Chapter 2

The Mind–Body Problem

William G. Lycan

Human beings, and perhaps other creatures, have minds as well as bodies. But what is a mind, and what is its relation to body, or to the physical in general?

2.1 Mind–Body Dualism

The first answer to the mind–body question proposed since medieval times was that of Descartes, who held that minds are wholly distinct from bodies and from physical objects of any sort. According to *Cartesian dualism*, minds are purely spiritual and radically non-spatial, having neither size nor location. On this view, a normal living human being or person is a duality, a mind and a body paired (though there can be bodies without minds, and minds can survive the destruction of their corresponding bodies). Mysteriously, despite the radical distinctness of minds from bodies, they interact causally: bodily happenings cause sensations and experiences and thoughts in one's mind; conversely, mental activity leads to action and speech, causing the physical motion of limbs or lips.

Cartesian dualism has strong intuitive appeal, since from the inside our minds do not feel physical at all; and we can easily imagine their existing disembodied or, indeed, their existing in the absence of any physical world whatever. And until the 1950s, in fact, the philosophy of mind was dominated by Descartes's "first-person" perspective, our view of ourselves from the inside. With few exceptions, philosophers had accepted the following claims: (1) that one's own mind is better known than one's body, (2) that the mind is metaphysically in the body's driver's seat, and (3) that there is at least a theoretical problem of how we human intelligences can know that "external," everyday physical objects exist at all, even if there are tenable solutions to that problem. We human subjects are immured within a movie theatre of the mind, though we may have some defensible ways of inferring what goes on outside the theatre.

Midway through the past (twentieth) century, all this suddenly changed, for two reasons. The first reason was the accumulated impact of logical positivism and the verification theory of meaning. Intersubjective verifiability or testability 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 purposes and, it was felt, for serious metaphysical purposes also.

The second reason was the emergence of a number of pressing philosophical objections to Cartesian dualism, such as the following:

1 Immaterial Cartesian minds and ghostly non-physical events were increasingly seen to fit ill with our otherwise physical and scientific picture of the world, uncomfortably like spooks or ectoplasm themselves. They are not needed for the explanation of any publicly observable fact, for neurophysiology promises to explain the motions of our bodies in particular and to explain them completely. Indeed, ghost-minds could not very well help in such an explanation, since nothing is known of any properties of spookstuff that would bear on public physical occurrences.
2 Since human beings evolved over aeons, by purely physical processes of mutation and natural selection, from primitive creatures such as one-celled organisms which did not have minds, it is anomalous to suppose that at some point Mother Nature (in the form of population genetics) somehow created immaterial Cartesian minds in addition to cells and physical organs. The same point can be put in terms of the development of a single human zygote into an embryo, then a fetus, a baby, and finally a child.
3 If minds really are immaterial and utterly non-spatial, how can they possibly interact causally with physical objects in space? (Descartes himself was very uncomfortable about this. At one point he suggested gravity as a model for the action of something immaterial on a physical body; but gravity is spatial in nature even though it is not tangible in the way that bodies are.)
4 In any case it does not seem that immaterial entities could cause physical motion consistently with the conservation laws of physics, such as those regarding motion and matter-energy; physical energy would have to vanish and reappear inside human brains.

2.2 Behaviorism

What alternatives are there to dualism? First, Carnap (1932–3) and Ryle (1949) noted that the obvious verification conditions or tests 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 damage or disorder, or that you believe that parsnips are dangerous, save by your verbal avowals and your avoidance of parsnips? If the tests are behavioral, then (it was argued) the very meanings of the ascriptions, or at least the only facts genuinely described, are not ghostly or ineffable but behavioral. Thus behaviorism as a theory of mind and a paradigm for psychology.

In academic psychology, behaviorism took primarily a methodological form, and the psychologists officially made no metaphysical claims. But 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, "Leo is in pain" means, not anything about Leo's putative ghostly ego, or even about any episode taking place within Leo, but that either Leo is actually behaving in a wincing and groaning way or he is disposed so to behave (in that he would so behave were something not keeping him from doing so). "Leo believes that parsnips are dangerous" means just that, if asked, Leo would assent to that proposition, and, if confronted by a parsnip, Leo would shun it, and so forth.

