PHILOSOPHY OF MIND

Classical and Contemporary Readings

David J. Chalmers

New York Oxford
OXFORD UNIVERSITY PRESS
2002

Oxford University Press

Oxford New York
Auckland Bangkok Buenos Aires Cape Town Chennai
Dar es Salaam Delhi Hong Kong Istanbul Karachi Kolkata
Kuala Lumpur Madrid Melbourne Mexico City Mumbai Nairobi
São Paulo Shanghai Singapore Taipei Tokyo Toronto

and an associated company in Berlin

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York, 10016
http://www.oup-usa.org

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data
Chalmers, David John, 1966–
Philosophy of mind : classical and contemporary readings / David J. Chalmers.
p. cm.
Includes bibliographical references.
ISBN-13 978-0-19-514581-6
ISBN 0-19-514580-1 (hardback : alk. Paper)—ISBN 0-19-514581-X (pbk : alk.paper)
1. Philosophy of mind. I. Title.

BD418.3 .C435 2002
128′.2—dc21 2002072403

Printing number: 9 8 7 6 5

Printed in the United States of America
on acid-free papers

tity true whenever both its terms have the status of improper descriptions, false whenever one term has that status and the other does not. This might best be the theory of descriptions in Dana Scott, 'Existence and Description in Formal Logic,' in R. Schoenman, ed., *Bertrand Russell: Philosopher of the Century* (London: Allen & Unwin, 1967).

12. In general, or in the case of a given species, or in the case of a given person. It might turn out that the causal roles definitive of mental states are occupied by different neural (or other) states in different organisms. See my discussion of Hilary Putnam 'Psychological Predicates' in *Journal of Philosophy,* 66 (1969): 23–25.
13. It may be objected that the number of mental states is infinite, or at least enormous; for instance, there are as many states of belief as there are propositions to be believed. But it would be better to say that there is one state of belief, and it is a relational state, relating people to propositions. (Similarly, centigrade temperature is a relational state, relating objects to numbers.) The platitudes involving belief would, of course, contain universally quantified proposition-variables. Likewise for other mental states with intentional objects.
14. Wilfrid Sellars, 'Empiricism and the Philosophy of Mind,' in Feigl and Scriven, eds., *Minnesota Studies in the Philosophy of Science,* I (University of Minnesota Press, 1956): 309–20.
15. Two myths which cannot both be true together can nevertheless both be good together. Part of my myth says that names of color-sensations were T-terms, introduced using names of colors as O-terms. If this is a good myth, we should be able to define 'sensation of red' roughly as 'that state apt for being brought about by the presence of something red (before one's open eyes, in good light, etc.).' A second myth says that names of colors were T-terms introduced using names of color-sensations as O-terms. If this second myth is good, we should be able to define 'red' roughly as 'that property of things apt for bringing about the sensation of red.' The two myths could not both be true, for which came first: names of color-sensations or of colors? But they could both be good. We could have a circle in which colors are correctly defined in terms of sensations and sensations are correctly defined in terms of colors. We could not discover the meanings *both* of names of colors and of names of color-sensations just by looking at the circle of correct definitions, but so what?
16. See 'How to Define Theoretical Terms': 440–441.
17. By Armstrong, in *A Materialist Theory of the Mind,* pp. 100–13. He finds independent grounds for denying the infallibility of introspection.

Troubles with Functionalism

Ned Block

. . . One characterization of functionalism that is probably vague enough to be accepted by most functionalists is: each type of mental state is a state consisting of a disposition to act in certain ways *and to have certain mental states,* given certain sensory inputs and certain mental states. So put, functionalism can be seen as a new incarnation of behaviorism. Behaviorism identifies mental states with dispositions to act in certain ways in certain input situations. But as critics have pointed out (Chisholm, 1957; Putnam, 1963), desire for goal G cannot be identified with, say, the disposition to do A in input circumstances in which A leads to G, since, after all, the agent might not *know* A leads to G and thus might not be disposed to do A. Functionalism replaces behaviorism's "sensory inputs" with "sensory inputs and mental states"; and functionalism replaces behaviorism's "disposition to act" with "disposition to act and have certain mental states." Functionalists want to individuate mental states causally, and since mental states have mental causes and effects as well as sensory causes and behavioral effects, functionalists individuate mental states partly in terms of causal relations to other mental states. One consequence of this difference between functionalism and behaviorism is that there are organisms that according to behaviorism, have mental states but, according to functionalism, do not have mental states.

