Classic Text 19 - Philosophy of Mind: Functionalism

Functionalism in contemporary philosophy is a theory of mind that identifies mental states in terms of their functional role in causal relation to other mental states and their numerous sensory inputs and behavioural outputs. It avoids many of the difficulties beset by Cartesian Dualism on the one hand and Skinnerian behaviourism and type identity theory on the other hand, and as such "provides the conceptual underpinnings of much work in cognitive science." (Block, 1996)

The classic text for this study unit is "The Mind-Body Problem" by the American philosopher Jerry A. Fodor (1981) intended for a scientifically informed audience of non-philosophers. The article can be downloaded for free here. (South African copyright law allows for the reproduction of individual journal articles for the purposes of education.)

Fodor begins by describing an idealised version of the traditional role of philosophers of science before the advent of contemporary cognitive science in the 1970's and '80s. Before then, theoretical approaches to the mind-body problem fell into two exclusive categories: dualism and materialism. According to dualism, the mind is a non-physical entity or substance entirely distinct from that of the physical body. According to materialism in general, the mental is either not distinct from the physical or can be *identified* with certain physical states (within the brain or within the brain and the extended nervous system.) Several of these approaches have been discussed in Classic Texts 02, 06, 11 and 16.

Although mind-body dualism is the default belief for most people who have not given the matter much systematic thought, its difficulties are seemingly insurmountable. According to Fodor,

> The chief drawback of dualism is its failure to account adequately for mental causation. If the mind is nonphysical, it has no position in physical space. How, then, can a mental cause give rise to a behavioural effect that has a position in space? To put it another way, how can the nonphysical give rise to the physical without violating the laws of the conservation of mass, of energy and of momentum? (p. 124)

Despite difficulties with the notion of causation raised by Hume in his *An Enquiry Concerning Human Understanding* (See Classic Text 04) there are many clear examples of what we would want to call physical causation but none of what we might be inclined to call non-physical causation. Therefore, an appeal to scepticism concerning the nature of causality will not advance the case for dualism. As Fodor observes, "Most philosophers now agree that no argument has successfully demonstrated

why mind-body causation should not be regarded as a species of physical causation." Indeed the very practice of modern psychology presupposes that the experimental methods of physical science are applicable to the study of the mind. If, as Fodor points out, "mental processes were different in kind from physical processes, there would be no reason to expect these methods to work in the realm of the mental." (p. 124)

In stark contrast to dualism and beginning in the 1920's as a reaction to depth psychology and other traditional forms of psychology which had trouble testing their predictions experimentally, John B. Watson proposed that behaviour does not have internal mental causes. Eschewing the introspective method, Watson's behaviourism emphasises external observable behavioural reactions to given situations, rather than internal mental states. Watson's form of behaviourism was not without precedents: Ivan Pavlov in Russia had been investigating the nature of reflexes and classical conditioning while Edward Thorndike in the United States had pioneered the study of study operant (or instrumental) behaviour, both linking observable environmental stimuli and behavioural responses.

According to Fodor, "over the next thirty years psychologists such as B.F. Skinner… developed Watson's ideas into an elaborate worldview in which the role of psychology was to catalogue the laws that determine causal relations between stimuli and responses." (*l.c.*) While Skinner did not reject the presence of unobservable private mental events, including cognition and emotions, he proposed that they are subject to the same controlling variables as observable behaviours. For the radical behaviourist, the problem of explaining mind-body interaction simply vanishes because there is no such interaction - merely relations between stimuli and responses. (Wikipedia: Behaviorism)

However the idea of mental causation is deeply ingrained in our folk psychology and everyday language. We commonly attribute behaviour to knowledge, beliefs and expectations. Fodor provides the following seemingly uncontroversial examples:

> Brown puts gas in his tank because he believes the car will not run without it. Jones writes not "acheive" but "achieve" because he knows the rule about putting *i* before *e*. Even when a behavioral response is closely tied to an environmental stimulus, mental processes often intervene. Smith carries an umbrella because the sky is cloudy, but the weather is only part of the story. There are apparently also mental links in the causal chain: observation and expectation. The clouds affect Smith's behaviour only because he observes them and because they induce in him an expectation of rain. (*l.c.*)

