Handout 8: Functionalism

The Basic Idea behind various forms of *behaviorism* (see handout 6) is that to have a mind (or to be in such-and-such mental state) is to (be disposed to) behave in certain ways: how a creature's *inner organization* produces the behavior is irrelevant to whether it has a mind (or is in such-and-such mental state).

For many reasons -- some of which we have touched on in class -- behaviorism is not a popular view these days. Instead, many philosophers think some kind of *functionalism* is true. (See Chalmers, pp. 5-6, and Block, Functionalism.) The Basic Idea behind functionalism may be illustrated by the following three examples.

Example 1

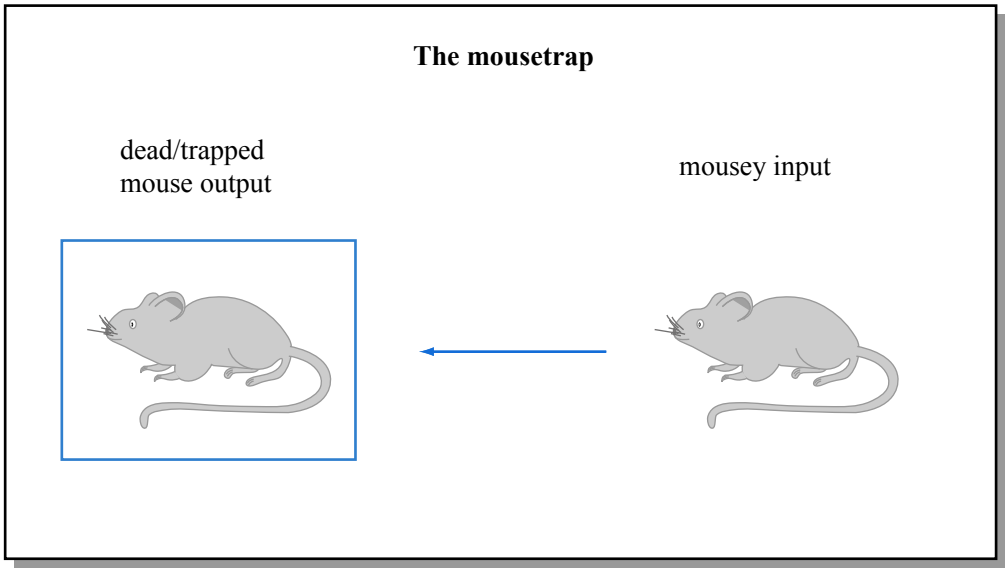

Image by MIT OpenCourseWare.

A mousetrap is anything that would respond to mousey input with trapped-mouse output. Provided that this is what something would do, it is a mousetrap no matter what it's made of: a mousetrap can be *multiply realized*. A mousetrap is the simplest sort of *functional kind*: the only constraints placed on the internal organization of a mousetrap is that it produce the right input-output behavior. So we might say that mousetrap is also a *behavioral* kind.

Cite as: Alex Byrne, course materials for 24.09 Minds and Machines, Fall 2011.
MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Example 2