So, necessary conditions for mentality that are postulated by functionalism are in one respect stronger than those postulated by behav-

Excerpted from C. W. Savage, ed., *Perception and Cognition* (University of Minnesota Press, 1978), pp. 261–325, with permission of the publisher.

iorism. According to behaviorism, it is necessary and sufficient for desiring that G that a system be characterized by a certain set (perhaps infinite) of input-output relations; that is, according to behaviorism, a system desires that G just in case a certain set of conditionals of the form 'It will emit O given I' are true of it. According to functionalism, however, a system might have these input-output relations, yet not desire that G; for according to functionalism, whether a system desires that G depends on whether it has internal states which have certain causal relations to other internal states (and to inputs and outputs). Since behaviorism makes no such "internal state" requirement, there are possible systems of which behaviorism affirms and functionalism denies that they have mental states.[1] One way of stating this is that, according to functionalism, behaviorism is guilty of *liberalism*—ascribing mental properties to things that do not in fact have them. . . .

By 'physicalism,' I mean the doctrine that pain, for example, is identical to a physical (or physiological) state.[2] As many philosophers have argued (notably Fodor, 1965, and Putnam, 1966; see also Block & Fodor, 1972), if functionalism is true, physicalism is false. The point is at its clearest with regard to Turing-machine versions of functionalism. Any given abstract Turing machine can be realized by a wide variety of physical devices; indeed, it is plausible that, given any putative correspondence between a Turing-machine state and a configurational physical (or physiological) state, there will be a possible realization of the Turing machine that will provide a counterexample to that correspondence. (See Kalke, 1969; Gendron, 1971; Mucciolo, 1974, for unconvincing arguments to the contrary; see also Kim, 1972.) Therefore, if pain is a functional state, it cannot, for example, be a brain state, because creatures without brains can realize the same Turing machine as creatures with brains. . . .

One way of expressing this point is that, according to functionalism, physicalism is a *chauvinist* theory: it withholds mental properties from systems that in fact have them. In saying mental states are brain states, for example, physicalists unfairly exclude those poor brainless creatures who nonetheless have minds. . . .

This chapter has three parts. The first [excerpted here—ed.] argues that functionalism is guilty of liberalism, the second that one way of modifying functionalism to avoid liberalism is to tie it more closely to empirical psychology, and the third that no version of functionalism can avoid both liberalism and chauvinism.

1.1. More about What Functionalism Is

. . . One can also categorize functionalists in terms of whether they regard functional identities as part of a priori psychology or empirical psychology. (Since this distinction crosscuts the machine/nonmachine distinction, I shall be able to illustrate nonmachine versions of functionalism in what follows.) The a priori functionalists (e.g., Smart, Armstrong, Lewis, Shoemaker) are the heirs of the logical behaviorists. They tend to regard functional analyses as analyses of the meanings of mental terms, whereas the empirical functionalists (e.g., Fodor, Putnam, Harman) regard functional analyses as substantive scientific hypotheses. In what follows, I shall refer to the former view as 'Functionalism' and the latter as 'Psychofunctionalism.' (I shall use 'functionalism' with a lowercase 'f' as neutral between Functionalism and Psychofunctionalism. When distinguishing between Functionalism and Psychofunctionalism, I shall always use capitals.)

