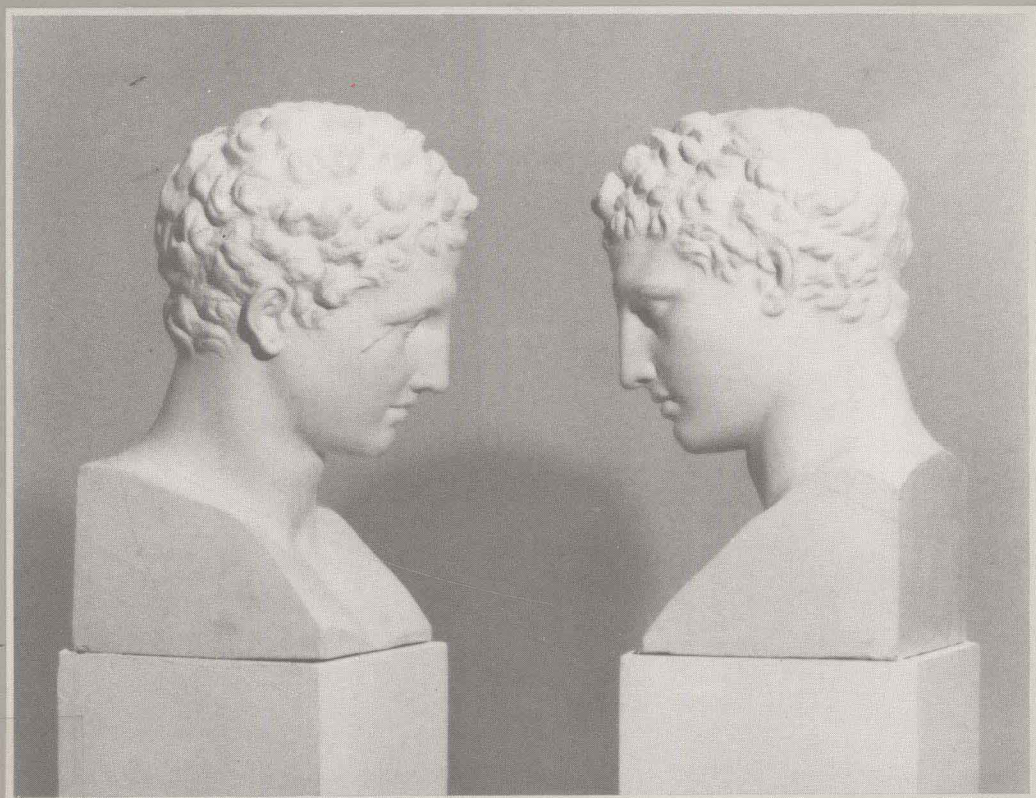


MIND AND COGNITION

A Reader



Edited by
WILLIAM G. LYCAN

MIND AND COGNITION

A Reader

Edited by
William G. Lycan



BLACKWELL
Oxford UK & Cambridge USA

Editorial matter and arrangement of contributions copyright © William G. Lycan, 1990; other matter as indicated in source footnote to each article; remaining matter copyright © Basil Blackwell Ltd., 1990

First published 1990

Reprinted (with corrections) 1990, 1991, 1992, 1994, 1995

Blackwell Publishers Ltd
108 Cowley Road, Oxford OX4 1JF, UK

Blackwell Publishers Inc.
238 Main Street
Cambridge, Massachusetts 02142, USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, re-sold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

British Library Cataloguing in Publication Data

A CIP catalogue record for this book is available from the British Library.

Library of Congress Cataloging in Publication Data

Mind and cognition.

Includes index.

1. Mind and body. 2. Thought and thinking.

3. Cognitive science. I. Lycan, William G.

BF171.M55 1989 128'.2 88-7915

ISBN 0-631-16763-3 (pbk.)

Typeset in 9½ on 11pt Ehrhardt

by Photo-graphics, Honiton, Devon

Printed and bound in Great Britain by Hartnolls Limited, Bodmin,
Cornwall

This book is printed on acid-free paper

Mind and Cognition



Preface

In the past thirty years, the philosophy of mind has seen a massive shift of doctrine, of method, and of perspective. Characteristic of this shift is the unprecedented attention of philosophers of mind to science: not only to psychology and linguistics, but to computer science, evolutionary biology and neuroanatomy as well. As a result, the mind-body problem is now better understood than at any previous point in human history – so I would contend, and the contention is borne out by the contents of this anthology. That is not to claim consensus for any one solution to the mind-body problem, for (of course) none exists. It is to claim a fairish consensus on questions of what the going arguments do and do not show, what the live options are, and what is at stake.

The essays and excerpts collected here are themselves predominantly philosophical. I would rather have assembled a more eclectic gathering, to include works written by empirical scientists with no speculative parsnips to butter, but the presentation of a significant number of such pieces as an integrated whole would have required a book-length introductory survey. The reader will have to rely on my bibliographies and (better) on my authors' footnotes.

Even regarding philosophy alone, my choice of headings for the various Parts, and of the readings themselves, reflects my no doubt tendentious view of the field and of what has happened in the philosophy of mind since the 1960s. Others may see things differently, and I am sure others would have included different items in the bibliographies.

This volume's closest and most distinguished predecessor is Ned Block (ed.), *Readings in Philosophy of Psychology*, vols One and Two (Cambridge: Harvard University Press, 1980). I thank Block for his unselfish encouragement of my own project, and I urge every reader who has already purchased a copy of this anthology to buy Block's as well.