The radical behaviourist however would, perhaps without a sense of irony, dismiss such thinking "as the residue of outworn creeds". Indeed he might contend that, as psychologists uncover more and complex relations between stimuli and responses, the need to postulate intervening mental causes would become redundant. As it turns out the opposite has happened. According to Fodor, "As psychology has matured, the framework of mental states and processes that is apparently needed to account for experimental observations has grown all the more elaborate." (p. 125)

Given that initially there seemed no materialist alternatives, for philosophers the choice in favour of radical behaviourism, however imperfect, over dualism was obvious. In the early 1960's however, doubts began to emerge that radical behaviourism and dualism exhausted all the possible

approaches to the philosophy of mind. Two materialist alternatives that allowed for mental causation began to be taken seriously, namely logical behaviourism and central-state identity theory.

According to **logical behaviourism**, attributing a mental state (such as thirst) to an organism is the same as saying that the organism is disposed to behave in a particular way (such as drinking water if it is available or seeking out water if it is not.) Logical behaviourism then is a **semantic theory** because it attempts to define what a (mental) term means. Accordingly, every mental ascription is equivalent in meaning to a conditional statement (known as a behavioural hypothetical,) in which the 'if...' part of the conditional issues in terms of stimuli only and the 'then...' part of the conditional issues in terms of behavioural responses only. Mental causation then is simply the manifestation of a behavioural disposition or as Fodor expresses it: "More precisely, mental causation is what happens when an organism has a behavioral disposition and the if-clause of the behavioral hypothetical expressing the disposition happens to be true." And because the stimuli and behavioural responses involved are physical events, logical behaviourism counts as a species of materialism. (p. 125)

Of course, it may not be practically possible to define every mental ascription by a unique behavioural hypothetical. Indeed, there may be an infinite number of behavioural hypotheticals that spell out even a simple mental ascription such as thirst. Thus, 'Smith is thirsty' might also be satisfied by, 'if there were orange juice available, Smith would drink some' and a host of other hypotheticals, too many to enumerate. The point is not that the logical behaviourist is in fact able to offer an exhaustive set of behavioural hypotheticals for every mental ascription, only that in principle the meaning of any mental term *can* be conveyed by such hypotheticals. (p. 126)

We have already seen in Classic Text 11 how David Armstrong introduced Gilbert Ryle's dispositional behaviourism, which equated mental terms with physical dispositions such as fragility or brittleness: The physical property of fragility is the disposition of such an object (say glass) to break when struck. According to Fodor, "By the same token the logical behaviorist's analysis of mental causation is similar to the received analysis of one kind of physical causation. The causal statement "The glass broke because it was fragile" is taken to mean something like "If the glass were struck, then it would break, and the glass was struck." (*l.c.*)

Because dispositional properties are philosophically uncontentious in the physical sciences, equating the meaning of mental terms with those of dispositional terms places the logical behaviourist's account of mental causation on a firmly materialist foundation. However as Fodor observes, "the analogy between mental causation as construed by the logical behaviorist and physical causation goes only so far." For one, logical behaviourism treats all mental causation as dispositional whereas there are other kinds of causation recognised by the physical sciences. The kind of causation where one physical event causes another is a case in point. When for example, we attribute the breaking of glass to its having been struck, we are explaining the manifestation of a disposition in terms of a more basic event-event causation and not the other way round. (*l.c.*)

On reflection, mental event-event causation is quite common. Except in the case of reflexes, it is very seldom that external stimuli issue directly in behavioural outcomes. There is usually an interaction between several mental causes that gives rise to behavioural effects. "For example, having a headache causes a disposition to take aspirin only if one also has the desire to get rid of the headache, the belief that aspirin exists, the belief that taking aspirin reduces headaches and so on."