Functionalism and Psychofunctionalism and the difference between them can be made clearer in terms of the notion of the Ramsey sentence of a psychological theory. Mental-state terms that appear in a psychological theory can be defined in various ways by means of the Ramsey sentence of the theory. All functional-state identity theories (and functional-property identity theories) can be understood as defining a set of functional states (or functional properties) by means of the Ramsey sentence of a psychological theory—with one functional state corresponding to each mental state (or one functional property corresponding to each mental property). The functional state corresponding to pain will be called the 'Ramsey functional correlate' of pain, with respect to the psychological theory. In terms of the notion of a Ramsey functional correlate with respect to a theory, the distinction between Functionalism and Psychofunctionalism can be defined as follows: Functionalism identifies mental state S with S's Ramsey functional correlate with respect to a *common-sense* psychological theory; Psychofunctionalism identifies S with S's Ramsey functional correlate with respect to a *scientific* psychological theory. . . .

1.2. Homunculi-Headed Robots

In this section I shall describe a class of devices that embarrass all versions of functionalism in that they indicate functionalism is guilty of liberalism—classifying systems that lack mentality as having mentality.

Consider the simple version of machine functionalism already described. It says that each system having mental states is described by at least one Turing-machine table of a certain kind, and each mental state of the system is identical to one of the machine-table states specified by the machine table. I shall consider inputs and outputs to be specified by descriptions of neural impulses in sense organs and motor-output neurons. This assumption should not be regarded as restricting what will be said to Psychofunctionalism rather than Functionalism. As already mentioned, every version of functionalism assumes *some* specification of inputs and outputs. A Functionalist specification would do as well for the purposes of what follows.

Imagine a body externally like a human body, say yours, but internally quite different. The neurons from sensory organs are connected to a bank of lights in a hollow cavity in the head. A set of buttons connects to the motor-output neurons. Inside the cavity resides a group of little men. Each has a very simple task: to implement a "square" of a reasonably adequate machine table that describes you. On one wall is a bulletin board on which is posted a state card, i.e., a card that bears a symbol designating one of the states specified in the machine table. Here is what the little men do: Suppose the posted card has a 'G' on it. This alerts the little men who implement G squares—'G-men' they call themselves. Suppose the light representing input I_{17} goes on. One of the G-men has the following as his sole task: when the card reads 'G' and the I_{17} light goes on, he presses output button O_{191} and changes the state card to 'M'. This G-man is called upon to exercise his task only rarely. In spite of the low level of intelligence required of each little man, the system as a whole manages to simulate you because the functional organization they have been trained to realize is yours. A Turing machine can be represented as a finite set of quadruples (or quintuples, if the output is divided into two parts)—current state, current input; next state, next output. Each little man has the task corresponding to a single quadruple. Through the efforts of the little men, the system realizes the same (reasonably adequate) machine table as you do and is thus functionally equivalent to you.

I shall describe a version of the homunculi-headed simulation, which is more clearly nomologically possible. How many homunculi are required? Perhaps a billion are enough; after all, there are only about a billion neurons in the brain.

Suppose we convert the government of China to functionalism, and we convince its officials that it would enormously enhance their international prestige to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to the artificial body mentioned in the previous example. We replace the little men with a radio transmitter and receiver connected to the input and output neurons. Instead of a bulletin board, we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China. Surely such a system is not physically impossible. It could be functionally equivalent to you for a short time, say an hour.

"But," you may object, "how could something be functionally equivalent to me for *an hour?* Doesn't my functional organization determine, say, how I would react to doing nothing for a week but reading *Reader's Digest?*" Remember that a machine table specifies a set of conditionals of the form: if the machine is in S_i and receives input I_j, it emits output O_k and goes into S_l. Any system that has a set of inputs, outputs, and states related in the way described realizes that machine table, even if it exists for only an instant. For the hour the Chinese system is "on," it *does* have a set of inputs, outputs, and states of which such conditionals are true. Whatever the initial state, the system will respond in whatever way the machine table directs. This is how *any* computer realizes the machine table it realizes.

