

JECTIVITY and in many respects opaque to its owner.

See also PSYCHOLOGY AND PHILOSOPHY.

BIBLIOGRAPHY

- Cavell, M. 1993. *The Psychoanalytic Mind: From Freud to Philosophy*. Cambridge, MA.: Harvard University Press.
- Davidson, D. 1982. Paradoxes of irrationality. In *Philosophical Essays on Freud*, ed. R. Wollheim and J. Hopkins. Cambridge University Press.
- Freud, S. 1953–74. *Standard Edition of the Complete Psychological Works of Sigmund Freud*, 24 vols, trans. under the general editorship of J. Strachey, in collaboration with A. Freud, assisted by A. Strachey and A. Tyson. London: Hogarth Press and the Institute of Psycho-Analysis.
- Gardner, S. 1993. *Irrationality and the Philosophy of Psychoanalysis*. Cambridge University Press.
- Hopkins, J. 1988. Epistemology and depth psychology: critical notes on *The Foundations of Psychoanalysis*. *Mind, Psychoanalysis and Science*, ed. P. Clark and C. Wright. Oxford: Basil Blackwell.
- Hopkins, J. 1991. The interpretation of dreams. In *The Cambridge Companion to Freud*, ed. J. Neu. Cambridge University Press.
- Hopkins, J. 1992. Psychoanalysis, interpretation, and science. In *Psychoanalysis, Mind and Art: Perspectives on Richard Wollheim*, ed. J. Hopkins and A. Savile. Oxford: Basil Blackwell.
- Klein, M. 1951, 1975. *The Writings of Melanie Klein*, 4 vols, ed. R. E. Money-Kyrle. London: Hogarth Press and the Institute of Psycho-Analysis.
- Wollheim, R. 1984. *The Thread of Life*. Cambridge University Press.
- Wollheim, R. 1991. *Freud*, 2nd edn. London: Harper Collins.
- Wollheim, R., ed. 1974. *Freud: a Collection of Critical Essays*. New York: Anchor Doubleday. Reprinted as *Philosophers on Freud: New Evaluations*. New York: Aronson (1977).
- Wollheim, R., and Hopkins, J. eds. 1982. *Philosophical Essays on Freud*. Cambridge University Press.

psychology and philosophy The last two decades have been a period of extraordinary change in psychology. COGNITIVE PSYCHOLOGY, which focuses on higher mental processes like reasoning, decision making, problem solving, language processing and higher-level visual processing, has become a – perhaps the – dominant paradigm among experimental psychologists, while behaviouristically oriented approaches have gradually fallen into disfavour (see BEHAVIOURISM). Largely as a result of this paradigm shift, the level of interaction between the disciplines of philosophy and psychology has increased dramatically. The goal of this article is to sketch some of the areas in which these interactions have been most productive, or at least most provocative. The interactions I will discuss fall into three categories, though the boundary between the first two is sometimes rather fuzzy.

COGNITIVE PSYCHOLOGY AS A SUBJECT FOR 'DESCRIPTIVE' PHILOSOPHY OF SCIENCE

One of the central goals of the philosophy of science is to provide explicit and systematic accounts of the theories and explanatory strategies exploited in the sciences. Another common goal is to construct philosophically illuminating analyses or explications of central theoretical concepts invoked in one or another science. In the philosophy of biology, for example, there is a rich literature aimed at understanding teleological explanations, and there has been a great deal of work on the structure of evolutionary theory and on such crucial concepts as fitness and biological function (see DRETSKE; TELEOLOGY). The philosophy of physics is another area in which studies of this sort have been actively pursued. (For an excellent example in the philosophy of biology, see Sober, 1984; in the philosophy of physics, see Sklar, 1974). In undertaking this work, philosophers need not (and typically do not) assume that there is anything wrong with the science they are studying. Their goal is simply to provide accounts of

SEBASTIAN GARDNER

the theories, concepts and explanatory strategies that scientists are using – accounts that are more explicit, systematic and philosophically sophisticated than the often rather rough-and-ready accounts offered by the scientists themselves.