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 tests, to the extent of fooling a limited observer (say, a human being corresponding with it by mail) 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 non-verbal responses, any behaviorist, psychological or philosophical, would agree that psychological differences cannot outrun behavioral tests; organisms (including machines) whose actual and hypothetical behavior is just the same are psychologically just alike.

Besides solving the methodological problem of intersubjective verification, philosophical behaviorism also adroitly avoided a number of the objections to Cartesian dualism, including all of (1)–(4) listed above. It dispensed with immaterial Cartesian egos and ghostly non-physical events, writing them off as metaphysical excrescences. It disposed of Descartes's admitted problem of mind–body interaction, since it posited no immaterial, non-spatial causes of behavior. It raised no scientific mysteries concerning the intervention of Cartesian substances in physics or biology, since it countenanced no such intervention. Thus it is a *materialist* view, as against Descartes's immaterialism.

Yet some theorists were uneasy; they felt that in its total repudiation of the inner, the private, and the subjective, behaviorism was leaving out something real and important. When this worry was voiced, the behaviorists often replied with mockery, assimilating the doubters to old-fashioned dualists who believed in

ghosts, ectoplasm, or the Easter bunny; behaviorism was the only (even halfway sensible) game in town. Nonetheless, the doubters made several lasting points against it. First, people who are honest and not anesthetized know perfectly well that they experience, and can introspect, actual inner mental episodes or occurrences, that are neither actually accompanied by characteristic behavior nor merely static hypothetical facts of how they would behave if subjected to such-and-such a stimulation. Place (1956) spoke 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). Secondly, contrary to the Turing Test, it seems perfectly possible for two people to differ psychologically despite total similarity of their actual and hypothetical behavior, as in a case of "inverted spectrum" as hypothesized by John Locke: it might be that when you see a red object, you have the sort of color experience that I have when I see a green object, and vice versa. For that matter, a creature might exhibit all the appropriate stimulus-response relations and lack a mental life entirely; we can imagine building a "zombie" or stupid robot that behaves in the right ways but does not really feel or think anything at all (Block and Fodor 1972; Kirk 1974; Block 1981; Campbell 1984). Thirdly, 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: 8; Block 1981); for example, if Leo believes that parsnips are dangerous and he is offered parsnips, he would shun them only if he does not want to die. Therefore, the behaviorist analyses are either circular or radically incomplete, so far as they are supposed to exhaust the mental generally.

So matters stood in stalemate between dualists, behaviorists, and doubters, until the late 1950s, when U. T. Place (1956) and J. J. C. Smart (1959) proposed a middle way, a conciliatory compromise solution.

2.3 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 neither ghostly nor non-physical. 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, for example, to have one's c-fibers, or more likely a-fibers, firing in the central nervous system; 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 objections to dualism. But it also

brilliantly accommodated the inner and the episodic as behaviorism 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 can differ mentally despite total outward behavioral similarity, since clearly organisms can differ neurophysiologically in mediating their outward stimulus-response regularities; that would afford the possibility of inverted spectrum. 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 (1966, 1972) and D. M. Armstrong (1968). Lewis and Armstrong maintained that mental terms were defined causally, in terms of mental items' typical causes and effects. For instance, the word "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." More elaborately, Lewis held that mental terms are the theoretical terms of a commonsensical "folk theory," and with the positivists that all theoretical terms are implicitly defined by the theories in which they occur. That common-sense theory has since come to be called "folk psychology.") Now if, by definition, pain is whatever state occupies a certain causal niche, and if, as is overwhelmingly likely, scientific research will reveal that that particular niche is in fact occupied by such-and-such a neurophysiological state, it follows straightaway 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 convoluted elaboration of this already hybrid view is developed by Lewis 1980; for meticulous discussion, see Block 1978; Shoemaker 1981; Tye 1983; Owens 1986.)

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 was itself no such claim, 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; and temperature 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 non-physical but right in thinking them inner and episodic; the behaviorists were right in their materialism but wrong to repudiate inner mental episodes. A delightful synthesis. But alas, it was too good to be true.

