Philosophy of Mind and Metaphysics

Lecture XI: Mind as a Computer: Machine Functionalism

Tim Black
California State University, Northridge
Spring 2004

I. Multiple Realizability and Functionalism

a. Mental states (and events and properties) are ***multiply realizable***. That is, one and the same mental state can be realized in different ways in different physical systems. "Mental property *M*—say, being in pain—may be such that in humans C-fiber activation realizes it but in other species (think of octopuses and reptiles) the physiological mechanism that realizes pain may well be vastly different" (p. 74). Moreover, the multiple realizability of mental states creates a problem for the Psychoneural (Type) Identity Theory. For, if one and the same mental state can be realized in different ways in different creatures, then there is no *one* type of physical state with which any particular type of mental state is identical. Pain, for example, is not identical to C-fiber activation since pain is realized in other creatures by some *other* type of physical state.

b. The multiple realizability of mental states pushes us away from the Psychoneural (Type) Identity Theory, and toward ***Functionalism***. "According to functionalism, a mental kind is a *functional kind*, or a *causal-functional kind*, since the "function" involved is to fill a certain causal role. ... Thus, the concept of pain is defined in terms of its function, and the function involved is to serve as a causal intermediary between typical pain inputs (tissue damage, trauma, etc.) and typical pain outputs (winces, groans, escape behavior, etc.). The functionalist will say that this is generally true of all mental kinds: Mental kinds are causal-functional kinds, and what all instances of a given mental kind have in common is the fact that they serve a certain causal role distinctive of that kind. ... the concept of a mental state is that of an internal state apt to be caused by certain sensory inputs and apt for causing certain behavioral outputs" (pp. 76-7).

II. Functionalism and Behaviorism

a. Similarities: "Both functionalism and behaviorism speak of sensory input and behavioral output—or "stimulus" and "response"—as central to the concept of mentality" (p. 77).

b. Differences: Behaviorism says that behavior is constitutive of mentality, but functionalism says that mentality is whatever internal state is the causal intermediary between (sensory) input and (behavioral) output. According to functionalism, "what makes a mental event the kind of

mental event it is is the way it is causally linked to other mental-event kinds and input/output conditions" (p. 79).

III. TURING MACHINES

a. A Turing machine is a computing machine that is designed to perform certain tasks—for example, adding positive integers—and that can be defined by its "machine table," which is a complete set of instructions for performing the assigned tasks.

b. "Suppose that we give the machine table for our simple adding machine, TM_1, to an engineering class as an assignment: Each student is to build an actual physical device that will do the computations as specified by its machine table. What we are asking the students to build, therefore, are "physical realizations" of TM_1, actual adding machines that will operate in accordance with the machine table of TM_1. We can safely predict that a huge, heterogeneous variety of machines will be turned in. Some of them may really look and work like the Turing machine as described: They may have a paper tape neatly divided into squares, with an actual physical "head" that can scan, erase, and print symbols. Some will perhaps use magnetic tapes and heads that read, write, and erase electrically. Some machines will have no "tapes" or "heads" but instead use spaces on a computer disk or memory locations in its CPU to do the computation. A clever student with a sense of humor (and lots of time and other resources) might try to build a hydraulically operated device with pipes and valves instead of wires and switches. The possibilities are endless" (p. 85). This means that Turing machines are *multiply physically realizable*.

IV. MACHINE FUNCTIONALISM

a. According to machine functionalists, we can think of the mind as a Turing machine. "The central idea is that what it is for something to have mentality—that is, to have a psychology—is for it to be a physically realized Turing machine of appropriate complexity, with its mental states (that is, mental-state types) identified with the internal states of the machine table. Another way of explaining this idea is to use the notion of machine description: An organism has mentality just in case there is a Turing machine of appropriate complexity that is a machine description of it, and its mental-state kinds are to be identified with the internal states of that Turing machine" (p. 88).