Cognitive psychology is in many ways a curious and puzzling science. Many of the theories put forward by cognitive psychologists make use of a family of 'intentional' concepts – like *believing* that p, *desiring* that q, and *representing* r – which don't appear in the physical or biological sciences, and these intentional concepts play a crucial role in many of the explanations offered by these theories (see INTENTIONALITY; PROPOSITIONAL ATTITUDES). People's decisions and actions are explained by appeal to their beliefs and desires. Perceptual processes, some of which may themselves be representational, are said to result in mental states which represent (or sometimes *misrepresent*) one or another aspect of the cognitive agent's environment (see PERCEPTION; REPRESENTATION). While cognitive psychologists occasionally say a bit about the nature of intentional concepts and the explanations that exploit them, their comments are rarely systematic or philosophically illuminating. Thus it is hardly surprising that many philosophers have seen cognitive psychology as fertile ground for the sort of careful descriptive work that is done in the philosophy of biology and the philosophy of physics. Jerry FODOR's *The Language of Thought* (1975) was a pioneering study in this genre, one that continues to have a major impact on the field. Robert Cummins (1983, 1989), Daniel DENNETT (1978a, 1987) and John Haugeland (1978) have also done important and widely discussed work in the what might be called the 'descriptive' philosophy of cognitive psychology.

PHILOSOPHY AS A SOURCE OF PROPOSALS FOR IMPROVING COGNITIVE PSYCHOLOGY

The goal of the projects discussed in the previous section is to provide accurate, illum-

inating descriptions of what is going on in cognitive psychology. These philosophical accounts of cognitive theories and the concepts they invoke are generally much more explicit than the accounts provided by psychologists, and they inevitably smooth over some of the rough edges of scientists' actual practice. But if the account they give of cognitive theories diverges significantly from the theories that psychologists actually produce, then the philosophers have just gotten it wrong. There is, however, a very different way in which philosophers have approached cognitive psychology. Rather than merely trying to characterize what cognitive psychology is actually doing, some philosophers try to say what it *should* and *should not* be doing. Their goal is not to explicate scientific practice, but to criticize and improve it. The most common target of this critical approach is the use of intentional concepts in cognitive psychology. Intentional notions have been criticized on various grounds. The two that I will consider here are that they fail to supervene on the physiology of the cognitive agent, and that they cannot be 'naturalized'.

Perhaps the easiest way to make the point about SUPERVENIENCE is to use a thought experiment of the sort originally proposed by Hilary PUTNAM (1975). Suppose that in some distant corner of the universe there is a planet, Twin Earth, which is very similar to our own. On Twin Earth there is a person who is an atom-for-atom replica of President Clinton. Now the President Clinton who lives on Earth believes that Vice President Gore was born in Tennessee. If you asked him, 'Was Gore born in Tennessee?' he'd say, 'Yes'. Twin-Clinton would respond in the same way. But it is not because he believes that our Gore was born in Tennessee. Twin-Clinton has no beliefs at all about our Gore. His beliefs are about Twin-Gore, and Twin-Gore was certainly *not* born in Tennessee. Indeed, we may even suppose that Twin-Gore was not born in Twin-Tennessee, and thus that Clinton's belief is true while Twin-Clinton's is false. What all this is supposed to show is that two people can share all their physio-

logical properties without sharing all their intentional properties. To turn this into a problem for cognitive psychology, two additional premises are needed. The first is that cognitive psychology attempts to explain behaviour by appeal to people's intentional properties. The second is that psychological explanations should not appeal to properties that fail to supervene on an organism's physiology. (Variations on this theme can be found in Stich (1978, 1983) and in Fodor (1987, ch. 2).)