2.4 Machine Functionalism

Quite soon, Hilary Putnam (1960, 1967a, 1967b) and Jerry Fodor (1968b) pointed out a presumptuous implication of the identity theory understood as a theory of "types" or kinds of mental item: 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 are in order to have what can be accurately recognized as pain? The identity theorist had overreacted to the behaviourists' difficulties and focused too narrowly on the specifics of biological humans' actual inner states, and in so doing they had fallen into species chauvinism.

Putnam and Fodor 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 they 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 – in other words, with the states that happen to be playing the appropriate roles; this is the thesis of *token identity* or "token" materialism or 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, behavioral responses, and other intervening psychological states.

Functionalism, then, is the doctrine that what makes a mental state the type of state it is – a pain, a smell of violets, a belief that koalas are venomous – is its distinctive set of functional relations, its role in its subject's behavioral economy.

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 can a psychological "program" 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, asserts 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 it 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, presumably, there are no non-physical 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. Moreover, unlike behaviorism and the identity theory, functionalism does not strictly entail that minds are physical; it might be true of non-physical minds, so long as those minds realized the relevant programs.

2.5 Homuncular Functionalism and Other Teleological Theories

Machine functionalism has been challenged on a number of points, which together motivate a specifically teleological notion of "function": we are to think of a thing's function as what the thing is for, what its job is, what it is supposed to do. Here are three reasons for thus "putting the function back into functionalism" (Sober 1985).

First, the machine functionalist still conceived psychological explanation in the logical positivists' terms of subsuming observed data under wider and wider

universal laws. But Fodor (1968a), Dennett (1978), and Cummins (1983) 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 thought of as a little agent or bureaucrat operating within its containing subject; 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 neurophysiological structures. Thus biological and mechanical systems alike are hierarchically organized. (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.) But nothing in this pattern of explanation corresponds to the subsumption of data under wider and wider universal generalizations.

The second reason is that 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); any middle-sized physical object has some set of component molecular motions that happen to correspond one–one to a given machine program. 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. (Machine functionalism took "function" in its spare mathematical sense rather than in a genuinely functional sense. One should note that, as used here, the term "machine functionalism" is tied to the original liberal conception of "realizing;" so to impose a teleological restriction is to abandon machine functionalism.)

Thirdly, Van Gulick (1980), Millikan (1984), Dretske (1988), Fodor (1990a), and others have argued powerfully that teleology must enter into any adequate analysis of the intentionality or aboutness or referential character of mental states such as beliefs and desires, by reference to the states' psychobiological functions.

Beliefs, desires, and other propositional attitudes such as suspecting, intending, and wishing are directed upon states of affairs which may or may not actually obtain (for instance, that the Republican candidate will win), and are about individuals who may or may not exist (such as King Arthur or Sherlock Holmes). Franz Brentano (1973 [1874]) drew a distinction between psychological phenomena, which are directed upon objects and states of affairs, even non-existing ones, and physical objects, which are not so directed. If mental items are physical, however, the question arises how any purely physical entity or state could have the property of being "directed upon" or about a non-existent state of affairs or object; that is not the sort of feature that ordinary, purely physical objects (such as bricks) can have. 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. If teleology is needed to explicate intentionality, and machine functionalism affords no teleology, then machine functionalism is not adequate to explicate intentionality.

All this talk of teleology and biological function seems to presuppose that biological and other "structural" states of physical systems really do have functions in the teleological sense. The latter claim is, to say the least, controversial. But, fortunately for the teleological functionalist, there is a 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 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 (Wright 1973; Millikan 1984).

Functionalism inherits some of the same difficulties that earlier beset behaviorism and the identity theory. These remaining obstacles fall into two main categories: qualia problems and intentionality problems.

2.6 Problems over Qualia and Consciousness

The quale of a mental state or event (particularly a sensation) is that state or event's feel, its introspectible "phenomenal character," its nature as it presents itself to consciousness. Many philosophers have objected that neither functionalist metaphysics nor any of the allied doctrines aforementioned can "explain consciousness," or illuminate 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.

