realization," the relation between an individual physical organism and the abstract program it was said to instantiate, as a simple matter of one-to-one correspondence between the organism's repertoire of physical stimuli, structural states, and behavior, on the one hand, and the program's defining input/state/output function on the other. But this criterion of realization was seen to be too liberal; since virtually anything bears a one–one correlation of some sort to virtually anything else, "realization" in the sense of mere one–one correspondence is far too easily come by (Block (1978), Lycan (1987, ch. 3)). Some theorists have proposed to remedy this defect by imposing a teleological requirement on realization: a physical state of an organism will count as realizing such-and-such a functional description only if the organism has genuine organic integrity and the state plays its functional role properly *for* the organism, in the teleological sense of "for" and in the teleological sense of "function." The state must do what it does as a matter of, so to speak, its biological purpose.

(iii) Machine Functionalism's two-leveled picture of human psychobiology is unbiological in the extreme. Neither living things nor even computers themselves are split into a purely "structural" level of biological/physiochemical description and any one "abstract" computational level of machine/psychological description. Rather, they are all hierarchically organized at many levels, each level "abstract" with respect to those beneath it but "structural" or concrete as it realizes those levels above it. The "functional"/"structural" or "software"/"hardware" distinction is entirely relative to one's chosen level of organization. This relativity has repercussions for Functionalist solutions to problems in the philosophy of mind (Lycan 1987, ch. 5), and for current controversies surrounding Connectionism and neural modeling (see Part III of this volume).

(iv) The teleologizing of functional realization has helped functionalists to rebut various objections based on the "qualia" or "feels" or experienced phenomenal characters of mental states (Lycan 1981; Sober, this volume).

(v) Millikan (1984, this volume), Van Gulick (1980), Fodor (1984, 1990), Dretske, (1988), and others have argued powerfully that teleology must enter into any adequate analysis of the intentionality or aboutness of mental states such as beliefs and desires. According to the teleological theorists, a neurophysiological state should count as *a belief that broccoli will kill you*, and in particular as *about broccoli*, only if that state has the representing of broccoli as in some sense one of its psychobiological functions.

All this talk of teleology and biological function seems to presuppose that biological and other

"structural" states of physical systems really have functions in the teleological sense. The latter claim is controversial to say the least. Some philosophers dismiss it as hilariously false, as a superstitious relic of primitive animism or Panglossian theism or at best the vitalism of the nineteenth century; others tolerate it but only as a useful metaphor; still others take teleological characterizations to be literally but only interest-relatively true, true *modulo* a convenient classificatory or interpretive scheme (Cummins 1975). Only a few fairly recent writers (Wimsatt 1972, Wright 1973, Millikan 1984, and a few others) have taken teleological characterizations to be literally and categorically true. This may seem to embarrass teleologized Functionalist theories of mind.

Yes and no. Yes, because if a Homuncular and/or Teleological Functionalist type-identifies mental items with teleologically characterized items, and teleological characterizations are not literally true, then mental ascriptions cannot be literally true either. Equivalently, if people really do have mental states and events, on their own and not merely in virtue of anyone's superstitious or subjective interpretation of them, but their physical states do not have objectively teleological functions, then mental states cannot be type-identified with teleological states.

Fortunately for the Teleological Functionalist there is now a small but vigorous industry whose purpose is to explicate biological teleology in naturalistic terms, typically in terms of etiology. For example, a trait may be said to have the function of doing *F* in virtue of its having been selected for because it did *F*; a heart's function is to pump blood because hearts' pumping blood in the past has given them a selection advantage and so led to the survival of more animals with hearts. Actually, no simple etiological explication will do (Cummins 1975, Boorse 1976, Bigelow and Pargetter 1987, Davies 1994), but philosophers of biology have continued to refine the earlier accounts and to make them into adequate naturalistic analyses of genuine function (Neander 1991, Godfrey-Smith 1994).

It should be noted that the correctness of type-identifying mental items with teleological items does not strictly depend on the objectivity or even the truth of teleological descriptions. For corresponding to each metaphysical view of teleology, including deflationary and flatly derisive ones, there is a tenable view of mind. Just as teleology may be a matter of interest-relative interpretation, so, after all, may mental ascriptions be (see Part II of this volume). For that matter, just as teleology may be only metaphorical, fictional, or illusory, so may mental ascriptions be; some philosophers now hold that mental ascriptions are in the end false (see Part III). But we shall consider those possibilities in due course.

Chronic Problems

Functionalism, cognitive psychology considered as a complete theory of human thought, and AI in the strongest sense all inherit some of the same problems that earlier beset Behaviorism and the Identity Theory. These remaining problems fall into two main categories, respectively headed, by philosophers, "qualia" and "intentionality," both mentioned in the previous section.