Logical behaviourism however fails to take into account such causal sequences of mental events. (p. 126)

While logical behaviourism does acknowledge the existence of mental states equivalent in meaning to behavioural hypotheticals, it is committed only to facts about relations between stimuli and responses in a way that is heuristic rather than substantive. What logical behaviourism offers is merely a construal of mental causation, rather than a metaphysical commitment to the existence of such causes. "In this respect" Fodor observes, "logical behaviorism is just radical behaviorism in a semantic form. [And…] What does not really exist cannot cause anything, and the logical behaviorist, like the radical behaviorist, believes deep down that mental causes do not exist." (*l.c.*)

In contrast to logical behaviourism (and dualism in its various forms,) **central-state identity theory** (or more broadly, physicalism) holds that mental events, states or processes are *identical to* neurophysiological events in the central nervous system and that the property of being in a certain mental state is *identical to* being in a certain neurophysiological state. A particular behavioural outcome can then be seen as contingent on a sequence of neurophysiological events. Moreover the central-sate identity theory allows for the interaction of various mental causes that result in no behavioural effect, such as when a person thinks about an action but on reflection decides to do nothing. Unlike the radical behaviourist who either dismisses mental causation or the logical behaviourist who merely construes mental causation as a set of relations among stimuli and responses, the identity theorist takes mental causation (and other explanatory constructs of psychology) as literal accounts of the causal history of behaviour. (p. 127)

Identity theory is also immune to the semantic arguments levelled against logical behaviourism. The fact that the sentences "John has a headache" and "John is disposed to behave in such and such a way" don't mean the same thing is a problem for a semantic theory. However, the fact that the sentences "John has a headache" and "John is in such and such a neurophysiological state" are not synonymous is not a problem for the identity theorist so long as they are rendered true (or false) together by the occurrence of the same neurophysiological phenomena. (*l.c.*)

As we saw in Classic Text 16 there are two versions of identity theory or physicalism that differ in strength and plausibility. According to **type identity theory**, there is a one-to-one correspondence between *types* of mental events and *types* of neurophysiological events.[1] Thus for every *token* instantiation of a *type* of mental event, there is an identical physical *token* event of a single physical *type*. For example, token instances of types of mental events like pain correspond to token instances of types of neurophysiological events, such as the firing of C-fibres. For some this association would be lawful, so that in theory there ought to exist discoverable psychophysical laws. **Token identity theory** by contrast merely states that token instances of mental events, such as there are, always correspond to *some or other* token physical events. For Donald Davidson, as we saw, there is no prospect of such psychophysical laws, hence the name of his stance: '**Anomalous Monism**'.

Type identity theory rules out the possibility of computer simulations or silicon-based aliens from ever having pains because they do not have the neurophysiological hardware (or "wet-ware") in which pain states are instantiated. Token identity theory, however allows for the possible existence

---

[1] Ned Block (1996) has another neat way for remembering the difference between types and tokens: "The word 'teeth' in this sentence has five letter tokens of three letter types."

of such mental states as pains and beliefs in other vastly different organisms, so long as they are appropriately organized or in the case of machines, appropriately programmed. If it is logically possible for aliens or machines to have mental properties, then mental states and neurophysiological states cannot be identical, although some of them may, of course, be coextensive. (p. 127)

According to Fodor,

> What it all comes down to is that there seems to be a level of abstraction at which the generalizations of psychology are most naturally pitched. This level of abstraction cuts across differences in the physical composition of the systems to which psychological generalizations apply. In the cognitive sciences, at least, the natural domain for psychological theorizing seems to be all systems that process information. The problem with type physicalism is that there are possible information-processing systems with the same psychological constitution as human beings but not the same physical organization. (p. 127 - 128)

What was needed at this impasse was a *relational* account of mental properties that abstracts away from the physical structure of their bearers, yet preserves the causal character of their mind-body interaction. Whereas the logical behaviourist got the relational aspect right and was silent on the causal character of mind-body interaction, the type identity theorist at least got the causal character right. Functionalism, on the other hand, resolves this dilemma by defining mental states in terms of their causal relation to other mental states as well as sensory inputs and behavioural outputs.