"The" problem of consciousness or qualia is familiar. Indeed, it is so familiar that we tend to overlook the most important thing about it: that its name is

legion, for it is many. There is no single problem of qualia; there are at least eleven quite distinct objections that have been brought against functionalism (some of them apply to materialist views generally). To mention a few:

1 Block (1978) and others have urged various "zombie"-style counterexample cases against functionalism – examples in which some entity seems to realize the right program but which lacks one of mentality's crucial qualitative aspects. (Typically the "entity" is a group of human beings, such as the entire population of China acting according to an elaborate set of instructions. It does not seem that such a group of individuals would collectively be feeling anything.) Predictably, functionalists have rejoined by arguing, for each example, either that the proposed group entity does not in fact succeed in realizing the right program (for example, because the requisite teleology is lacking) or that there is no good reason for denying that the entity does have the relevant qualitative states.
2 Nagel (1974) and Jackson (1982) have appealed to a disparity in knowledge, as a general anti-materialist argument: I can know what it is like to have such-and-such a sensation only if I have had that sensation myself; no amount of objective, third-person scientific information would suffice. In reply, functionalists have offered analyses of "perspectivalness," complete with accounts of "what it is like" to have a sensation, that make those things compatible with functionalism. Nagel and Jackson have argued, further, for the existence of a special, intrinsically perspectival kind of fact, the fact of "what it is like", which intractably and in principle cannot be captured or explained by physical science. Functionalists have responded that the arguments commit a logical fallacy (specifically, that of applying Leibniz's Law in an intensional context); some have added that in any case, to "know what it is like" is merely to have an ability, and involves no fact of any sort, while, contrariwise, some other theorists have granted that there are facts of "what it is like" but insisted that such facts can after all be explained and predicted by natural science.
3 Saul Kripke (1972) made ingenious use of modal distinctions against type or even token identity, arguing that unless mental items are necessarily identical with neurophysiological ones, which they are not, they cannot be identical with them at all. Kripke's close reasoning has attracted considerable critical attention. And even more sophisticated variants have been offered, e.g., by Jackson (1993) and Chalmers (1996).
4 Jackson (1977) and others have defended the claim that in consciousness we are presented with mental individuals that themselves bear phenomenal, qualitative properties. For instance, when a red flash bulb goes off in your face, your visual field exhibits a green blotch, an "after-image," a thing that is really green and has a fairly definite shape and exists for a few seconds before disappearing. If there are such things, they are entirely different from anything physical to be found in the brain of a (healthy) human subject. Belief in such "phenomenal individuals" as genuinely green after-images has been unpopular

among philosophers for some years, but it can be powerfully motivated (see Lycan 1987: 83–93).

This is a formidable quartet of objections, and, on the face of it, each is plausible. Materialists and particularly functionalists must respond in detail. Needless to say, materialists have responded at length; some of the most powerful rejoinders are formulated in Lycan (1987, 1996). Yet recent years have seen some reaction against the prevailing materialism, including a re-emergence of some neo-dualist views, as in Robinson (1988), Hart (1988), Strawson (1994), and Chalmers (1996).

2.7 Problems over Intentionality

The problem arising from our mention of Brentano was to explain how any purely physical entity or state could have the property of being about or "directed upon" a non-existent state of affairs. The standard functionalist reply is that propositional attitudes have Brentano's feature because the internal physical states and events that realize them represent actual or possible states of affairs. What they represent (their content) is determined at least in part by their functional roles.

There are two main difficulties. One is that of saying exactly how a physical item's supposed representational content is determined; in virtue of what does a neurophysiological state represent precisely that the Republican candidate will win? An answer to that general question is what Fodor has called a *psychosemantics*. Several attempts have been made (Dretske 1981; Millikan 1984; Fodor 1987, 1990a, 1990b, 1994), but none is very plausible. In particular, none applies to any content but that which involves actual and presently existing physical objects. Abstract entities such as numbers, future entities such as a child I hope one day to have, and Brentano's non-existent items, are just left out.

The second difficulty is that ordinary propositional attitude contents do not supervene on the states of their subjects' nervous systems, but are underdetermined by even the total state of that subject's head. Putnam's (1975) Twin Earth and indexical examples show that, surprising as it may seem, two human beings could be molecule-for-molecule alike and still differ in their beliefs and desires, depending on various factors in their spatial and historical environments. Thus we can distinguish between "narrow" properties, those that are determined by a subject's intrinsic physical composition, and "wide" properties, those that are not so determined. Representational contents are wide, yet functional roles are, ostensibly, narrow. How, then, can propositional attitudes be type-identified with functional roles, or for that matter with states of the brain under any narrow description?