Reactions to this argument have taken a variety of forms. Perhaps the most radical is the proposal that cognitive psychology should recast its theories and explanations in a way that does not appeal to intentional properties of mental states but only to their formal or 'syntactic' properties (Stich, 1983). Somewhat less radical is the suggestion that we can define a species of representation or *CONTENT* – often called 'narrow content' (see EXTERNALISM/INTERNALISM) – which *does* supervene on an organism's physiology, and that psychological explanations that appeal to ordinary ('wide') intentional properties can be replaced by explanations that invoke only their narrow counterparts (Fodor, 1987). Both of these proposals accept the conclusion of the argument sketched in the previous paragraph, and they go on to propose ways in which cognitive psychology might be modified. But many philosophers have urged that the problem lies in the argument, not in the way that cognitive psychology goes about its business. The most common critique of the argument focuses on the normative premise – the one that insists that psychological explanations ought not to appeal to 'wide' properties that fail to supervene on physiology. Why shouldn't psychological explanations appeal to wide properties, the critics ask? What, exactly, is wrong with psychological explanations invoking properties that don't supervene on physiology? (See Burge, 1979, 1986.) Various answers have been proposed in the literature, though they typically end up invoking metaphysical principles that are less clear and less plausible than the normative thesis

they are supposed to support. (See, for example, Fodor, 1987, 1991.)

My own view is that the extensive literature in this area is mostly a tempest in a teapot, though I'm afraid I bear some of the responsibility for provoking it. I know of no clear or persuasive argument for excluding wide properties from psychological theories and explanations. But if you are inclined to demand that psychology invoke only properties that supervene on physiology, the demand is easy enough to satisfy. Given any psychological property that fails to supervene on physiology, it is trivial to characterize a narrow correlate property that does supervene. The extension of the correlate property includes all actual and possible objects in the extension of the original property, plus all actual and possible physiological duplicates of those objects. Theories originally stated in terms of wide psychological properties can be recast in terms of their narrow correlates with no obvious loss in their descriptive or explanatory power. It might be protested that when characterized in this way, narrow belief and narrow content are not really species of belief and content at all. But it is far from clear how this claim could be defended, or why we should care if it turns out to be right. (For more details see Stich (1991) and Stich and Laurence (forthcoming).)

The worry about the 'naturalizability' of intentional properties is much harder to pin down (see NATURALISM). According to Fodor, the worry derives from 'a certain ontological intuition: that there is no place for intentional categories in a physicalistic view of the world,' (1987, p. 97) and thus that 'the semantic (and/or the intentional) will prove permanently recalcitrant to integration in the natural order' (Fodor, 1984, p. 32). If intentional properties can't be integrated into the natural order, then presumably they ought to be banished from serious scientific theorizing. Psychology should have no truck with them. Indeed, if intentional properties have no place in the natural order, then nothing in the natural world has intentional properties, and intentional states do not exist at all. So goes the

worry. Unfortunately, neither Fodor nor anyone else has said anything very helpful about what is required to 'integrate' intentional properties into the natural order. There are, to be sure, various proposals to be found in the literature. But all of them seem to suffer from a fatal defect. On each account of what is required to naturalize a property or integrate it into the natural order, there are lots of perfectly respectable non-intentional scientific or commonsense properties that fail to meet the standard. Thus all the proposals that have been made so far end up throwing out the baby with the bath water. (For the details, see Stich and Laurence (forthcoming).)

Now, of course, the fact that no one has been able to give a plausible account of what is required to 'naturalize' the intentional may indicate nothing more than that the project is a difficult one. Perhaps with further work a more plausible account will be forthcoming. But one might also offer a very different diagnosis of the failure of all accounts of 'naturalizing' that have so far been offered. Perhaps the 'ontological intuition' that underlies the worry about integrating the intentional into the natural order is simply muddled. Perhaps there is no coherent criterion of naturalizability that all properties invoked in respectable science must meet. My own guess is that this diagnosis is the right one. Until those who are worried about the naturalizability of the intentional provide us with some plausible account of what is required of intentional categories if they are to find a place in 'a physicalistic view of the world' I think we are justified in refusing to take their worry seriously.