The "quale" of a mental state or event is that state or event's *feel*, its introspectible "phenomenal character." Many philosophers have objected that neither Functionalist metaphysics nor cognitive psychology nor AI nor the computer model of the mind can explain, illuminate, acknowledge, or even tolerate the notion of *what it feels like* to be in a mental state of such-and-such a sort. Yet, say these philosophers, the feels are quintessentially mental – it is the feels that make the mental states the mental states they are. Something, therefore, must be drastically wrong with Functionalism, cognitive psychology, AI in the strongest sense, and the computer model of the mind. Such "qualia"-based objections and responses to them will be the topic of Part VI below.

"Intentionality" is a feature common to most mental states and events, particularly the "propositional attitudes," those cognitive and conative states that are described in everyday language with the use of "that"-clauses. One believes *that broccoli is lethal*, desires *that visitors should wipe their feet*, hopes *that the Republican candidate will win*, etc. Other propositional attitudes include thoughts, intentions, rememberings, doubts, wishes, and wonderings.

A "that"-clause contains what is itself grammatically a sentence; intuitively that internal sentence expresses the "content" of the belief, desire, or other attitude in question. This is because propositional attitudes *represent* actual or possible states of affairs. That indeed is what makes them propositional attitudes, and accordingly they are described in terms of their respective representational contents.

The objects and states of affairs upon which our propositional attitudes are directed may actually obtain, in the real world. But equally they may not: beliefs are often false, desires can be frustrated, hopes may be dashed. The attitudes may also be about "things" that do not exist: Sherlock Holmes, the Easter Bunny, the free lunch. Franz Brentano raised the question of how any purely physical entity or state could have the property of being about or "directed upon" a nonexistent state of affairs or object; that is not the sort of feature that ordinary, purely physical objects can have. Many philosophers, including Chisholm (1957), have argued that no purely physical account of a system or organism, human or computer, could explain Brentano's property. That difficulty for Functionalism et al. will be addressed in Parts IV and V.

In alluding to sensory states and to mental states with intentional content, we have said nothing specifically about the emotions. Since the rejection of Behaviorism, theories of mind have tended not to be applied directly to the emotions; rather, the emotions have been generally thought to be conceptually analyzable as complexes of more central or "core" mental states, typically propositional attitudes such as belief and desire (and the intentionality of emotions has accordingly been traced back to that of attitudes). Kenny (1963) took this line, as do Armstrong (1968, ch. 8, sec. III), Solomon (1977), and Gordon (1987). However, there is a nascent literature on Functionalism and the emotions; see Part VII below.

It may be wondered whether materialist theories of the mind and/or functionalist theories in particular have any interesting implications for morality and ethics. Three materialists take this up explicitly: Smart (1963, ch. VIII), tries to exhibit a materialist basis for morals; Michael Levin (1979, ch. VII) addresses the specific charge that materialists cannot allow freedom of the will or whatever else may be necessary to make room for moral responsibility; Lycan (1985) explores some moral consequences of the computational view of the mind. A main purpose of Dennett (1978) is also to show why moral responsibility and the mental vernacular that supports it are possible despite Dennett's instrumentalist – sometimes fictionalist – treatment of the mental (see Part III of this volume).

Further Reading

Useful general works on theories of mind

Campbell, K. K. (1984) *Body and Mind* (2nd edn), University of Notre Dame Press.

Churchland, P. M. (1988) *Matter and Consciousness* (2nd edn), Bradford Books/MIT Press.

Braddon-Mitchell, D. and Jackson, F. (1996) *Philosophy of Mind and Cognition*, Blackwell Publishers.

Kim, J. (1996) *Philosophy of Mind*, Westview Press.

Rey, G. (1997) *Contemporary Philosophy of Mind*, Blackwell Publishers.

All five books contain very clear discussions of Dualism, Behaviorism, the Identity Theory, and Functionalism, as well as bibliographies. Churchland's takes up some of the newer developments in "neurophilosophy" (see Part III). See also McGinn, C. (1982) *The Character of Mind*, Oxford University Press; Smith, P. and Jones, O. R. (1986) *The Philosophy of Mind*, Cambridge University Press; Bechtel, W. (1988) *Philosophy of Mind: An Overview for Cognitive Science*, Laurence Erlbaum; Flanagan, O. (1991) *The Science of Mind*, Bradford Books/MIT Press; Graham, G. (1993) *Philosophy of Mind: An Introduction*, Blackwell Publishers. An excellent general reference work in the field is Guttenplan, S. (ed.) (1994) *A Companion to the Philosophy of Mind*, Blackwell Publishers.

Psychological Behaviorism

Skinner, B. F. (1933) *Science and Human Behavior*, Macmillan.

Chomsky, N. (1959) Review of B. F. Skinner's *Verbal Behavior*, *Language* 35, 26–58.

Behavioral and Brain Sciences 7, 4 (1984), a special issue on the "Canonical papers of B. F. Skinner." [See particularly Skinner's "Representations and misrepresentations," 655–65, his response to "Open peer commentary" on his article "Behaviorism at fifty."]