b. PROBLEMS

i. INPUT/OUTPUT INAPPROPRIATENESS: Suppose that two systems realize the same Turing machine. One of those systems is a human being, Larry, while the other is a computer, MAX. "So the inputs and outputs specified [for MAX] are the usual inputs and outputs

appropriate for a computing machine, perhaps strings of symbols entered on the keyboard and strings of symbols on the computer monitor or its printout" (p. 93). Here's the problem, though: "While [MAX's] input/output specification is isomorphic to Larry's, it seems entirely inappropriate for psychology. Although it may not be easy to characterize the differences precisely, we would not consider inputs and outputs consisting merely of strings of symbols as fitting for something with true mentality" (p. 93). It seems, then, that we need not think of the mind as a Turing machine. For (a) two things can realize the same Turing machine, (b) one of those two things can obviously be something that enjoys mentality, while (c) the other has an input/output specification that seems entirely inappropriate for mentality.

ii. COMMITMENT TO ISOMORPHIC TOTAL PSYCHOLOGIES: According to machine functionalism, to be in the same psychological state is to be in the same internal mental state, and two things can be in the same internal mental state only if they realize the same Turing machine, that is, only if they have identical total psychologies. It follows that two things can be in the same psychological state only if they have identical total psychologies (see pp. 92-3, 94-6). Yet this seems mistaken. Consider a counterexample: "For an octopus and a human to be in the same pain state, they must share an isomorphic psychology—an unlikely possibility, to say the least! And for two humans to share a single mental state, they must have an exactly identical total psychology ... This consequence of machine functionalism clearly goes against our expectations" (p. 95).

V. THE TURING TEST

a. The Turing test is designed to determine whether a system is *complex enough* to count as having a mind.[1] Here's the test: "By asking questions (or just holding a conversation) via keyboard terminals, can we find out whether we are talking to a human or a computing machine? If there is a computer that can consistently fool us so that our success in guessing its identity is no better than what could be achieved by random guesses, we must concede, it seems, that this machine has the kind of mentality that we grant to humans" (p. 97).

b. PROBLEMS

i. INTERNAL PROCESSING MAKES NO DIFFERENCE: Imagine two machines, each of which performs basic arithmetic operations

[1] Things that fail the test might nevertheless have a mind, but the things that pass the test certainly will have a mind, or so the story goes.

involving integers up to 100, but only one of which "calculates ("figures out") the answer by applying the usual algorithms we use for these operations ... the other [machine] has a file in which answers are stored for all possible problems of addition, multiplication, subtraction, and division for integers up to 100, and its computation consists in "looking up" the answer for any problem given to it" (p. 99). In this case, the two machines are input/output equivalent. Only one of them, though, "thinks things through." Thus, it seems that the two machines do *not* have the same psychological status. This suggests that even if some machine is the input/output equivalent of a human being, that machine need not have the same psychological status as a human being. If this is the case, then the Turing test does not do what it purports to do.

ii. THE "CHINESE ROOM" ARGUMENT: (1) Imagine that someone, S, who understands no Chinese is confined in a room ("the Chinese room") with a set of rules for systematically trading certain strings of symbols for certain other strings of symbols. These symbol strings are in fact Chinese expressions, and the transformation rules are purely *formal*, in the sense that their application depends solely on the shapes of the symbols involved. Once S becomes familiar enough with the rules, then from the perspective of someone outside the room, the input/output relationships are exactly the same as they would be if someone with a genuine understanding of Chinese were locked inside the room. (2) Moreover, what goes on inside as a computer is exactly like what goes on inside the Chinese room: rule-governed manipulation of symbols based on their shapes. (3) Therefore, mentality is more than rule-governed syntactic manipulation of symbols, and the Turing test, therefore, is invalid a test for mentality. (4) Searle goes on to suggest that "no amount of syntactic symbol-pushing will generate meaning" (p. 100), and mentality is not possible unless meanings are generated. It seems to follow that machine functionalism is false: If computers, no matter how complex, are only symbol-pushing engines, and if mentality requires something more than mere symbol-pushing, then the mind cannot be a computer, no matter how complex.