

Functionalism

According to *functionalism*, the essential or defining feature of any type of mental state is the set of causal relations it bears to (1) environmental effects on the body, (2) other types of mental states, and (3) bodily behavior. Pain, for example, characteristically results from some bodily damage or trauma; it causes distress, annoyance, and practical reasoning aimed at relief; and it causes wincing, blanching, and nursing of the traumatized area. Any state that plays exactly that functional role is a pain, according to functionalism. Similarly, other types of mental states (sensations, fears, beliefs, and so on) are also defined by their unique causal roles in a complex economy of internal states mediating sensory inputs and behavioral outputs.

This view may remind the reader of behaviorism, and indeed it is the heir to behaviorism, but there is one fundamental difference between the two theories. Where the behaviorist hoped to define each type of mental state solely in terms of environmental input and behavioral output, the functionalist denies that this is possible. As he sees it, the adequate characterization of almost any mental state involves an ineliminable reference to a variety of other mental states with which it is causally connected, and so a reductive definition solely in terms of publicly observable inputs and outputs is quite impossible. Functionalism is therefore immune to one of the main objections against behaviorism.

Thus the difference between functionalism and behaviorism. The difference between functionalism and the identity theory will emerge from the following argument raised against the identity theory.

Imagine a being from another planet, says the functionalist, a being with an alien physiological constitution, a constitution based on the chemical element silicon, for example, instead of on the element carbon, as ours is. The chemistry and even the physical structure of the alien's brain would have to be systematically different from ours. But even so, that alien brain could well sustain a functional economy of internal states whose mutual relations parallel perfectly the mutual *relations* that define our own mental states. The alien may have an internal state that meets all the conditions for being a pain state, as outlined earlier. That state, considered from a purely physical point of view, would have a very different makeup from a human pain state, but it could nevertheless be identical to a human pain state from a purely functional point of view. And so for all of his functional states.

If the alien's functional economy of internal states were indeed *functionally isomorphic* with our own internal economy—if those states were causally connected to inputs, to one another, and to behavior in ways that parallel our own internal connections—then the alien would have pains, and desires, and hopes, and fears just as fully as we, despite the differences in the physical system that sustains or realizes those functional states. What is important for mentality is not the matter of which the creature is made, but the structure of the internal activities which that matter sustains.

If we can think of one alien constitution, we can think of many, and the point just made can also be made with an artificial system. Were we to create an electronic system—a computer of some kind—whose internal economy were functionally isomorphic with our own in all the relevant ways, then it too would be the subject of mental states.

What this illustrates is that there are almost certainly many more ways than one for nature, and perhaps even for man, to put together a thinking, feeling, perceiving creature. And this raises a problem for the identity theory, for it seems that there is no single type of physical state to which a given type of mental state must always correspond. Ironically, there are *too many* different kinds of physical systems that can realize the functional economy characteristic of conscious intelligence. If we consider the universe at large, therefore, and the future as well as the present, it seems quite unlikely that the identity theorist is going to find the one-to-one match-ups between the concepts of our common-sense mental taxonomy and the concepts of an overarching theory that encompasses all of the relevant physical systems. But that is what intertheoretic reduction is standardly said to require. The prospects for universal identities, between types of mental states and types of brain states, are therefore slim.

If the functionalists reject the traditional 'mental-type = physical type' identity theory, virtually all of them remain committed to a weaker 'mental token = physical token' identity theory, for they still maintain that each *instance* of a given type of mental state is numerically identical with some specific physical state in some physical system or other. It is only universal (type/type) identities that are rejected. Even so, this rejection is typically taken to support the claim that the science of psychology is or should be *methodologically autonomous* from the various physical sciences such as physics, biology, and even neurophysiology. Psychology, it is claimed, has its own irreducible laws and its own abstract subject matter.

As this book is written, functionalism is probably the most widely held theory of mind among philosophers, cognitive psychologists, and artificial intelligence researchers. Some of the reasons are apparent from the preceding discussion, and there are further reasons as well. In characterizing mental states as essentially functional states, functionalism places the concerns of psychology at a level that abstracts from the teeming detail of a brain's neurophysiological (or crystallographic, or microelectronic) structure. The science of psychology, it is occasionally said, is methodologically autonomous from those other sciences (biology, neuroscience, circuit theory) whose concerns are with what amount to engineering details. This provides a rationale for a great deal of work in cognitive psychology and artificial intelligence, where researchers postulate a system of abstract functional states and then test the postulated system, often by way of its computer simulation, against human behavior in similar circumstances. The aim of such work is to discover in detail the functional organization that makes us what we are. (Partly in order to evaluate the prospects for a functionalist philosophy of mind, we shall examine some of the recent research in artificial intelligence in chapter 6.)

Arguments against Functionalism

Current popularity aside, functionalism also faces difficulties. The most commonly posed objection cites an old friend: sensory qualia. Functionalism may escape one of behaviorism's fatal flaws, it is said, but it still falls prey to the other. By attempting to make its *relational* properties the definitive feature of any mental state, functionalism ignores the 'inner' or qualitative nature of our mental states. But their qualitative nature is the essential feature of a great many types of mental state (pain, sensations of color, of temperature, of pitch, and so on), runs the objection, and functionalism is therefore false.