Analytical Behaviorism

Carnap, R. (1932/33) "Psychology in physical language," *Erkenntnis* 3, 107–42. [In the full text, Carnap considers possible objections at length.]

Hempel, C. G. (1949) "The logical analysis of psychology," in H. Feigl and W. Sellars (eds), *Readings in Philosophical Analysis*, Appleton Century Crofts.

Ryle, G. (1949) *The Concept of Mind*, Barnes and Noble.

Chisholm, R. M. (1957) *Perceiving*, Cornell University Press.

Geach, P. (1957) *Mental Acts*, Routledge & Kegan Paul.

Putnam, H. (1965) "Brains and behaviour," in R. J. Butler (ed.), *Analytical Philosophy, Part II*, Basil Blackwell.

The Turing Test

Turing, A. M. (1964) "Computing machinery and intelligence," reprinted in A. R. Anderson (ed.), *Minds and Machines*, Prentice-Hall.

Gunderson, K. (1985) *Mentality and Machines* (2nd edn), University of Minnesota Press.

Block, N. J. (1981) "Psychologism and Behaviorism," *Philosophical Review* 90, 5–43.

Rosenberg, J. F. (1982) "Conversation and intelligence," in B. de Gelder (ed.), *Knowledge and Representation*, Routledge & Kegan Paul.

Dennett, D. C. (1985) "Can machines think?", in M. Shafto (ed.), *How We Know*, Harper & Row.

The Identity Theory

Smart, J. J. C. (1959) "Sensations and brain processes," *Philosophical Review* 68, 141–56.

Feigl, H. (1967) *The "Mental" and the "Physical": The Essay and a Postscript*, University of Minnesota Press.

Presley, C. F. (ed.) (1967) *The Identity Theory of Mind*, University of Queensland Press.

Borst, C. V. (ed.) (1970) *The Mind/Brain Identity Theory*, Macmillan.

The Common-Sense Causal Theory

Armstrong, D. M. (1968) *A Materialist Theory of the Mind*, Routledge & Kegan Paul.

Armstrong, D. M. (1981) "Epistemological foundations for a materialist theory of the mind," reprinted in *The Nature of Mind and Other Essays*. Cornell University Press.

Lewis, D. (1972) "Psychophysical and theoretical identifications," *Australasian Journal of Philosophy* 50, 249–58.

Nagel, T. (1970) "Armstrong on the mind," *Philosophical Review* 79, 394–403.

Pappas, G. (1977) "Armstrong's Materialism," *Canadian Journal of Philosophy* 7, 569–92.

Block, N. J. (1978) "Troubles with Functionalism," in W. Savage (ed.), *Perception and Cognition: Minnesota Studies in the Philosophy of Science*, vol. IX, University of Minnesota Press. [N.b., Block's section on Lewis is included in neither the short version of his paper reprinted in this volume nor that reprinted in his own anthology, *Readings in Philosophy of Psychology*, vol. 1 (Harvard University Press, 1980).]

Lewis, D. (1980) "Mad pain and Martian pain," in N. Block (ed.), *Readings in Philosophy of Psychology*, vol. 1, ibid.

Shoemaker, S. (1981) "Some varieties of Functionalism," *Philosophical Topics* 12, 93–119.

Tye, M. (1983) "Functionalism and Type Physicalism," *Philosophical Studies* 44, 161–74.

Hill, C. S. (1991) *Sensations: A Defense of Type Materialism*, Cambridge University Press.

Machine Functionalism

Putnam, H. (1960) "Minds and machines," in S. Hook (ed.), *Dimensions of Mind*, Collier Books.

Putnam, H. (1967) "The mental life of some machines," in H.-N. Castañeda (ed.), *Intentionality, Minds, and Perception*, Wayne State University Press.

Fodor, J. A. (1968) *Psychological Explanation*, Random House.

Kalke, W. (1969) "What is wrong with Fodor and Putnam's Functionalism?" *Noûs* 3, 83–94.

Rorty, R. (1972) "Functionalism, machines, and incorrigibility," *Journal of Philosophy* 69, 203–20.

Fodor, J. A. and Block, N. J. (1972) "What psychological states are not," *Philosophical Review* 81, 159–81.

Lycan, W. (1974) "Mental states and Putnam's Functionalist hypothesis," *Australasian Journal of Philosophy* 52, 48–62.

Cognitive Psychology

Johnson-Laird, P. N. and Wason, P. C. (1977) *Thinking: Readings in Cognitive Science*, Cambridge University Press.

Anderson, J. R. (1985) *Cognitive Psychology and its Implications* (2nd edn), W. H. Freeman.

Glass. A. L. and Holyoak, K. J. (1986) *Cognition* (2nd edn), Random House.