According to Fodor, functionalists can make sense of both the causal and the relational character of the mental by stressing the distinction computer science draws between hardware and software. (p. 128) Perhaps this analogy was helpful at the time of publication, however we prefer to avoid it here as it invites dualistic comparisons of the "ghost in the machine" variety, to borrow a phrase from Gilbert Ryle. Besides which, Fodor himself demonstrates with his Coke machines (see below) that it is theoretically easy to construct even relatively simple machines that are able to instantiate a number of functional states without running any software.

Further advantages of functionalism are that it is both compatible with token identity theory and that it is neither reductionist nor eliminativist. Indeed functionalism takes mentalistic concepts seriously in the explanatory apparatus of psychological theories. Moreover, even if the central-state identity theory were one day to turn out to be true and mental causation is shown to be a species of physical causation, functionalism will still be able to assert that mental properties are typically defined in terms of their relations, and that mind-body interactions are typically causal, "in however robust a notion of causality is required by psychological explanations." (p. 129)

Fodor uses two simple machines to provide examples of functional states and the idea that they may be realised by multiple systems. Both of the Coke machines in the illustration on page 131 dispense a can of Coke for 10 dollar cents (ignoring inflation). The functional states of the machines are completely defined by reference to their causal roles. The one on the left has only one functional state (S0) completely specified in terms of its input and output. S0 is the function state such that, if and only if, given an input of a dime (10 cents), it dispenses a can of Coke as an output.

The machine on the right is only a little more complex in that it accepts both nickels (5 cents) and dimes, and gives change. This machine has two inter-defined functional states (S1 and S2). S1 is the state the machine is in such that, if and only if, given a nickel it dispenses nothing and goes to state S2 OR given a dime, it dispenses a Coke and stays in state S1. S2 is the state the machine is in such that, if and only if, given a nickel it dispenses a Coke and goes to S1 OR given a dime it dispenses a nickel and a Coke and goes to S1. The machine diagrams for both Coke machines are redrawn below.

	State S0
Dime input	Dispense Coke

	State S1	State S2
Nickel input	No output Go to S2	Dispense Coke Go to S1
Dime input	Dispense Coke Stay in S1	Dispense Coke and Nickle Go to S1

The machine on the left is analogous to what would satisfy a behaviourist account in that its single state (S0) is completely specified in terms of stimuli (input) and responses (output). The machine on the right however with its inter-defined states (S1 and S2) is characteristic of a functional account where the internal state the machine is already in has a material effect on the outcome given a certain input, something the behaviourist would not countenance.

Nothing about the way either machine functions dictates what it might be made of. As long as the proper relations among inputs, output and internal states can be realised, either machine could be constructed of gears and pulleys or transistors, diodes and an actuator, hence the term **multiple realisability**. That such machines may be multiply realised is an instance of token physicalism for Coke machines. Similarly, that minds like ours are neurophysiological realised is an instance of token physicalism for human beings. (p. 129)

Functionalism further entails that any substance with the right causal properties and inter-defined states could realise a Coke machine. Thus, we could have disembodied Coke machines or ones made out of ghostly ectoplasm, if such stuff exits. In the same way, functionalism does not logically require that minds be neural based, allowing for functionally comparable disembodied minds. On the other hand, just because S1 and S2 are inter-defined and multiply realisable does not mean that Coke machines have mental states. So although being inter-defined and multiply realisable are features of mental states, they are not sufficient for mentality. (*l.c.*)

Before returning to the issue on page 130, Fodor takes a slightly technical detour to address the objection that functional explanations might not be so much untrue as trivial.

> Since functionalism licenses the individuation of states by reference to their causal role, it appears to allow a trivial explanation of any observed event *E*, that is, it appears to postulate an *E*-causer. For example, what makes the valves in a machine open? Why, the operation of a valve opener. And what is a valve opener? Why, anything that has the functionally defined property of causing valves to open. (p. 129)

In psychology it is it is all too easy to explain intellectual capacities by postulating mechanisms (or *homunculi*) that have the selfsame intellectual capacities. In the case of visual perception, for example, one might postulate a mechanism that processes visual information. According to Fodor,