Functionalists have responded in either of two ways to the second difficulty. The first is to understand "function" widely as well, specifying functional roles historically and/or by reference to features of the subject's actual environment.

The second is simply to abandon functionalism as an account of content in particular, giving some alternative psychosemantics for propositional attitudes, but preserving functionalism in regard to attitude types. (Thus what makes a state a desire that P is its functional role, even if something else makes the state a desire that P).

2.8 The Emotions

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). Armstrong (1968: ch. 8, secn III) essentially took this line, as do Solomon (1977) and Gordon (1987). However, there is a literature on functionalism and the emotions; see Rey (1980) and some of the other papers collected in Rorty (1980). Griffiths (1997) takes a generally functionalist view, but argues that "the emotions" do not constitute a single kind.

2.9 Instrumentalism

The identity theorists and the functionalists, machine or teleological, joined common sense (and current cognitive psychology) in understanding mental states and events both as internal to human subjects and as causes. Beliefs and desires in particular are thought to be caused by perceptual or other cognitive events and as in turn conspiring from within to cause behavior. If Armstrong's or Lewis's theory of mind is correct, this idea is not only common-sensical but a conceptual truth; if functionalism is correct, it is at least a metaphysical fact.

In rallying to the inner-causal story, as we saw in section 2.3, the identity theorists and functionalists broke with the behaviorists, for behaviorists did not think of mental items as entities, as inner, or as causes in any stronger sense than the bare hypothetical. Behaviorists either dispensed with the mentalistic idiom altogether, or paraphrased mental ascriptions in terms of putative responses to hypothetical stimuli. More recently, other philosophers have followed them in rejecting the idea of beliefs and desires as inner causes and in construing them in a more purely operational or instrumental fashion. D. C. Dennett (1978, 1987) has been particularly concerned to deny that beliefs and desires are causally active inner states of people, and maintains instead that belief-ascriptions and desire-ascriptions are merely calculational devices, which happen to have predictive

usefulness for a reason that he goes on to explain. Such ascriptions are often objectively true, he grants, but not in virtue of describing inner mechanisms.

Thus Dennett is an *instrumentalist* about propositional attitudes such as belief and desire. (According to a contemporary interpretation, an "instrumentalist" about Xs is a theorist who claims that although sentences about "Xs" are often true, they do not really describe entities of a special kind, but only serve to systematize more familiar phenomena. For instance, we are all instrumentalists about "the average American homeowner," who is white, male, and the father of exactly 2.2 children.) To ascribe a "belief" or a "desire" is not to describe some segment of physical reality, Dennett says, but is more like moving a group of beads in an abacus. (It should be noted that Dennett has more recently moderated his line: see 1991.)

Dennett offers basically four grounds for his rejection of the common-sensical inner-cause thesis:

1 He thinks it quite unlikely that any science will ever turn up any distinctive inner-causal mechanism that would be shared by all the possible subjects that had a particular belief.
2 He compares the belief-desire interpretation of human beings to that of lower animals, chess-playing computers, and even lightning-rods, arguing that (a) in their case we have no reason to think of belief-ascriptions and desire-ascriptions as other than mere calculational-predictive devices and (b) we have no more reason for the case of humans to think of belief-ascriptions and desire-ascriptions as other than that.
3 Dennett argues from the verification conditions of belief-ascriptions and desire-ascriptions – basically a matter of extrapolating rationally from what a subject ought to believe and want in his or her circumstances – and then he boldly just identifies the truth-makers of those ascriptions with their verification conditions, challenging inner-cause theorists to show why instrumentalism does not accommodate all the actual evidence.
4 He argues that in any case, if a purely normative assumption (the "rationality assumption," which is that people will generally believe what they ought to believe and desire what they should desire) is required for the licensing of an ascription, then the ascription cannot itself be a purely factual description of a plain state of affairs.

Stich (1981) explores and criticizes Dennett's instrumentalism at length (perhaps oddly, Stich (1983) goes on to defend a view nearly as deprecating as Dennett's, though clearly distinct from it). Dennett (1981) responds to Stich, bringing out more clearly the force of the "rationality assumption" assumption. (Other criticisms are levelled against Dennett by commentators in the *Behavioral and Brain Sciences* symposium that is headed by Dennett 1988.)