N. A. Stillings et al. (1987) *Cognitive Science: An Introduction*, Bradford Books/MIT Press.

D. N. Osherson et al. (eds) (1995) *An Invitation to Cognitive Science*, 2nd edn, especially vol. 3, *Thinking* (ed. by E. E. Smith and D. N. Osherson), Bradford Books/MIT Press.

Artificial Intelligence and the computer model of the mind

A. R. Anderson (ed.) (1964) *Minds and Machines*, Prentice Hall.

Hofstadter, D. (1979) *Gödel, Escher, Bach: An Eternal Golden Braid*, Basic Books. [See also Hofstadter, D. and Dennett, D. C. (eds) (1981) *The Mind's I: Fantasies and Reflections on Self and Soul*, Basic Books.]

Winston, P. H. (1984) *Artificial Intelligence* (2nd edn), Addison-Wesley.

Pylyshyn, Z. W. (1984) *Computation and Cognition: Toward a Foundation for Cognitive Science*, MIT Press.

Charniak, E. and McDermott, D. (1985) *Introduction to Artificial Intelligence*, Addison-Wesley.

Haugland, J. (1985) *Artificial Intelligence: The Very Idea*, Bradford Books/MIT Press.

Dennett, D. C. (1986) "The logical geography of computational approaches: a view from the East Pole," in M. Brand and R. M. Harnish (eds), *The Representation of Knowledge and Belief*, University of Arizona Press.

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

Haugland, J. (ed.) (1997) *Mind Design II: Philosophy, Psychology, Artificial Intelligence*, Bradford Books/MIT Press.

Anomalous Monism

Davidson, D. (1973) "The material mind," in P. Suppes, L. Henkin, and A. Joja (eds), *Logic, Methodology and Philosophy of Science*, vol. 4, North-Holland.

Davidson, D. (1974) "Psychology as philosophy," in S. C. Brown (ed.), *Philosophy of Psychology*, Macmillan.

Van Gulick, R. (1980) "Rationality and the anomalous nature of the mental," *Philosophical Research Archives* 7, 1404.

Lycan, W. (1981) "Psychological laws," *Philosophical Topics* 12, 9–38.

Johnston, M. (1985) "Why having a mind matters," in E. Lepore and B. McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell.

Kim, J. (1985) "Psychophysical laws," in E. Lepore and B. McLaughlin (eds), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell.

Antony, L. (1989) "Anomalous Monism and the problem of explanatory force," in *Philosophical Review* 98, 153–87.

Homuncular Functionalism

Attneave, F. (1960) "In defense of homunculi," in W. Rosenblith (ed.), *Sensory Communication*, MIT Press.

Simon, H. (1969) "The architecture of complexity," in *The Sciences of the Artificial*, MIT Press.

Wimsatt, W. C. (1976) "Reductionism, levels of organization, and the mind–body problem," in G. Globus, G. Maxwell, and I. Savodnik (eds), *Consciousness and the Brain*, Plenum.

Dennett, D. C. (1978) *Brainstorms*, Bradford Books.

Haugeland, J. (1978) "The nature and plausibility of Cognitivism," *Behavioral and Brain Sciences* 1, 215–26.

Lycan, W. (1981) "Form, function, and feel," *Journal of Philosophy* 78, 24–50.

Cummins, R. (1983) *The Nature of Psychological Explanation*, Bradford Books/MIT Press.

Lycan, W. G. (1987) *Consciousness*, Bradford Books/MIT Press.

Teleological Functionalism

Dennett, D. C. (1969) *Content and Consciousness*, Routledge & Kegan Paul, chs III and IV.

Van Gulick, R. (1980) "Functionalism, Information and Content," *Nature and System*, 139–62; reprinted in the 1990 first edition of this anthology.

Fodor, J. A. (1984) "Semantics, Wisconsin Style," *Synthese* 59, 231–50.

Millikan, R. G. (1984) *Language, Thought, and Other Biological Categories*, Bradford Books/MIT Press.

Papineau, D. (1987) *Reality and Representation*, Basil Blackwell, ch. 4.

Fodor, J. A. (1980) "Psychosemantics," in the 1990 first edition of this anthology.

Dretske, F. (1988) *Explaining Behavior*, Bradford Books/MIT Press.

Naturalistic theories of teleology

Wimsatt, W. C. (1972) "Teleology and the logical structure of function statements," *Studies in History and Philosophy of Science* 3, 1–80.

Wright, L. (1973) "Functions," *Philosophical Review* 82, 139–68.

Cummins, R. (1975) "Functional analysis," *Journal of Philosophy* 72, 741–64.

Bennett, J. (1976) *Linguistic Behaviour*, Cambridge University Press, ch. 1.

Boorse, C. (1976) "Wright on functions," *Philosophical Review* 85, 70–86.

Bigelow, J. and Pargetter, R. (1987) "Functions," *Journal of Philosophy* 84, 181–96.