"The behaviorist has often charged the mentalist, sometimes justifiably, of mongering this kind of question-begging pseudo explanation." Therefore if functionally defined metal states are to have non-question-begging explanatory roles to play in psychological theories, the functionalist will have to meet the charge of triviality. Fodor believes that such an objection can be met by requiring that for functionally defined theoretical constructs, mechanisms must exist that can carry out the function and then only where we have some notion of what such mechanisms might be like. (p. 129)

Named after the English Computer Scientist and Mathematician Alan Turing (1912 - 1954), a **Turing machine** is an abstract machine with a finite number of programme states that manipulates symbols on a strip of tape of infinite length, according to some table of rules. The tape is divided into squares and the input and output of the machine are written on the tape in the form of symbols from a finite alphabet, one on each square. As the machine scans the tape, one square at time, it can erase the previous symbol and/or print a new one in its place and/or move the tape to an adjacent square.
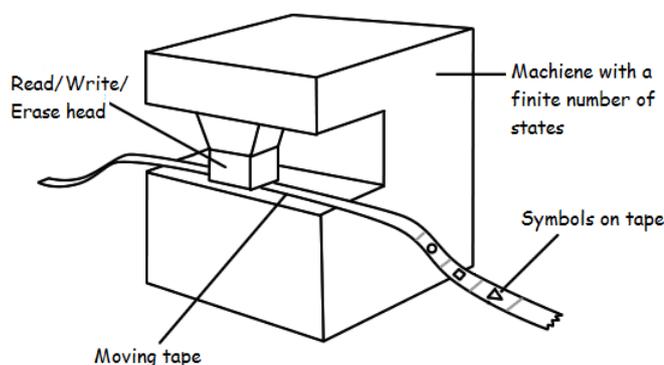
A Schematic Diagram of a Possible Turing Machine. See Text

Because the program states of a Turing machine are defined solely by the input and output symbols on the tape and by the elementary operations of the machine, as well as other states of the program, the functional role of a state is dependent on its relation to other states as well as its inputs and outputs. According to Fodor, "the relational character of the mental is captured by the Turing-machine version of functionalism." Since a program state is defined without reference to the hardware running the program, be it Lego, Silicon, neurons or a room full of people, the Turing machine version of functionalism captures the idea multiple realisability of mental states. (*l.c.*)

If we restrict functional definitions of psychological states to those that can be expressed in terms of program states of Turing machines we can provide that, in theory, such a mechanism could be realised. Thus, formulating a psychological explanation as a Turing machine program ensures that the explanation is mechanistic rather than question begging or merely semantic. Doing so also leaves open the option of how such states may be realised, whether *in vivo*, *in silico* or otherwise. Note that Fodor is not suggesting that mental processes are instantiations of Turing machine programs, instead being able to formulate a functionalist psychological theory in Turing-machine notation provides only a sufficient condition for the theory's being mechanically realizable. (p. 130)

Moreover, because Turing machines operate by the manipulation of discreet symbols, they provide a bridge between the functionalist explanations of the cognitive sciences that treat the mind as a symbolic device, on the one hand and mechanistic explanations, on the other. Another virtue of formulating a psychological explanation as a Turing machine program is to ensure that no operations have been hypothesised other than those that can ordinarily be carried out by mechanised operations. Fodor refers to this practice as "exorcising the homunculi" *cf.* the operations of "little men" invoked in non-mechanistic explanations such as Cartesian dualism. (*l.c.*)

In practice however, research psychologists cannot specify a Turing machine reduction for every functionally individuated process in every theory that they take seriously. Instead the reasoning goes, if some mental operation is essential to one of their cherished psychological explanations, then, if such reductions are considered at all, it is assumed that a Turing machine must exist that is capable of carrying out such an operation. Quite commonly we simply encounter "black boxes" in psychologists' flowcharts that serve to indicate that the Turing machine reduction of the hypothesised mental process is wanting. According to Fodor,

> Even so, the possibility in principle of such reductions serves as a methodological constraint on psychological theorizing by determining what functional definitions are to be allowed and what it would be like to know that everything has been explained that could possibly need explanation. (p. 130)