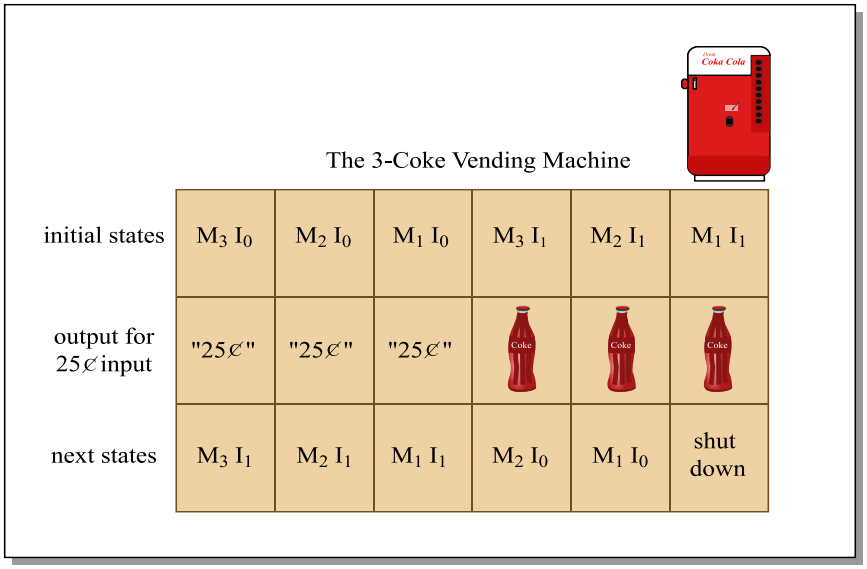

Image by MIT OpenCourseWare.

At any one time the vending machine is in two states (the six combinations of the three states M_3, M_2, M_1, with the two states I_1, I_0), or else it's shut down. Think of M_n as the "n-drinks available state" and I_n as the "n-quarters-received state". The causal relations between the inputs (25¢), the output (the dispensing of a Coke, or the illumination of the symbols "25¢"), and the various internal states are displayed in the above table:

So, if the machine is in state M_2 and state I_0 and receives input of 25¢, it illuminates the symbols "25¢", and goes into state M_2 and state I_1. If the machine is in state M_2 and state I_1 and receives input of 25¢, it dispenses a Coke, and goes into state M_1 and state I_1. And so on.

A vending machine is also a functional kind, and the table gives a functional theory of a particular machine machine. Like the mousetrap, provided something has internal states that are causally related to each other and to the inputs and outputs as specified by the table (provided, that is, that something has inner states with the appropriate *causal roles*), it is an instance of this particular vending machine, no matter what it's made of. Notice that to be this kind of vending machine a thing's internal organization must have a particular causal structure. So this vending machine is not a behavioral kind. Behaviorism is a limiting case of functionalism (e.g. things like mousetraps). But what *are* the states M_3, I_0, etc.? The table tells you everything you need to know. More explicitly:

A system S is in M_3 iff S is in the third of five states X_1, X_2, X_3, X_4, X_5, that are related to one another and to the possible inputs and outputs of S as follows:

> being in X_1 and X_4 and receiving 25¢ causes S to go into X_2 and remain in X_4; being in X_2 and X_4 and receiving 25¢ causes S to go into X_3 and remain in X_4;... being in X_1 and X_5 and receiving 25¢ causes S to dispense a Coke and to shut down.

The Basic Idea of functionalism is that mental states are functional states like M_3 and I_0. They are inner states that can be specified by their causal relations to inputs, outputs, and to each other. (See Putnam, "The nature of mental states".)

Cite as: Alex Byrne, course materials for 24.09 Minds and Machines, Fall 2011.
MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Example 3