Neander, K. (1991) "Functions as Selected Effects: The Conceptual Analyst's Defense," *Philosophy of Science* 58, 168–84.

Davies, P. S. (1994) "Troubles for direct proper functions," *Noûs* 28, 363.

Godfrey-Smith, P. (1994) "A modern history theory of functions," *Noûs* 28, 344–62.

Emotions

Kenny, A. (1963) *Action, Emotion, and Will*, Routledge & Kegan Paul.

de Sousa, R. (1987) *The Rationality of Emotion*, Bradford Books/MIT Press.

Gordon, R. M. (1987) *The Structure of Emotions*, Cambridge University Press.

Solomon, R. (1977) *The Passions, Doubleday*.

Morality

Levin, M. (1979) *Metaphysics and the Mind–Body Problem*, Oxford University Press.

Lycan, W. (1985) "Abortion and the civil rights of machines," in N. T. Potter and M. Timmons (eds), *Morality and Universality*, D. Reidel.

Smart, J. J. C. (1963) *Philosophy and Scientific Realism*, Routledge & Kegan Paul.

1

The Identity Theory

Is Consciousness a Brain Process?

U. T. Place

The thesis that consciousness is a process in the brain is put forward as a reasonable scientific hypothesis, not to be dismissed on logical grounds alone. The conditions under which two sets of observations are treated as observations of the same process, rather than as observations of two independent correlated processes, are discussed. It is suggested that we can identify consciousness with a given pattern of brain activity, if we can explain the subject's introspective observations by reference to the brain processes with which they are correlated. It is argued that the problem of providing a physiological explanation of introspective observations is made to seem more difficult than it really is by the "phenomenological fallacy," the mistaken idea that descriptions of the appearances of things are descriptions of the actual state of affairs in a mysterious internal environment.

I Introduction

The view that there exists a separate class of events, mental events, which cannot be described in terms of the concepts employed by the physical sciences no longer commands the universal and unquestioning acceptance among philosophers and psychologists which it once did. Modern physicalism, however, unlike the materialism of the seventeenth and eighteenth centuries, is behavioristic. Consciousness on this view is either a special type of behavior, "sampling" or "running-back-and-forth" behavior as Tolman has it,[1] or a disposition to behave in a certain way, an itch, for example, being a temporary propensity to scratch. In the case of cognitive concepts like "knowing," "believing," "understanding," "remembering," and volitional concepts like "wanting" and "intending," there can be little doubt, I think, that an analysis in terms of dispositions to behave is fundamentally sound.[2] On the other hand, there would seem to be an intractable residue of concepts clustering around the notions of consciousness, experience, sensation, and mental imagery, where some sort of inner process story is unavoidable.[3] It is possible, of course, that a satisfactory behavioristic account of this conceptual residuum will ultimately be found. For our present purposes, however, I shall assume that this cannot be done and that statements about pains and twinges, about how things look, sound, and feel, about things dreamed of or pictured in the mind's eye, are statements referring to events and processes which are in some sense private or internal to the individual of whom they are predicated. The question I wish to raise is whether in making this assumption we are inevitably committed to a dualist position in which sensations and mental images form a separate category of processes over and above the physical and physiological processes with which they are known to be correlated. I shall argue that an acceptance of inner processes does not entail dualism and that the thesis that consciousness is a process in the brain cannot be dismissed on logical grounds.

II The "Is" of Definition and the "Is" of Composition

I want to stress from the outset that in defending the thesis that consciousness is a process in the brain, I am not trying to argue that when we describe our dreams, fantasies, and sensations we are talking about processes in our brains. That is, I am not claiming that statements about sensations and mental images are reducible to or analyzable into statements about brain processes, in the way in which "cognition statements" are analyzable into statements about behavior. To say that statements about consciousness are statements about brain processes is manifestly false. This is shown (a) by the fact that you can describe your sensations and mental imagery without knowing anything about your brain processes or even that such things exist, (b) by the fact that statements about one's consciousness and statements about one's brain processes are verified in entirely different ways, and (c) by the fact that there is nothing self-contradictory about the statement "X has a pain but there is nothing going on in his brain." What I do want to assert, however, is that the statement "Consciousness is a process in the brain," although not necessarily true, is not necessarily false. "Consciousness is a process in the brain" in my view is neither self-contradictory nor self-evident; it is a reasonable scientific hypothesis, in the way that the statement "Lightning is a motion of electric charges" is a reasonable scientific hypothesis.