Resuming form his earlier digression, Fodor raises a fairly obvious objection to functionalism as a theory of mind. If Coke machines, catalysts, valve openers, pencil sharpeners, mouse traps and ministers of finance are all concepts that can be functionally defined without being mental concepts the way that, say, pain or belief are, then what is it that characterises the mental as such? And can it, whatever it may be, be captured by functionalism? (*l.c.*)

Historically, mental states have been characterised by their subjective quality or qualitative content; something that can only be known by direct experience. Suppose that you look at a blank wall through a red filter. What was your experience? Suppose further that you leave everything as before and switch to a green filter instead. Something about your experience will have changed. This something is what philosophers have referred to as subjective quality or qualitative content. Suppose also that you have an application on your phone that detects frequencies of light via its inbuilt camera. If you point the camera at the wall through first the red and then the green filter the phone would register a change in the frequency of the light. However we would be very reluctant to say of the phone that there was any subjective quality or qualitative content associated with the detection, which after all, is merely the functional output of a software application running on the phone's hardware.

The **inverted spectrum argument**, introduced by Fodor on page 130 but dating back to John Locke, purports to show that it is possible for two observers to have all the same causal relations and yet differ in their subjective quality or qualitative content. Imagine two observers alike in all relevant psychological respects, except that when the one observer has the experience of seeing red, the other has the experience of seeing green. Of course, when viewing ripe tomatoes or flaming sunsets they would both name the colour "red". Perhaps they would even both think about the story of Little Red Riding Hood when the saw the colour red and verdant pastures when they saw the colour green. So if it is possible for the causal roles of the experiences of two observers to be the same, and yet differ in terms of their subjective quality or qualitative content of their mental state, then the functionalist account does not work for mental states that have subjective quality or qualitative content. (*l.c.*)

It could be answered that an inverted spectrum cannot, in fact, be consistently imagined. A stronger retort is that inverted spectra are a biological impossibility, given the way that red, green and blue cones in our retina are primed to detect differing ranges of frequency of visible light as a function of the slightly different light sensitive pigments that each of them contain. A similar objection could be

made against the possibility of an inverted sound spectrum: the cochlea would have to be differently structured or differently innervated. The conclusion that Fodor does not take seriously is that it is not possible to physically (biologically) instantiate two functionally identical percepts that yet differ in subjective quality or qualitative content.

Turning to the intentional content of mental states: According to Fodor,

> To say that a mental state has intentional content is to say that it has certain semantic properties. For example, for Enrico to believe Galileo was Italian apparently involves a three-way relation between Enrico, a belief and a proposition that is the content of the belief (namely the proposition that Galileo was Italian). In particular it is an essential property of Enrico's belief that it is about Galileo (and not about, say, Newton) and that it is true if, and only if, Galileo was indeed Italian. (p. 130)

While it is widely agreed that that beliefs involve semantic properties, such as expressing propositions which are capable of being true or false in a given context, the question as to whether semantic properties are fundamental or whether and how they are derived from other non-semantic states has not yet been settled. Nevertheless beliefs, however construed, are important in the cognitive sciences because, for example, they feature in theories of learning and perception as accounts of how organisms are determined by their experiences and hereditary. Still, functionalist accounts of mental states do not by themselves account for semantic properties, as Fodor remarks: "Mousetraps are functionally defined, yet mousetraps do not express propositions and they are not true or false." (*l.c.*)

Symbols, on the other hand, do have intentional content - they are about certain things other than themselves. Thus the utterance "Galileo was Italian", like Enrico's belief, is about Galileo only and is true or false depending on Galileo's nationality. There is therefore something quite alike about the symbolic and the mental that led Fodor to propose the existence of mental symbols or mental representations that have semantic properties. Thus, having a belief involves a relation to a mental symbol from which the belief derives its semantic properties. Cognitive processes then, such as thinking, perceiving, learning *etc.* "involve causal interactions among relational states such as having a belief. The semantic properties of the words and sentences we utter are in turn inherited from the semantic properties of the mental states that language expresses." (p. 131)