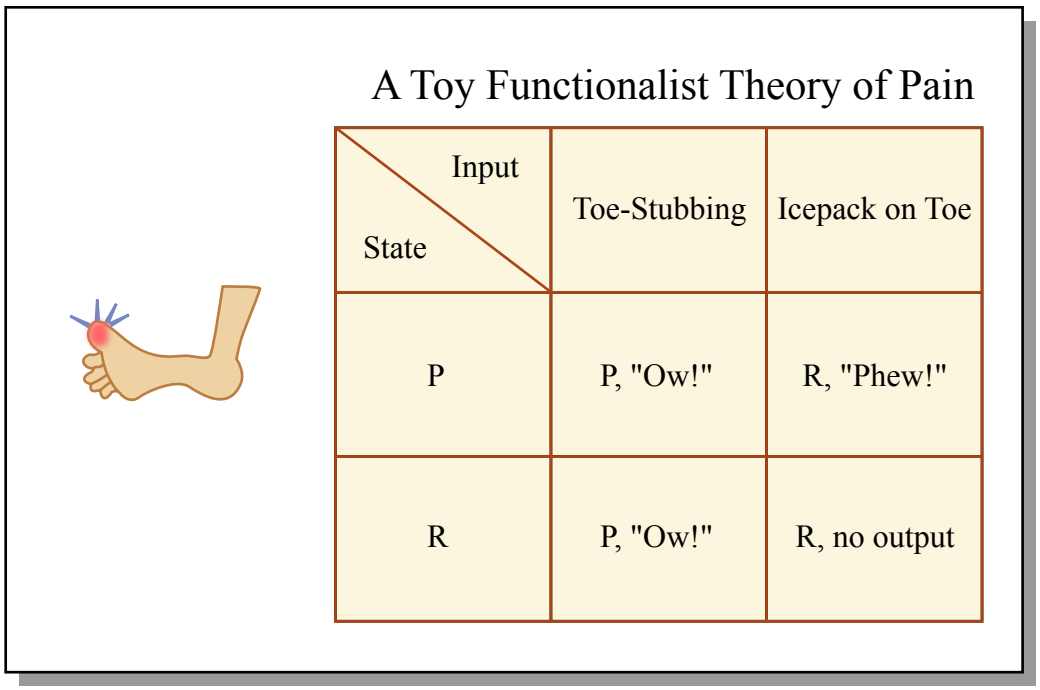

Image by MIT OpenCourseWare.

This is a (simplistic!) functionalist theory of pain (in the toe). The mental states of Pain and Relief are those functional states P and R specified by the above table. To be in pain, on this theory, is to have the above causal organization and to be in state P.

More explicitly:

A system S is in Pain iff S is in the first of two (exclusive) states X and Y that are related to one another and to the possible inputs and outputs of S as follows:

> being in X and stubbing its toe causes S to remain in X and say "Ow!"; being in X and having an icepack on its toe causes S to go into Y and say "Phew!"; being in Y and stubbing its toe causes S to go into X and say "Ow!"; being in Y and having an icepack on its toe causes S to remain in Y.

Functionalism and Physicalism

Is functionalism a version of physicalism or materialism? Unfortunately the terminology here can be very confusing: see Block, Functionalism, and contrast Block, "Troubles with functionalism", and Putnam, "The nature of mental states". Although functionalists are happy to concede that minds could be made from "non-physical" components, and in that sense are not physicalists, they invariably think that minds in the actual world are "realized" entirely physically, and in that sense are physicalists. On the standard way of defining 'physicalism' these days, functionalists are physicalists.

Representationalism and Computationalism

It is important to distinguish the following two theses that are sometimes conflated with functionalism:

Representationalism (a.k.a. the representational theory of the mind)

Cite as: Alex Byrne, course materials for 24.09 Minds and Machines, Fall 2011.
MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

Mental states require inner representations: a subject is in such-and-such mental state with representational content that *p* only if a proper part of the subject has the representational content that *p*. For example, if the subject believes that snow is white, then according to representationalism some proper part of the subject (part of her brain, presumably) has the representational content *that snow is white*. (Denied, for instance, by Dennett: see "True believers") [1]

Computationalism (a.k.a. the computer model of the mind)

A combination of (a) functionalism[2], (b) representationalism, and (c) the idea that mental processes are *computational*: they involve the algorithmic manipulation of symbols (i.e. the inner representations). Note that computationalism is not committed to the view that a mind is the realization of some Turing machine.[3] Although any algorithmic manipulation of symbols can be carried out by some Turing machine, it's not true that any *functional state* can be realized in some Turing machine. In other words, if being in a mental state involves having an inner causal organization of some sort, there is no guarantee that a realization of some Turing machine will have the required causal organization. And, on any plausible account of the causal organization required, no Turing machine will do the job. What might be required instead is a machine with a massively parallel architecture.

These two theses, and functionalism, come in various degrees of strength: e.g. one might be a functionalist about all mental states, or only about mental states like belief and desire, leaving pain for another kind of theory.