The all but universally accepted view that an assertion of identity between consciousness and brain processes can be ruled out on logical grounds alone derives, I suspect, from a failure to distinguish between what we may call the "is" of definition and the "is" of composition. The distinction I have in mind here is the difference between the function of the word "is" in statements like "A square is an equilateral rectangle," "Red is a color," "To understand an instruction is to be able to act appropriately under the appropriate circumstances," and its function in statements like "His table is an old packing case," "Her hat is a bundle of straw tied together with string," "A cloud is a mass of water droplets or other particles in suspension." These two types of "is" statements have one thing in common. In both cases it makes sense to add the qualification "and nothing else." In this they differ from those statements in which the "is" is an "is" of predication; the statements "Toby is eighty years old and nothing else," "Her hat is red and nothing else," or "Giraffes are tall and nothing else," for example, are nonsense. This logical feature may be described by saying that in both cases both the grammatical subject and the grammatical predicate are expressions which provide an adequate characterization of the state of affairs to which they both refer.

In another respect, however, the two groups of statements are strikingly different. Statements like "A square is an equilateral rectangle" are necessary statements which are true by definition. Statements like "His table is an old packing-case," on the other hand, are contingent statements which have to be verified by observation. In the case of statements like "A square is an equilateral rectangle" or "Red is a color," there is a relationship between the meaning of the expression forming the grammatical predicate and the meaning of the expression forming the grammatical subject, such that whenever the subject expression is applicable the predicate must also be applicable. If you can describe something as red then you must also be able to describe it as colored. In the case of statements like "His table is an old packing-case," on the other hand, there is no such relationship between the meanings of the expressions "his table" and "old packing-case"; it merely so happens that in this case both expressions are applicable to and at the same time provide an adequate characterization of the same object. Those who contend that the statement "Consciousness is a brain process" is logically untenable, base their claim, I suspect, on the mistaken assumption that if the meanings of two statements or expressions are quite unconnected, they cannot both provide an adequate characterization of the same object or state of affairs: if something is a state of consciousness, it cannot be a brain process, since there is nothing self-contradictory in supposing that someone feels a pain when there is nothing happening inside his skull. By the same token we might be led to conclude that a table cannot be an old packing-case, since there is nothing self-contradictory in supposing that someone has a table, but is not in possession of an old packing-case.

III The Logical Independence of Expressions and the Ontological Independence of Entities

There is, of course, an important difference between the table/packing-case and the consciousness/brain

process case in that the statement "His table is an old packing-case" is a particular proposition which refers only to one particular case, whereas the statement "Consciousness is a process in the brain" is a general or universal proposition applying to all states of consciousness whatever. It is fairly clear, I think, that if we lived in a world in which all tables without exception were packing-cases, the concepts of "table" and "packing-case" in our language would not have their present logically independent status. In such a world a table would be a species of packing-case in much the same way that red is a species of color. It seems to be a rule of language that whenever a given variety of object or state of affairs has two characteristics or sets of characteristics, one of which is unique to the variety of object or state of affairs in question, the expression used to refer to the characteristics or set of characteristics which defines the variety of object or state of affairs in question will always entail the expression used to refer to the other characteristic or set of characteristics. If this rule admitted of no exception it would follow that any expression which is logically independent of another expression which uniquely characterizes a given variety of object or state of affairs must refer to a characteristic or set of characteristics which is not normally or necessarily associated with the object or state of affairs in question. It is because this rule applies almost universally, I suggest, that we are normally justified in arguing from the logical independence of two expressions to the ontological independence of the states of affairs to which they refer. This would explain both the undoubted force of the argument that consciousness and brain processes must be independent entities because the expressions used to refer to them are logically independent and, in general, the curious phenomenon whereby questions about the furniture of the universe are often fought and not infrequently decided merely on a point of logic.

The argument from the logical independence of two expressions to the ontological independence of the entities to which they refer breaks down in the case of brain processes and consciousness, I believe, because this is one of a relatively small number of cases where the rule stated above does not apply. These exceptions are to be found, I suggest, in those cases where the operations which have to be performed in order to verify the presence of the two sets of characteristics inhering in the object or state of affairs in question can seldom if ever be performed simultaneously. A good example here is the case of the cloud and the mass of droplets or other particles in suspension. A cloud is a large semi-transparent mass with a fleecy texture suspended in the atmosphere whose shape is subject to continual and kaleidoscopic change. When observed at close quarters, however, it is found to consist of a mass of tiny particles, usually water droplets, in continuous motion. On the basis of this second observation we conclude that a cloud is a mass of tiny particles and nothing else. But there is no logical connection in our language between a cloud and a mass of tiny particles; there is nothing self-contradictory in talking about a cloud which is not composed of tiny particles in suspension. There is no contradiction involved in supposing that clouds consist of a dense mass of fibrous tissue; indeed, such a consistency seems to be implied by many of the functions performed by clouds in fairy stories and mythology. It is clear from this that the terms "cloud" and "mass of tiny particles in suspension" mean quite different things. Yet we do not conclude from this that there must be two things, the mass of particles in suspension and the cloud. The reason for this, I suggest, is that although the characteristics of being a cloud and being a mass of tiny particles in suspension are invariably associated, we never make the observations necessary to verify the statement "That is a cloud" and those necessary to verify the statement "This is a mass of tiny particles in suspension" at one and the same time. We can observe the micro-structure of a cloud only when we are enveloped by it, a condition which effectively prevents us from observing those characteristics which from a distance lead us to describe it as a cloud. Indeed, so disparate are these two experiences that we use different words to describe them. That which is a cloud when we observe it from a distance becomes a fog or mist when we are enveloped by it.