Recently, John SEARLE (1992) has offered a new set of philosophical arguments aimed at showing that certain theories in cognitive psychology are profoundly wrong-headed. The theories that are the targets of Searle's critique offer purely formal or computational explanations of various psychological capacities – like the capacity to recognize grammatical sentences, or the capacity to judge which of two objects in one's visual field is further away. Typically these the-

ories are set out in the form of a computer program – a set of rules for manipulating symbols – and the explanation offered for the exercise of the capacity in question is that people's brains are executing the program. The central claim in Searle's critique is that being a symbol or a computational state is not an 'intrinsic' physical feature of a computer state or a brain state. Rather, being a symbol is an 'observer relative' feature. But, Searle maintains, only intrinsic properties of a system can play a role in causal explanations of how they work. Thus appeal to symbolic or computational states of the brain could not possibly play a role in a 'causal account of cognition'.

There is something quite paradoxical about Searle's argument. To see the point, imagine that we find an unfamiliar object lying on the beach. After playing with it for a while, we discover that it has some remarkable capacities. If you ask it mathematical questions like, 'How much is 345 times 678?' or 'What is the square root of 1492?' the correct answer appears on a video display screen. How does the object do it? In order to find out, we turn it over to a group of scientists and engineers. After studying it for a while, they report that the object is a remarkable computer whose program includes a sophisticated algorithm for processing English along with a set of algorithms for various mathematical tasks. The report from the research team includes a detailed specification of the program they believe the object is using. Now most of us would be inclined to think that this report provides deep insight into how the object manages to produce answers to the questions we ask it. But if Searle is right, the program couldn't possibly explain how the object works. Obviously something has gone wrong somewhere. As I see it, the problem with Searle's argument lies in his assumption that cognitive theories of the sort he is criticizing aim at providing a 'causal account of cognition'. One of the important lessons of what I earlier called the 'descriptive' philosophy of science is that there are lots of different strategies of explanation in

science. One of the most useful explanatory strategies is functional decomposition, in which a complicated capacity is explained by showing how it can be accomplished by assembling a number of simpler capacities in an appropriate way. This strategy is widely used in biology. And, as Fodor (1968), Dennett (1978b), Cummins (1983) and others have argued, it is also central to the explanatory approach in cognitive psychology. Thus, even if we grant Searle's claim that being a computational state is not an intrinsic property of a state, and that only intrinsic properties can play a role in causal explanations, we will have no reason to conclude that the sort of computational explanations that Searle is criticizing are in any way problematic.

COGNITIVE PSYCHOLOGY SUGGESTS WAYS TO RESOLVE PHILOSOPHICAL PROBLEMS

The last section surveyed some of the philosophical arguments aimed at showing that cognitive psychology is confused and in need of reform. My reaction to those arguments was none too sympathetic. In each case, I maintained, it is the philosophical argument that is problematic, not the psychology it is criticizing. In this section I want to turn the tables and consider some of the proposals that have been made for using psychological findings to criticize philosophical theories and to resolve traditional philosophical problems. The tone in this section will be much more optimistic than in the previous section, since in this area I think there is some real progress to report.

Perhaps the most impressive example of the way in which psychological research can contribute to the resolution of philosophical disputes is to be found in the venerable debate between empiricist and rationalist accounts of knowledge. Though this debate is complex and multifaceted, one central issue has been the extent to which our knowledge in various domains is derived from experience, and the extent to which it is innate (*see INNATENESS*). Empiricists typically claim that most or all of our knowl-

edge is derived from experience, while rationalists maintain that important aspects of what we know are innate. Rationalists generally recognize that some input from experience may be needed to activate our innate knowledge and make it useable. Without appropriate environmental 'triggers' our innate knowledge may lie dormant. To make this point, Leibniz uses the analogy of a deeply grained block of marble. The block may have the shape of a man or a horse within it, though a fair amount of hammering and chiselling may be necessary to turn the block into a statue.