So, the relations between representationalism, computationalism, and functionalism, are as follows:

Computationalism implies functionalism and representationalism. Neither of the converse implications holds. And neither functionalism nor representationism implies the other.

"Troubles with Functionalism"

Block's homunculi-head example is supposed to show that functionalism is too "liberal" -- that it counts some things as having minds that do not in fact have them.

> 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 an 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... 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'... 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... (p. 96)

Cite as: Alex Byrne, course materials for 24.09 Minds and Machines, Fall 2011.
MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

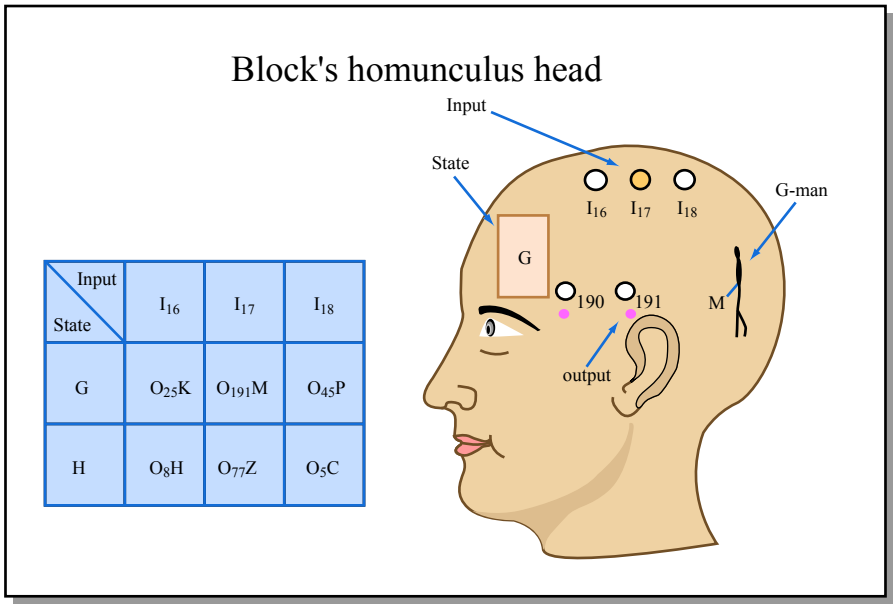

Image by MIT OpenCourseWare.

In a variant, the homunculi are replaced by the citizens of China (pp. 96-97). Block's objection is simple: in these examples, the system implemented by homunculi or by the nation of China has exactly the same functional organization as you. So, if functionalism is correct, these systems should share your mental life. But intuitively, the system doesn't have any mental states. It seems especially implausible, Block thinks, that the system could be in any "qualitative" mental states, like being in pain.

---

[1] A restriction here to "explicit" beliefs (and other intentional states) is required (see Block, The mind as ..., sect. 3.1)

[2] To avoid undue complexity, we are here blurring a distinction that Block makes in section 3 of "The mind as...". The distinction is between functionalism as a *partial* account of mental states (*excluding* their representational content), and functionalism as a *total* account (*including* representational content). On the second way of using 'functionalism' (this is the way Block uses it in "The mind as..."), it's possible to be a computationalist without being a functionalist.

[3] Be sure not to confuse Turing machines in the sense of *mathematical objects* about which various things are *proved*, with *concrete realizations* of Turing machines. But what is the relation between the abstract machine and its concrete realizations? In class, we have been assuming an intuitively correct answer to that question. See Block, "The mind as...", sect. 2.2, and Block, Functionalism.

Cite as: Alex Byrne, course materials for 24.09 Minds and Machines, Fall 2011.
MIT OpenCourseWare (http://ocw.mit.edu/), Massachusetts Institute of Technology. Downloaded on [DD Month YYYY].

MIT OpenCourseWare
http://ocw.mit.edu

24.09 Minds and Machines

Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.