IV When Are Two Sets of Observations Observations of the Same Event?

The example of the cloud and the mass of tiny particles in suspension was chosen because it is one of the few cases of a general proposition involving what I have called the "is" of composition which does not involve us in scientific technicalities. It is useful because it brings out the connection between the ordinary everyday cases of the "is" of composition like the table/packing-case exam-

ple and the more technical cases like "Lightning is a motion of electric charges" where the analogy with the consciousness/brain process case is most marked. The limitation of the cloud/tiny particles in suspension case is that it does not bring out sufficiently clearly the crucial problems of how the identity of the states of affairs referred to by the two expressions is established. In the cloud case the fact that something is a cloud and the fact that something is a mass of tiny particles in suspension are both verified by the normal processes of visual observation. It is arguable, moreover, that the identity of the entities referred to by the two expressions is established by the continuity between the two sets of observations as the observer moves towards or away from the cloud. In the case of brain processes and consciousness there is no such continuity between the two sets of observations involved. A closer introspective scrutiny will never reveal the passage of nerve impulses over a thousand synapses in the way that a closer scrutiny of a cloud will reveal a mass of tiny particles in suspension. The operations required to verify statements about consciousness and statements about brain processes are fundamentally different.

To find a parallel for this feature we must examine other cases where an identity is asserted between something whose occurrence is verified by the ordinary processes of observation and something whose occurrence is established by special procedures. For this purpose I have chosen the case where we say that lightning is a motion of electric charges. As in the case of consciousness, however closely we scrutinize the lightning we shall never be able to observe the electric charges, and just as the operations for determining the nature of one's state of consciousness are radically different from those involved in determining the nature of one's brain processes, so the operations for determining the occurrence of lightning are radically different from those involved in determining the occurrence of a motion of electric charges. What is it, therefore, that leads us to say that the two sets of observations are observations of the same event? It cannot be merely the fact that the two sets of observations are systematically correlated such that whenever there is lightning there is always a motion of electric charges. There are innumerable cases of such correlations where we have no temptation to say that the two sets of observations are observations of the same event. There is a systematic correlation, for example, between the movement of the tides and the stages of the moon, but this does not lead us to say that records of tidal levels are records of the moon's stages or vice versa. We speak rather of a causal connection between two independent events or processes.

The answer here seems to be that we treat the two sets of observations as observations of the same event in those cases where the technical scientific observations set in the context of the appropriate body of scientific theory provide an immediate explanation of the observations made by the man in the street. Thus we conclude that lightning is nothing more than a motion of electric charges, because we know that a motion of electric charges through the atmosphere, such as occurs when lightning is reported, gives rise to the type of visual stimulation which would lead an observer to report a flash of lightning. In the moon/tide case, on the other hand, there is no such direct causal connection between the stages of the moon and the observations made by the man who measures the height of the tide. The causal connection is between the moon and the tides, not between the moon and the measurement of the tides.

V The Physiological Explanation of Introspection and the Phenomenological Fallacy

If this account is correct, it should follow that in order to establish the identity of consciousness and certain processes in the brain, it would be necessary to show that the introspective observations reported by the subject can be accounted for in terms of processes which are known to have occurred in his brain. In the light of this suggestion it is extremely interesting to find that when a physiologist, as distinct from a philosopher, finds it difficult to see how consciousness could be a process in the brain, what worries him is not any supposed self- contradiction involved in such an assumption, but the apparent impossibility of accounting for the reports given by the subject of his conscious processes in terms of the known properties of the central nervous system. Sir Charles Sherrington has posed the problem as follows:

> The chain of events stretching from the sun's radiation entering the eye to, on the one hand,

> the contraction of the pupillary muscles, and on the other, to the electrical disturbances in the brain-cortex are all straightforward steps in a sequence of physical "causation," such as, thanks to science, are intelligible. But in the second serial chain there follows on, or attends, the stage of brain-cortex reaction an event or set of events quite inexplicable to us, which both as to themselves and as to the causal tie between them and what preceded them science does not help us; a set of events seemingly incommensurable with any of the events leading up to it. The self "sees" the sun; it senses a two-dimensional disc of brightness, located in the "sky," this last a field of lesser brightness, and overhead shaped as a rather flattened dome, coping the self and a hundred other visual things as well. Of hint that this is within the head there is none. Vision is saturated with this strange property called "projection," the unargued inference that what it sees is at a "distance" from the seeing "self." Enough has been said to stress that in the sequence of events a step is reached where a physical situation in the brain leads to a psychical, which however contains no hint of the brain or any other bodily part . . . The supposition has to be, it would seem, two continuous series of events, one physico-chemical, the other psychical, and at times interaction between them.[4]