A close cousin of Dennett's view, in that it focuses on the rationality assumption, is Donald Davidson's (1970) *anomalous monism*. Unlike Dennett's instrumentalism,

it endorses token physicalism and insists that individual mental tokens are causes, but it rejects on similarly epistemological grounds the possibility of any interesting materialistic type-reduction of the propositional attitudes.

2.10 Eliminativism and Neurophilosophy

Dennett's instrumentalism breaks with common sense and with philosophical tradition in denying that propositional attitudes such as belief and desire are real inner-causal states of people. But Dennett concedes – indeed, he urgently insists – that belief-ascriptions and desire-ascriptions are true, and objectively true, nonetheless. Other philosophers have taken a less conciliatory, more radically uncommon-sensical view: that mental ascriptions are not true after all, but are simply false. Common sense is just mistaken in supposing that people believe and desire things, and perhaps in supposing that people have sensations and feelings, disconcerting as that nihilistic claim may seem.

Following standard usage, let us call the nihilistic claim "eliminative materialism," or "eliminativism" for short. It is important to note a customary if unexpected alliance between the eliminativist and the token physicalist: the eliminativist, the identity theorist, and the functionalist all agree that mental items are, if anything, real inner-causal states of people. They disagree only on the empirical question of whether any real neurophysiological states of people do in fact answer to the common-sensical mental categories of "folk psychology." Eliminativists praise identity theorists and functionalists for their forthright willingness to step up and take their empirical shot. Both eliminativists and token physicalists scorn the instrumentalist's sleazy evasion. (But eliminativists agree with instrumentalists that functionalism is a pipe-dream, and functionalists agree with instrumentalists that mental ascriptions are often true and obviously so. The three views form an eternal triangle of a not uncommon sort.)

Paul Feyerabend (1963a, 1963b) was the first to argue openly that the mental categories of folk psychology simply fail to capture anything in physical reality and that everyday mental ascriptions were therefore false. (Rorty (1965) took a notoriously eliminativist line also, but, following Sellars (1963), tried to soften its nihilism; Lycan and Pappas (1972) argued that the softening served only to collapse Rorty's position into incoherence.) Feyerabend attracted no great following, presumably because of his view's outrageous flouting of common sense. But eliminativism was resurrected by Paul Churchland (1981) and others, and defended in more detail.

Churchland argues mainly from the poverty of "folk psychology;" he claims that historically, when other primitive theories such as alchemy have done as badly on scientific grounds as folk psychology has, they have been abandoned, and rightly so. P. S. Churchland (1986) and Churchland and Sejnowski (1990) emphasize the comparative scientific reality and causal efficacy of neurobiological

mechanisms: given the scientific excellence of neurophysiological explanation and the contrasting diffuseness and type-irreducibility of folk psychology, why should we suppose – even for a minute, much less automatically – that the platitudes of folk psychology express truths?

Reasons for rejecting eliminativism are obvious. First, we think we know there are propositional attitudes because we introspect them in ourselves. Secondly, the attitudes are indispensable to prediction, reasoning, deliberation, and understanding, and to the capturing of important macroscopic generalizations. We could not often converse coherently without mention of them. But what of P. M. Churchland's and P. S. Churchland and Sejnowski's arguments?

One may dispute the claim that folk psychology is a failed or bad theory; Kitcher (1984) and Horgan and Woodward (1985) take this line. Or one may dispute the more basic claim that folk psychology is a theory at all. Ryle (1949) and Wittgenstein (1953) staunchly opposed that claim before it had explicitly been formulated. More recent critics include Morton (1980), Malcolm (1984), Baker (1988), McDonough (1991), and Wilkes (1993).

References

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

Baker, L. R. (1988). *Saving Belief*. Princeton, NJ: Princeton University Press.

Block, N. J. (1978). "Troubles with Functionalism." In W. Savage (ed.), *Minnesota Studies in the Philosophy of Science, Vol. IX: Perception and Cognition*. Minneapolis: University of Minnesota Press: 261–325. Excerpts reprinted in Lycan (1990, 1999).