My greatest debts are to Stephan Chambers, of Basil Blackwell Ltd, who suggested this anthology and has supported my work unstintingly, to Kim Sterelny for valuable discussions on choice of contents, and of course to the authors, especially to those who have contributed new or substantially revised essays.

I have written a brief synoptic introduction to each of the Parts. Citations in those introductions refer to items in the "Further reading" lists at the end of each introduction.

Contents

Preface	ix
I: Ontology from Behaviorism to Functionalism	1
Introduction	3
1 <i>Behaviorism</i>	14
An excerpt from "Talking and Thinking"	
J. B. WATSON	14
An excerpt from "Psychology in Physical Language"	
RUDOLF CARNAP	23
2 <i>The Identity Theory</i>	29
Is Consciousness a Brain Process?	
U. T. PLACE	29
3 <i>Early Causal and Functionalist Views</i>	37
The Causal Theory of the Mind	
D. M. ARMSTRONG	37
The Nature of Mental States	
HILARY PUTNAM	47
II: Homuncular Functionalism and Other Teleological Theories	57
Introduction	59
4 <i>Homuncular Functionalism</i>	63
Why the Law of Effect Will Not Go Away	
DANIEL C. DENNETT	63
The Continuity of Levels of Nature	
WILLIAM G. LYCAN	77

5	<i>Teleological Functionalism</i>	97
	Putting the Function Back into Functionalism	
	ELLIOTT SOBER	97
6	<i>Teleological Views of Intentionality</i>	107
	Functionalism, Information and Content	
	ROBERT VAN GULICK	107
	Misrepresentation	
	FRED DRETSKE	129
III: Instrumentalism		145
	Introduction	147
7	<i>An Instrumentalist Theory</i>	150
	True Believers: The Intentional Strategy and Why it Works	
	DANIEL C. DENNETT	150
	Dennett on Intentional Systems	
	STEPHEN P. STICH	167
	Making Sense of Ourselves	
	DANIEL C. DENNETT	184
IV: Eliminativism and Neurophilosophy		199
	Introduction	201
8	<i>Classical Eliminativism</i>	204
	Mental Events and the Brain	
	PAUL K. FEYERABEND	204
9	<i>Current Eliminativism</i>	206
	Eliminative Materialism and the Propositional Attitudes	
	PAUL M. CHURCHLAND	206
10	<i>Neurophilosophy and Connectionism</i>	224
	Neural Representation and Neural Computation	
	PATRICIA SMITH CHURCHLAND AND TERRENCE J. SEJNOWSKI	224
	Connectionism and the Philosophy of Mind: An Overview	
	WILLIAM BECHTEL	252
V: The "Language of Thought" Hypothesis		275
	Introduction	277
11	<i>Defending the "Language of Thought"</i>	282
	Why There Still Has to Be a Language of Thought	
	JERRY A. FODOR	282

12	<i>Attacking the "Language of Thought"</i>	300
	Stalking the Wild Epistemic Engine	
	PAUL M. CHURCHLAND AND PATRICIA SMITH CHURCHLAND	300
13	<i>Psychosemantics</i>	312
	Psychosemantics, or: Where Do Truth Conditions Come From?	
	JERRY A. FODOR	312
VI: The Status of "Folk Psychology"		339
	Introduction	341
14	<i>Attacking "Folk Psychology"</i>	345
	Autonomous Psychology and the Belief–Desire Thesis	
	STEPHEN P. STICH	345
	An excerpt from "The Syntactic Theory of Mind"	
	STEPHEN P. STICH	361
	A Narrow Representational Theory of the Mind	
	MICHAEL DEVITT	371
15	<i>Defending "Folk Psychology"</i>	399
	Folk Psychology is Here to Stay	
	TERENCE HORGAN AND JAMES WOODWARD	399
	Banish DisContent	
	JERRY A. FODOR	420
VII: Consciousness, "Qualia" and Subjectivity		439
	Introduction	441
16	<i>"Qualia"-Based Objections to Functionalism</i>	444
	An excerpt from "Troubles with Functionalism"	
	NED BLOCK	444
	Epiphenomenal Qualia	
	FRANK JACKSON	469
17	<i>Functionalist Responses</i>	478
	Could Love Be Like a Heatwave? Physicalism and the Subjective	
	Character of Experience	
	JANET LEVIN	478
	Physicalism and the Cognitive Role of Acquaintance	
	LAURENCE NEMIROW	490
	What Experience Teaches	
	DAVID LEWIS	499
	Quining Qualia	
	DANIEL C. DENNETT	519

VIII: Special Topics	549
Introduction	551
18 <i>Perception and Physical Theory</i>	555
Color and Illusion	
C. L. HARDIN	555
The Implications of Land's Theory of Color Vision	
KEITH CAMPBELL	567
19 <i>The Imagery Issue</i>	577
Mental Pictures and Cognitive Science	
NED BLOCK	577
The Imagery Debate	
KIM STERELNY	607
20 <i>Language and Innateness</i>	627
On the Nature, Use and Acquisition of Language	
NOAM CHOMSKY	627
Modeling Language Development	
DAVID LIGHTFOOT	646
21 <i>The Prospects for Artificial Intelligence</i>	660
Understanding Natural Language	
JOHN HAUGELAND	660
Index	671

Part I

Ontology from Behaviorism to 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.”

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.

Behaviorism

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. None the less, 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 VI 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. Artificial Intelligence (AI) is, roughly, the project of getting computing machines to perform tasks that would usually be