In the mid-1960s, Noam CHOMSKY began developing a set of arguments aimed at showing how considerations from linguistics and psycholinguistics might be used to resolve the dispute between the rationalists and the empiricists. (See, for example, Chomsky 1965, 1980.) The basic strategy in Chomsky's argument is obvious enough. What we should do, he urged, is look at the input to the process of language acquisition and the output of that process. If there is a significant amount of information in the output that cannot be found in the input, then the only plausible hypothesis is that the excess information is innate. What was novel and striking in Chomsky's argument was the empirical evidence he offered about the richness and complexity of the information that competent speakers of a language possess. Chomsky and his followers argued that there are lots of examples of grammatical rules that people acquire quite reliably, though the evidence available to them is not adequate to select the rule actually acquired over various alternatives that are not acquired. (For details, see Hornstein and Lightfoot, 1981.) In the years since Chomsky first advanced this argument, a number of other lines of evidence have been developed that underscore the extent to which human knowledge is strongly influenced by information-rich, domain-specific innate mental capacities. Some of the most impressive studies have demonstrated that very young children recognize phoneme boundaries and other subtle features that are essential for language mastery (Mehler

and Fox, 1984). Studies of visual perception in young children have also revealed a great deal of innate structure (Spelke, 1990). Though the details about what is innate and what is acquired in many domains remains to be determined, I think it is now quite clear that the sort of radical empiricist view often associated with Locke and with behaviourism is simply untenable.

Another area in which psychological studies have made important contributions to philosophy is in the branch of epistemology that attempts to characterize the notion of RATIONALITY and the related notion of justified inference. Here there are two quite different lines of influence to report. Some philosophers have offered accounts of rationality that build upon the actual inferences that people make and endorse. The best-known account of this sort is the one developed by Nelson Goodman (1965). According to Goodman, the justified inferences are the ones that would be sanctioned by a certain process. That process begins with the inferences we are actually inclined to accept, and it attempts to provide the simplest and most satisfying set of rules that will capture those inferences. As the process proceeds, certain inferences that we are initially inclined to accept may have to be thrown out, and certain rules that we initially find appealing may have to be amended or rejected entirely. Obviously, the sort of rules that will ultimately be sanctioned by this process will depend to a significant extent on the sorts of inferences that those using the process are initially inclined to make. And, while they rarely stress the point, it is clear that Goodman and others who are attracted to this account of justification typically assume that people's untutored inclinations are generally pretty good. But during the last twenty years or so, psychologists studying inference have accumulated a substantial body of data that casts serious doubt on this assumption. In many studies of both deductive and probabilistic reasoning it has been shown that normal subjects regularly draw inferences that would be classified as invalid (or worse) by the prevailing normative theory. (See, for

example, Nisbett and Ross, 1980; Kahneman et al., 1982.) Moreover, there is good reason to think that at least some of these patterns of inference are quite robust enough to survive the pruning process that Goodman describes. If this is right, it would not constitute a knock-down argument against Goodmanian accounts of justified inference. But the psychological findings do indicate that resolute Goodmanians are going to have some unpleasant bullets to bite. If they want to hang on to their account of justification, they are going to have to classify some pretty weird inferences as 'justified'. (For further details, see Stich, 1990, ch. 4; for a defence of the Goodmanian strategy, see Cohen, 1981.)

Quite a different approach to the assessment of reasoning is one that grows out of the pragmatist tradition. On this approach, reasoning is viewed as a tool for achieving various ends, and good strategies of reasoning are those that do a good job in enabling people to achieve their ends. Which strategies will facilitate which goals is an empirical question, not a matter to be determined by philosophical argument. So if we want to know what good reasoning in a given domain is like, the best way to find out is to locate people who have been particularly successful in that domain, and study the way in which they reason. In recent years, this strategy has been pursued in a particularly sophisticated way by Herbert Simon and his co-workers (Langley, Simon, Bradshaw and Zytkow, 1987.) Simon and his colleagues are interested in characterizing good scientific reasoning. They proceed by locating clear examples of people who have been successful in science (the people whose pictures appear in the science textbooks, as Simon sometimes puts it) and then trying to construct computer models that will simulate the reasoning of these successful scientists. There is, of course, no guarantee that this strategy will work, since it might turn out that the reasoning patterns of successful scientists have little or nothing in common – that successful scientific reasoning is not a 'natural kind' in psychology. However, Simon's work so far suggests that there are

important patterns to be discovered in the thinking of successful scientists. To the extent that Simon's project succeeds, it will constitute a particularly exciting sort of 'naturalized epistemology'.