One immediate objection to what has become known as the "Language of Thought Hypothesis" is that it merely shifts the burden of explanation for linguistic intentional content to the level of the mental in the form of a "Language of Thought" or "mentalese" (měn'tl-ēz') which itself stands in need of explanation. However there are further parallels between the semantic properties of mental states and those of postulated mental symbols that make them compatible with the computer metaphor of the mind. After all, a computer is a device that manipulates symbols and a computation is a causal chain of computer states where the links in the chain are operations on semantically interpreted formulae in a machine code. Therefore, the analogy between computers and minds actually requires that there must be mental symbols, since there can be "no computation without representation." (*l.c.*)

Skipping over the now discredited resemblance theory of representation, Fodor offers the functionalist suggestion that the semantic properties of mental representation are causally

determined by aspects of their functional role. There are three types of causal relations among psychological states involving mental representations: those between stimuli and mental states, those between mental sates and responses and those between mental sates and other mental states. The hope then is that such relations will be sufficient to "fix" the semantic properties of mental representations. (p. 132)

Consider Fodor's example of the belief that John is tall:

> Presumably the following facts, which correspond respectively to the three types of causal relation, are relevant to determining the semantic properties of the mental representation involved in the belief. First, the belief is a normal effect of certain stimulations, such as seeing John in circumstances that reveal his height. Second, the belief is the normal cause of certain behavioral effects, such as uttering "John is tall." Third, the belief is a normal cause of certain other beliefs and a normal effect of certain other beliefs. For example, anyone who believes John is tall is very likely also to believe someone is tall. Having the first belief is normally causally sufficient for having the second belief. And anyone who believes everyone in the room is tall and also believes John is in the room will very likely believe John is tall. The third belief is a normal effect of the first two. In short, the functionalist maintains that the proposition expressed by a given mental representation depends on the causal properties of the mental states in which that mental representation figures. (*l.c.*)

Whether or not one buys into the Language of Thought Hypothesis, no one (other than a behaviourist) would deny that we do form mental representations that have semantic content. If functionalism is broadly correct in identifying mental states with their functional role in causal relation to other mental states as well as their sensory inputs and behavioural outputs, then the semantic properties of such representations must also be determined by aspects of their functional role. Furthermore, the concept of mental representation has theoretical work to do in the cognitive sciences in formulating and testing hypotheses in empirical theories of mind.

Fodor does not claim that the functional account of the semantic properties of mental representations is necessarily true, especially in the "bald and unelaborated way" that he sketched it. However, at the time, the science of mental representation was a flourishing field of research, ranging from phonetics to computer vision. Despite behaviourist's methodological scruples over which mechanisms may or may not figure in psychological theories, Fodor points out that, as the history of science reveals, "when a successful theory comes into conflict with a methodological scruple, it is generally the scruple that gives way." So instead of deciding what, according to one's philosophical scruples, is methodologically permissible in science, it is surely better to investigate what a successful science requires.

Tasks

1. We are all familiar with certain non-psychological functional concepts which we accept without controversy. Can you think of any?
2. How does functionalism fit in with the other philosophical theories of mind that we have dealt with so far?
3. Except for the inverted spectrum thought experiment, Fodor (in this article) has not provided a critique of functionalism. Via a quick on-line search, find one or more classic

arguments against functionalism. State them and say whether you think they are serious concerns. (Block, 1980 is also a good source)

4. Given the alternatives would you endorse a functionalist analysis of mental causation?

Feedback

1. According to Block (1996) "One motivation behind functionalism can be appreciated by attention to artifact concepts like *carburetor* and biological concepts like *kidney*." It is not necessary or indeed desirable to define the concept of a carburettor according to one structural model or another when what we mean is any device (in an internal combustion engine) that mixes fuel and air. Similarly because the anatomical structure of a kidney varies so considerably in nature from creatures as diverse as snails and humans, the concept of a kidney can only be functionally defined as an organ that filters blood and regulates salt balance. If we were to discover a species of animal entirely new to science biologists would have no trouble in deciding whether or not one (or more) of its organs were a kidney based its function.