Just as the physiologist is not likely to be impressed by the philosopher's contention that there is some self-contradiction involved in supposing consciousness to be a brain process, so the philosopher is unlikely to be impressed by the considerations which lead Sherrington to conclude that there are two sets of events, one physico-chemical, the other psychical. Sherrington's argument, for all its emotional appeal, depends on a fairly simply logical mistake, which is unfortunately all too frequently made by psychologists and physiologists and not infrequently in the past by the philosophers themselves. This logical mistake, which I shall refer to as the "phenomenological fallacy," is the mistake of supposing that when the subject describes his experience, when he describes how things look, sound, smell, taste, or feel to him, he is describing the literal properties of objects and events on a peculiar sort of internal cinema or television screen, usually referred to in the modern psychological literature as the "phenomenal field." If we assume, for example, that when a subject reports a green after-image he is asserting the occurrence inside himself of an object which is literally green, it is clear that we have on our hands an entity for which there is no place in the world of physics. In the case of the green after-image there is no green object in the subject's environment corresponding to the description that he gives. Nor is there anything green in his brain; certainly there is nothing which could have emerged when he reported the appearance of the green after-image. Brain processes are not the sort of things to which color concepts can be properly applied.

The phenomenological fallacy on which this argument is based depends on the mistaken assumption that because our ability to describe things in our environment depends on our consciousness of them, our descriptions of things are primarily descriptions of our conscious experience and only secondarily, indirectly, and inferentially descriptions of the objects and events in our environments. It is assumed that because we recognize things in our environment by their look, sound, smell, taste, and feel, we begin by describing their phenomenal properties, i.e. the properties of the looks, sounds, smell, tastes, and feels which they produce in us, and infer their real properties from their phenomenal properties. In fact, the reverse is the case. We begin by learning to recognize the real properties of things in our environment. We learn to recognize them, of course, by their look, sound, smell, taste, and feel; but this does not mean that we have to learn to describe the look, sound, smell, taste, and feel of things before we can describe the things themselves. Indeed, it is only after we have learned to describe the things in our environment that we learn to describe our consciousness of them. We describe our conscious experience not in terms of the mythological "phenomenal properties" which are supposed to inhere in the mythological "objects" in the mythological "phenomenal field," but by reference to the actual physical properties of the concrete physical objects, events, and processes which normally, though not perhaps in the present instance, give rise to the sort of conscious experience which we are trying to describe. In other words when we describe the after-image as green, we are not saying that there is something, the after-image, which is green; we are saying that we are having the sort of experience which we normally have when, and which we have learned to describe as, looking at a green patch of light.

Once we rid ourselves of the phenomenological fallacy we realize that the problem of explaining introspective observations in terms of brain processes is far from insuperable. We realize that there is nothing that the introspecting subject says about his conscious experiences which is inconsistent with anything the physiologist might want to say about the brain processes which cause him to describe the environment and his consciousness of that environment in the way he does. When the subject describes his experience by saying that a light which is in fact stationary appears to move, all the physiologist or physiological psychologist has to do in order to explain the subject's introspective observations is to show that the brain process which is causing the subject to describe his experience in this way is the sort of process which normally occurs when he is observing an actual moving object and which therefore normally causes him to report the movement of an object in his environment. Once the mechanism whereby the individual describes what is going on in his environment has been worked out, all that is required to explain the individual's capacity to make introspective observations is an explanation of his ability to discriminate between those cases where his normal habits of verbal descriptions are appropriate to the stimulus situation and those cases where they are not, and an explanation of how and why, in those cases where the appropriateness of his normal descriptive habits is in doubt, he learns to issue his ordinary descriptive protocols preceded by a qualificatory phrase like "it appears," "seems," "looks," "feels," etc.[5]

Notes

1 E. C. Tolman, *Purposive Behaviour in Animals and Men* (Berkeley 1932).

2 L. Wittgenstein, *Philosophical Investigations* (Oxford 1953); G. Ryle, *The Concept of Mind* (1949).

3 Place, "The Concept of Heed," *British Journal of Psychology* XLV (1954), 243–55.

4 Sir Charles Sherrington, *The Integrative Action of the Nervous System* (Cambridge 1947), pp. xx–xxi.

5 I am greatly indebted to my fellow-participants in a series of informal discussions on this topic which took place in the Department of Philosophy, University of Adelaide, in particular to Mr C. B. Martin for his persistent and searching criticism of my earlier attempts to defend the thesis that consciousness is a brain process, to Professor D. A. T. Gasking, of the University of Melbourne, for clarifying many of the logical issues involved, and to Professor J. J. C. Smart for moral support and encouragement in what often seemed a lost cause.