Before closing I want to mention one other domain in which psychological research promises to have an important impact on philosophical theorizing. In ethics there are a number of views that presuppose substantive theses about human psychology. Often these psychological theses are taken to be part of received common sense about the mind, and thus no defence is offered. In psychology, however, received common sense has a distressing tendency to be mistaken. Consider, for example, those versions of utilitarian theory that rank actions on the basis of how well they do in satisfying people's preferences. These proposals make little sense unless we make the commonsensical assumption that people have determinate and reasonably stable preferences which may be elicited in a variety of ways. However, much recent work on the psychology of choice and preference suggests that this seemingly innocuous assumption may well be mistaken (Fischhoff et al., 1980; Slovic, 1990.) As Goldman has noted in a recent discussion of this literature, 'subtle aspects of how problems are posed, questions are phrased, and responses are elicited can have a substantial effect on people's expressed judgements and preferences. This leads some researchers to doubt whether, in general, there are stable and precise values or preferences antecedent to an elicitation procedure' (Goldman, 1993). If these doubts turn out to be justified, then a great deal of work in moral theory may well turn out to be indefensible, perhaps even incoherent.

See also ARTIFICIAL INTELLIGENCE; COMPUTATIONAL MODELS; DEVELOPMENTAL PSYCHOLOGY.

BIBLIOGRAPHY

Burge, T. 1979. Individualism and the mental. *Midwest Studies in Philosophy*, Vol. 4, 73–121.

- Burge, T. 1986. Individualism and psychology. *Philosophical Review*, 95, 3–46.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA.: MIT Press.
- Chomsky, N. 1980. *Rules and Representations*. New York: Columbia University Press.
- Cohen, J. 1981. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4.
- Cummins, R. 1983. *The Nature of Psychological Explanation*. Cambridge, MA.: MIT Press/Bardford Books.
- Cummins, R. 1989. *Meaning and Mental Representation*. Cambridge, MA.: MIT Press/Bardford Books.
- Dennett, D. 1978a. *Brainstorms*. Cambridge, MA.: MIT Press/Bardford Books.
- Dennett, D. 1978b. Artificial intelligence as philosophy and as psychology. In Dennett, 1978a.
- Dennett, D. 1987. *The Intentional Stance*. Cambridge, MA.: MIT Press/Bardford Books.
- Fischhoff, P., Slovic, P., and Lichtenstein, S. 1980. Knowing what you want: Measuring labile values. In *Cognitive Processes in Choice and Decision Behavior*, ed. T. Wallsten. Hillsdale, NJ.: Erlbaum.
- Fodor, J. 1968. The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*, 65.
- Fodor, J. 1975. *The Language of Thought*. New York: Thomas Y. Crowell.
- Fodor, J. 1984. Semantics, wisconsin style. *Synthèse*, 59. Reprinted in Fodor, 1990; page reference is to the latter.
- Fodor, J. 1987. *Psychosemantics*. Cambridge, MA.: MIT Press/Bardford Books.
- Fodor, J. 1990. *A Theory of Content and Other Essays*. Cambridge, MA.: MIT Press/Bardford Books.
- Fodor, J. 1991. A modal argument for narrow content. *Journal of Philosophy*, 88.
- Goldman, A. 1993. Ethics and cognitive science. In *Philosophical Applications of Cognitive Science*. Boulder, CO: Westview Press.
- Goodman, N. 1965. *Fact, Fiction and Forecast*. Indianapolis: Bobbs-Merrill.
- Haugeland, J. 1978. The Nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, 1.
- Hornstein, N., and Lightfoot, D. 1981. *Explanation in Linguistics*. London: Longman.
- Kahneman, D., Slovic, P., and Tversky, A., eds