Of course, there are signals the system would respond to that you would not respond to, e.g., massive radio interference or a flood of the Yangtze River. Such events might cause a malfunction, scotching the simulation, just as a bomb in a computer can make it fail to realize the machine table it was built to realize. But just

as the computer *without* the bomb *can* realize the machine table, the system consisting of the people and artificial body can realize the machine table so long as there are no catastrophic interferences, e.g., floods, etc.

"But," someone may object, "there is a difference between a bomb in a computer and a bomb in the Chinese system, for in the case of the latter (unlike the former), inputs as specified in the machine table can be the cause of the malfunction. Unusual neural activity in the sense organs of residents of Chungking Province caused by a bomb or by a flood of the Yangtze can cause the system to go haywire."

Reply: the person who says what system he or she is talking about gets to say what counts as inputs and outputs. I count as inputs and outputs only neural activity in the artificial body connected by radio to the people of China. Neural signals in the people of Chungking count no more as inputs to this system than input tape jammed by a saboteur between the relay contacts in the innards of a computer count as an input to the computer.

Of course, the object consisting of the people of China + the artificial body has *other* Turing machine descriptions under which neural signals in the inhabitants of Chungking *would* count as inputs. Such a new system (i.e., the object under such a new Turing-machine description) would not be functionally equivalent to you. Likewise, any commercial computer can be redescribed in a way that allows tape jammed into its innards to count as inputs. In describing an object as a Turing machine, one draws a line between the inside and the outside. (If we count only neural impulses as inputs and outputs, we draw that line inside the body if *we* count only peripheral stimulations as inputs and only bodily movements as outputs, we draw that line at the skin.) In describing the Chinese system as a Turing machine, I have drawn the line in such a way that it satisfies a certain type of functional description—one that you *also* satisfy, and one that, according to functionalism, justifies attributions of mentality. Functionalism does not claim that every mental system has a machine table of a sort that justifies attributions of mentality with respect to *every* specification of inputs and outputs, but rather, only with respect to *some* specification.

Objection: The Chinese system would work too slowly. The kind of events and processes with which we normally have contact would pass by far too quickly for the system to detect them. Thus, we would be unable to converse with it, play bridge with it, etc.[3]

Reply: It is hard to see why the system's time scale should matter. What reason is there to believe that *your* mental operations could not be very much slowed down, yet remain mental operations? Is it really contradictory or nonsensical to suppose we could meet a race of intelligent beings with whom we could communicate only by devices such as time-lapse photography? When we observe these creatures, they seem almost inanimate. But when we view the time-lapse movies, we see them conversing with one another. Indeed, we find they are saying that the only way they can make any sense of us is by viewing movies greatly slowed down. To take time scale as all important seems crudely behavioristic. Further, even if the time-scale objection is right, I can elude it by retreating to the point that a homunculus-head that works in normal time is *metaphysically* possible, even if not nomologically possible. Metaphysical possibility is all my argument requires (see Section 1.3).[4]

What makes the homunculi-headed system (count the two systems as variants of a single system) just described a prima facie counter example to (machine) functionalism is that there is prima facie doubt whether it has any mental states at all—especially whether it has what philosophers have variously called "qualitative states," "raw feels," or "immediate phenomenological qualities." (You ask: What is it that philosophers have called qualitative states? I answer, only half in jest. As Louis Armstrong said when asked what jazz is, "If you got to ask, you ain't never gonna get to know.") In Nagel's terms (1974), there is a prima facie doubt whether there is anything which it is like to be the homunculi-headed system.

The force of the prima facie counterexample can be made clearer as follows: Machine functionalism says that each mental state is identical to a machine-table state. For example, a particular qualitative state, Q, is identical to a machine-table state, S_q. But if there is nothing it is like to be the homunculi-headed system, it cannot be in Q even when it is in S_q. Thus, if there is prima facie doubt about the homunculi-headed system's mentality, there is prima facie doubt that $Q = S_q$, i.e., doubt that the kind of functionalism under consideration is true.[5] Call this argument the Absent Qualia Argument.