—— (ed.) (1980). *Readings in Philosophy of Psychology*, 2 vols. Cambridge, MA: Harvard University Press.

—— (1981). "Psychologism and Behaviorism." *Philosophical Review*, 90: 5–43.

Block, N. J. and Fodor, J. A. (1972). "What Psychological States Are Not." *Philosophical Review*, 81: 159–81. Reprinted in Block (1980).

Brentano, F. (1973 [1874]). *Philosophy from an Empirical Standpoint*. London: Routledge and Kegan Paul.

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

Carnap, R. (1932–3). "Psychology in Physical Language." *Erkenntnis*, 3: 107–42. Excerpt reprinted in Lycan (1990).

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

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

Churchland, P. M. (1981). "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy*, 78: 67–90. Reprinted in Lycan (1990, 1999).

Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: Bradford Books/MIT Press.

Churchland, P. S. and Sejnowski, T. (1990). "Neural Representation and Neural Computation." In Lycan (1990): 224–52. Reprinted in Lycan (1999).

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

Davidson, D. (1970). "Mental Events." In L. Foster and J. W. Swanson (eds.), *Experience and Theory*. Amherst, MA: University of Massachusetts Press: 79–101. Reprinted in Block (1980) and in Lycan (1999).

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

—— (1981). "Making Sense of Ourselves." *Philosophical Topics*, 12: 63–81. Reprinted in Lycan (1990).

—— (1987). *The Intentional Stance*. Cambridge, MA: Bradford Books/MIT Press.

—— (1988). "Précis of *The Intentional Stance*." *Behavioral and Brain Sciences*, 11: 495–505.

—— (1991). "Real Patterns." *Journal of Philosophy*, 88: 27–51.

Dretske, F. (1981). *Knowledge and the Flow of Information*. Cambridge, MA: Bradford Books/MIT Press.

—— (1988). *Explaining Behavior*. Cambridge, MA: Bradford Books/MIT Press.

Feyerabend, P. (1963a). "Materialism and the Mind–Body Problem." *Review of Metaphysics*, 17: 49–66.

—— (1963b). "Mental Events and the Brain." *Journal of Philosophy*, 60: 295–6.

Fodor, J. A. (1968a). "The Appeal to Tacit Knowledge in Psychological Explanation." *Journal of Philosophy*, 65: 627–40.

—— (1968b). *Psychological Explanation*. New York, NY: Random House.

—— (1987). *Psychosemantics*. Cambridge, MA: MIT Press.

—— (1990a). "Psychosemantics." In Lycan (1990): 312–37.

—— (1990b). *A Theory of Content*. Cambridge, MA: Bradford Books/MIT Press.

—— (1994). *The Elm and the Expert*. Cambridge, MA: Bradford Books/MIT Press.

Geach, P. (1957). *Mental Acts*. London: Routledge and Kegan Paul.

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

Griffiths, P. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.

Hart, W. D. (1988). *Engines of the Soul*. Cambridge: Cambridge University Press.

Horgan, T. and Woodward, J. (1985). "Folk Psychology is Here to Stay." *Philosophical Review*, 94: 197–226. Reprinted in Lycan (1990, 1999).

Jackson, F. (1977). *Perception*. Cambridge: Cambridge University Press.

—— (1982). "Epiphenomenal Qualia." *Philosophical Quarterly*, 32: 127–36. Reprinted in Lycan (1990, 1999).

—— (1993). "Armchair Metaphysics." In J. O'Leary-Hawthorne and M. Michael (eds.), *Philosophy in Mind*. Dordrecht: Kluwer Academic Publishing.

Kirk, R. (1974). "Zombies vs. Materialists." *Aristotelian Society Supplementary Volume*, 48: 135–52.

Kitcher, P. (1984). "In Defense of Intentional Psychology." *Journal of Philosophy*, 81: 89–106.

Kripke, S. (1972). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. (1966). "An Argument for the Identity Theory." *Journal of Philosophy*, 63: 17–25.

—— (1972). "Psychophysical and Theoretical Identifications." *Australasian Journal of Philosophy*, 50: 249–58. Reprinted in Block (1980).

—— (1980). "Mad Pain and Martian Pain." In Block (1980).

Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press/Bradford Books.