1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press.
- Langley, P., Simon, H., Bradshaw, G., and Zytkow, J. 1987. *Scientific Discovery: Computational Explorations of the Creative Processes*. Cambridge, MA.: MIT Press.
- Mehler, J., and Fox, R., eds. 1984. *Neonate Cognition: Beyond the Blooming, Buzzing Confusion*. Hillsdale, N.J.: L. Erlbaum Associates.
- Nisbett, R., and Ross, L. 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, N.J.: Prentice-Hall.
- Putnam, H. 1975. The meaning of 'Meaning'. In *Language, Mind and Knowledge: Minnesota Studies in the Philosophy of Science*, Vol. 7, ed. K. Gunderson. Minneapolis: University of Minnesota Press.
- Searle, J. 1992. *The Rediscovery of the Mind*. Cambridge, MA.: MIT Press/Bradford Books.
- Sklar, L. 1974. *Space, Time and Spacetime*. Berkeley, CA: University of California Press.
- Slovic, P. 1990. Choice. In *Thinking*, ed. D. Osherson and E. Smith. Cambridge, MA.: MIT Press/Bradford Books.
- Sober, E. 1984. *The Nature of Selection*. Cambridge, MA.: MIT Press/Bradford Books.
- Spelke, E. 1990. Origins of visual knowledge. In *An Invitation to Cognitive Science: Visual Cognition and Action*, ed. Hollerbach, D. Osherson, S. Kosslyn, and J. Cambridge, MA.: MIT Press/Bradford Books.
- Stich, S. 1978. Autonomous psychology and the belief-desire thesis. *Monist*, 61.
- Stich, S. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA.: MIT Press/Bradford Books.
- Stich, S. 1990. *The Fragmentation of Reason*. Cambridge, MA.: MIT Press/Bradford Books.
- Stich, S. 1991. Narrow content meets fat syntax. In *Meaning in Mind: Fodor and His Critics*, ed. B. Loewer and G. Rey. Oxford: Basil Blackwell.
- Stich, S., and Laurence, S. Forthcoming. Intentionality and Naturalism.

STEPHEN STICH

Putnam, Hilary In 1960 I published a paper titled 'Minds and Machines' (1975, ch. 18), which suggested a possible new option in the philosophy of mind, and in 1967 I published two papers (1975, ch. 20

and 21) which became, for a time, the manifestos of the 'functionalist' current. FUNCTIONALISM (as many of my readers doubtless already know) holds that we are analogous to computers, and that our psychological states are simply our 'functional states', that is, they are the states that would figure in an ideal description of our 'program'. In the present 'self-portrait' of myself as a philosopher of mind I shall review the reasons that led me to propose functionalism and the reasons that subsequently led me to abandon it.

ROBOT CONSCIOUSNESS AND THE PROBLEM OF MAKING FUNCTIONALISM PRECISE

'Functionalism' views us as *automata*; that is as computers that happen to be made of flesh and blood. According to the functionalist view, a robot with the same program as a human being would *ipso facto* be conscious (*see CONSCIOUSNESS*). Although in a talk to the American Philosophical Association in 1964 (1975, ch. 19), I had drawn back from that view, arguing that the question whether any automaton was *conscious* was not really a question of fact but called for a 'decision' on our part, a decision 'to treat robots as fellow members of our linguistic community', when I came to write the two papers I described as 'functionalist manifestos'. I considered both the question as to whether psychological states are really 'functional' (i.e. computational) in nature and the question as to whether an automaton could be conscious to be factual questions. The earlier talk, I had come to see, contained an error.

In the 1964 paper, I assumed that if an 'IDENTITY THEORY' (a theory to the effect that psychological states are identical either with brain states or with functional states) were true, then it would have to be true as a consequence of (1) the meanings of psychological words, and (2) empirical facts that do not themselves beg the question as to whether a robot could be conscious. But the same line of reasoning, I saw in the 1967 papers, if applied to the question