NOTES

1. The converse is also true.
2. State type, not state token. Throughout the chapter, I shall mean by 'physicalism' the doctrine that says each distinct type of mental state is identical to a distinct type of physical state, for example, pain (the universal) is a physical state. Token physicalism, on the other hand, is the (weaker) doctrine that each particular datable pain is a state of some physical type or other. Functionalism shows that type physicalism is false, but it does not show that token physicalism is false.

   By 'physicalism,' I mean *first order* physicalism, the doctrine that, e.g., the property of being in pain is a first-order (in the Russell-Whitehead sense) physical property. (A first-order property is one whose definition does not require quantification over properties; a second-order property is one whose definition requires quantification over first-order properties.) The claim that being in pain is a second-order physical property is actually a (physicalist) form of functionalism. See Putnam, 1970.

   'Physical property' could be defined for the purposes of this chapter as a property expressed by a predicate of some true physical theory or, more broadly, by a predicate of some true theory of physiology, biology, chemistry, or physics. Of course, such a definition is unsatisfactory without characterizations of these branches of science. See Hempel, 1970, for further discussion of this prohlem.
3. This point has been raised with me by persons too numerous to mention.
4. One potential difficulty for Functionalism is provided by the possibility that one person may have two radically different Functional descriptions of the sort that justify attribution of mentality. In such a case, Functionalists might have to ascribe two radically different systems of belief, desire, etc., to the same person, or suppose that there is no fact of the matter about what the person's propositional attitudes are. Undoubtedly, Functionalists differ greatly on what they make of this possibility, and the differences reflect positions on such issues as indeterminacy of translation.
5. Shoemaker, 1975, argues (in reply to Block & Fodor, 1972) that absent qualia are logically impossible, that is, that it is logically impossible that two systems be in the same functional state yet one's state have and the other's state lack qualitative content. If Shoemaker is right, it is wrong to doubt whether the homunculi-headed system has qualia. I attempt to show Shoemaker's argument to he fallacious in Block, 1980.

REFERENCES

Block, N. Are absent qualia impossible? *Philosophical Review,* 1980, 89, 257–74.

Block, N. & Fodor, J. What psychological states are not. *Philosophical Review,* 1972, 81, 159–81.

Chisholm, Roderick. *Perceiving.* Ithaca: Cornell University Press, 1957.

Fodor, J. Explanations in psychology. In M. Black (Ed.), *Philosophy in America,* London: Routledge & Kegan Paul, 1965.

Gendron, B. On the relation of neurological and psychological theories: A critique of the hardware thesis. In R. C. Buck and R. S. Cohen (Eds.), *Boston studies in the philosophy of Science VIII.* Dordrecht: Reidel, 1971.

Hempel, C. Reduction: Ontological and linguistic facets. In S. Morgenbesser, P. Suppes & M. White (Eds.), *Essays in honor of Ernest Nagel.* New York: St. Martin's Press, 1970.

Kalke, W. What is wrong with Fodor and Putnam's functionalism? *Nous,* 1969, 3, 83–93.

Kim, J. Phenomenal properties, psychophysical laws, and the identity theory. *The Monist,* 1972, 56(2), 177–92.

Mucciolo, L. F. The identity thesis and neuropsychology. *Nous,* 1974, 8, 327–42.

Nagel, T. What is it like to be a bat? *Philosophical Review,* 1974, 83, 435–50.

Putnam, H. Brains and behavior. 1963. Reprinted as are all Putnam's articles referred to here (except "On properties") in *Mind, language and reality; philosophical papers,* Vol. 2). London: Cambridge University Press, 1975.

———. The mental life of some machines. 1966.

———. On properties. In *Mathematics, matter and method; philosophical papers,* Vol. 1. London: Cambridge University Press, 1970.

Shoemaker, S. Functionalism and qualia. *Philosophical studies,* 1975, 27, 271–315.