—— (ed.) (1990). *Mind and Cognition: A Reader*. Oxford: Blackwell.

—— (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press/Bradford Books.
—— (ed.) (1999). *Mind and Cognition: An Anthology*. Oxford: Blackwell.
Lycan, W. and Pappas, G. (1972). "What is Eliminative Materialism?" *Australasian Journal of Philosophy*, 50: 149–59.
Malcolm, N. (1984). "Consciousness and Causality." In D. Armstrong and N. Malcolm, *Consciousness and Causality: A Debate on the Nature of Mind*. Oxford: Blackwell.
McDonough, R. (1991). "A Culturalist Account of Folk Psychology." In J. Greenwood (ed.), *The Future of Folk Psychology*. Cambridge: Cambridge University Press: 263–88.
Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: Bradford Books/MIT Press.
Morton, A. (1980). *Frames of Mind*. Oxford: Oxford University Press.
Nagel, T. (1974). "What Is It Like to be a Bat?" *Philosophical Review*, 83: 435–50. Reprinted in Block (1980).
Owens, J. (1986). "The Failure of Lewis' Functionalism." *Philosophical Quarterly*, 36: 159–73.
Place, U. T. (1956). "Is Consciousness a Brain Process?" *British Journal of Psychology*, 47: 44–50. Reprinted in Lycan (1990, 1999).
Putnam, H. (1960). "Minds and Machines." In S. Hook (ed.), *Dimensions of Mind*. New York: Collier Books: 136–64.
—— (1967a). "The Mental Life of Some Machines." In H.-N. Castañeda (ed.), *Intentionality, Minds, and Perception*. Detroit, MI: Wayne State University Press: 177–200.
—— (1967b). "Psychological Predicates." In W. H. Capitan and D. Merrill (eds.), *Art, Mind, and Religion*, Pittsburgh, PA: University of Pittsburgh Press: 37–48. Reprinted in Block (1980) under the title "The Nature of Mental States."
—— (1975). "The Meaning of 'Meaning'." In *Philosophical Papers*. Cambridge: Cambridge University Press.
Rey, G. (1980). "Functionalism and the Emotions." In Rorty (1980): 163–95.
Robinson, W. S. (1988). *Brains and People*. Philadelphia, PA: Temple University Press.
Rorty, A. O. (ed.) (1980). *Explaining Emotions*. Berkeley and Los Angeles, CA: University of California Press.
Rorty, R. (1965). "Mind–Body Identity, Privacy, and Categories." *Review of Metaphysics*, 19: 24–54.
Ryle, G. (1949). *The Concept of Mind*. New York, NY: Barnes and Noble.
Sellars, W. (1963). *Science, Perception and Reality*. London: Routledge and Kegan Paul.
Shoemaker, S. (1981). "Some Varieties of Functionalism." *Philosophical Topics*, 12: 93–119.
Smart, J. J. C. (1959). "Sensations and Brain Processes." *Philosophical Review*, 68: 141–56.
Sober, E. (1985). "Panglossian Functionalism and the Philosophy of Mind." *Synthese*, 64: 165–93. Revised excerpt reprinted in Lycan (1990, 1999) under the title "Putting the Function Back Into Functionalism."
Solomon, R. (1977). *The Passions*. New York, NY: Doubleday.
Stich, S. (1981). "Dennett on Intentional Systems." *Philosophical Topics*, 12: 39–62. Reprinted in Lycan (1990, 1999).
—— (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: Bradford Books/MIT Press.
Strawson, G. (1994). *Mental Reality*. Cambridge, MA: Bradford Books/MIT Press.

Turing, A. (1964). "Computing Machinery and Intelligence." In A. R. Anderson (ed.), *Minds and Machines*. Englewood Cliffs, NJ: Prentice-Hall: 4–30.

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

Van Gulick, R. (1980). "Functionalism, Information, and Content." *Nature and System*, 2: 139–62.

Wilkes, K. (1993). "The Relationship Between Scientific and Common Sense Psychology." In S. Christensen and D. Turner (eds.), *Folk Psychology and the Philosophy of Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates: 144–87.

Wittgenstein, L. (1953). *Philosophical Investigations*, trans. G. E. M. Anscombe. New York, NY: Macmillan.

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