2. Functionalism developed as alternative to Skinnerian behaviourism on the one hand and type identity theory and eliminative materialism on the other. Because functional states are multiply realisable, functionalism is compatible with token identity theory. Although the early protagonists of functionalism such as Putnam, Fodor, Armstrong and Lewis eschewed Cartesian dualism, functionalism does curiously allow for the instantiation of mental states by non-physical substance (if such stuff exists) so long as it realises the correct functional causal relations. Although we do have metaphysical qualms with the notion of physical causation, our ordinary understanding of the concept has been adequate, indeed necessary, for the articulation and develop of psychology and many other sciences, whereas we cannot even begin to imagine what non-physical causation might entail. Although functionalism and Davidson's anomalous monism are both species of token-identity theory they stem from very different concerns. The monism part of Davidson's Anomalous monism is based on an *a priori* argument from his three principles (See Classic Text: 16) however functionalism appears to take monism on as an assumption. Accordingly, "some physical media must play the realizing role—hence, monism." (Yalowitz, 2014 for further discussion.)

3. Most of the arguments against functionalism involve thought experiments of varying quality. Here we summarise one that has received a lot of coverage. John Searle's (1980) "Chinese Room" thought experiment involves a man (Searle) who only speaks English in a room with baskets of Chinese symbols and a manual, written in English, for manipulating the symbols. People outside the room are Chinese speakers and they communicate with the man by passing Chinese symbols into the room via a slot in the wall. The man looks up the symbols in his manual and then passes the appropriate symbols from the baskets back to the people outside the room. In this way, the people outside the room come to believe that they are holding a meaningful conversation with the man inside the room, whom they believe understands Chinese, even though he obviously doesn't. The thought experiment purports to show that it is possible to simulate intelligent conversation through the use of a purely functional system, without any understanding taking place. The "Chinese Room" in this

thought experiment has all features essential to a functionalist account of a mental process, such as understanding, constituted by its functional role in terms of causal relations among inputs and output via the slot, as well as those internal states within the room codified by the manual. Hypothetically, the man could be replaced by a Turing machine with an appropriately encoded tape that allows it to manipulate the symbols in such a way that those outside the room would be similarly fooled into believing that they are having a meaningful conversation, in Chinese, with a person inside. Searle's original paper was intended to refute the belief in the possibility of intentionality being realised within artificial intelligence systems, which are after all functional systems. So if the "Chinese Room" thought experiment is a good analogy of a functional system that appears to mimic understanding without understanding, then then a purely functional analysis of understanding must be wanting. There are multiple responses to Searle's thought experiment ranging from those who bite the bullet and insist that even if the man does not understand Chinese the room as a functional system does, to those who believe the analogy makes false assumptions about subjective conscious, and so proves nothing. (Wikipedia: Chinese room) Whether or not you are convinced by the "Chinese Room" thought experiment, the point of the exercise is to demonstrate that functionalism is not without its detractors, of which there are quite a few.

4. Given the alternatives examined so far (from ghosts to black boxes and beyond) one or other form of functionalism is lightly to be the closest to correct analysis of what makes certain physical states count as mental states, if not in detail, then in outline. Functionalism in its crudest form has nothing to say about the ontology of mental states, other than that they are multiply realisable. If however we add the Turing machine requirement that ensures that there exists, in theory, a (physical) mechanism by which psychological processes could be realised, then some form of Token Identity theory will be required. We have not yet considered the notion of supervenience but functionalism and Token Identity theory are compatible with, if not required, for the supervenience of psychological states.

In the next classic test we will consider the problem of universals.

References

BLOCK, N. (1980) "Troubles with Functionalism." **In** Block, N., *Readings in the Philosophy of Psychology*, Vol. 1. Harvard University Press: Cambridge, MA

BLOCK, N. (1996) "What is functionalism?" a revised version of the entry on functionalism in *The Encyclopedia of Philosophy Supplement*, Macmillan.

FODOR, J. (1981) The Mind-Body Problem. *Scientific American* **244** (1) p, 114 - 123

SEARLE, J. (1980) "Minds, Brains and Programs", *Behavioral and Brain Sciences*, **3** (3) p. 417 - 457

YALOWITZ, S. "Anomalous Monism", *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.) http://plato.stanford.edu/archives/win2014/entries/anomalous-monism/