The standard illustration of this apparent failing is called "the inverted spectrum thought-experiment". It is entirely conceivable, runs the story, that the range of color sensations that I enjoy upon viewing standard objects is simply inverted relative to the color sensations that you enjoy. When viewing a tomato, I may have what is really a sensation-of-green where you have the normal sensation-of-red; when viewing a banana, I may have what is really sensation-of-blue where you have the normal sensation-of-yellow; and so forth. But since we have no way of comparing our inner qualia, and since I shall make all the same observational discriminations among objects that you will, there is no way to tell whether my spectrum is inverted relative to yours.

The problem for functionalism arises as follows. Even if my spectrum is inverted relative to yours, we remain *functionally* isomorphic with one another. My visual sensation upon viewing a tomato is functionally identical with your visual sensation upon viewing a tomato. According to functionalism, therefore, they are the very same type of state, and it does not even make sense to suppose that my sensation is 'really' a sensation-of-green. If it meets the functional conditions for being a sensation-of-red, then by definition it is a sensation-of-red. According to functionalism, apparently, a spectrum inversion of the kind described is ruled out by definition. But such inversions are entirely conceivable, concludes the objection, and if functionalism entails that they are not conceivable, then functionalism is false.

Another qualia-related worry for functionalism is the so-called "absent qualia problem". The functional organization characteristic of conscious intelligence can be instantiated (= realized or instanced) in a considerable variety of physical systems, some of them radically different from a normal human. For example, a giant electronic computer might instantiate it, and there are more radical possibilities still. One writer asks us to imagine the people of China—all 10^9 of them—organized into an intricate game of mutual interactions so that collectively they constitute a giant brain which exchanges inputs and outputs with a single robot body. That system of the robot-plus- 10^9 -unit-brain could presumably instantiate the relevant functional organization (though no doubt it would be much slower in its activities than a human or a computer), and would therefore be the subject of mental states, according to functionalism. But surely, it is urged, the complex states that there play the functional roles of pain, pleasure, and sensations-of-color would not have intrinsic qualia as ours do, and would therefore fail to be genuine mental states. Again, functionalism seems at best an incomplete account of the nature of mental states.

It has recently been argued that both the inverted-qualia and the absent-qualia objections can be met, without violence to functionalism and without significant violence to our

common-sense intuitions about qualia. Consider the inversion problem first. I think the functionalist is right to insist that the type-identity of our visual sensations be reckoned according to their functional role. But the objector is also right in insisting that a relative inversion of two people's qualia, without functional inversion, is entirely conceivable. The apparent inconsistency between these positions can be dissolved by insisting that (1) our functional states (or rather, their physical realizations) do indeed have an intrinsic nature on which our introspective identification of those states depends; while also insisting that (2) such intrinsic natures are nevertheless not essential to the type-identity of a given mental state, and may indeed *vary* from instance to instance of the same type of mental state.

What this means is that the qualitative character of your sensation-of-red might be different from the qualitative character of my sensation-of-red, slightly or substantially, and a third person's sensation-of-red might be different again. But so long as all three states are standardly caused by red objects and standardly cause all three of us to believe that something is red, then all three states are sensations-of-red, whatever their intrinsic qualitative character. Such intrinsic qualia merely serve as salient features that permit the quick introspective identification of sensations, as black-on-orange stripes serve as a salient feature for the quick visual identification of tigers. But specific qualia are not essential to the type-identity of mental states, any more than black-on-orange stripes are essential to the type-identity of tigers.

Plainly, this solution requires the functionalist to admit the *reality* of qualia, and we may wonder how there can be room for qualia in his materialist world-picture. Perhaps they can be fit in as follows: *identify* them with physical properties of whatever physical states instantiate the mental (functional) states that display them. For example, identify the qualitative nature of your sensations-of-red with that physical feature (of the brain state that instantiates it) to which your mechanisms of introspective discrimination are in fact responding when you judge that you have a sensation-of-red. If materialism is true, then there must *be* some internal physical feature or other to which your discrimination of sensations-of-red is keyed: *that* is the quale of your sensations-of-red. If the pitch of a sound can turn out to be the frequency of an oscillation in air pressure, there is no reason why the quale of a sensation cannot turn out to be, say, a spiking frequency in a certain neural pathway. (More likely, it will be a peculiar group or set of spiking frequencies, as claimed by the *vectorial* or *across-fiber pattern* theory of sensory coding. 'Spikes' are the tiny electrochemical pulses by which our brain cells communicate with each other along the thin fibers that connect them. For more on all of this, see Chapter 7.)

This entails that creatures with a physical constitution different from ours may have qualia different from ours, despite being psychologically isomorphic with us. It does not entail that they *must* have different qualia, however. If the qualitative character of my sensation-of-red is really a spiking frequency of 90 hertz in a certain neural pathway, it is possible that an electromechanical robot might enjoy the very same qualitative character if, in reporting sensations-of-red, the robot were responding to a spiking frequency of 90 hertz in a corresponding *copper* pathway. It might be the spiking frequency that matters

to our respective mechanisms of discrimination, not the nature of the medium that carries it.

This proposal also suggests a solution to the absent qualia problem. So long as the physical system at issue is functionally isomorphic with us, to the last detail, then it will be equally capable of subtle introspective discriminations among its sensations. Those discriminations must be made on some systematic physical basis, that is, on some characteristic physical features of the states being discriminated. Those features at the objective focus of the system's discriminatory mechanisms, *those* are its sensory qualia—though the alien system is no more likely to appreciate their true physical nature than we appreciate the true physical nature of our own qualia. Sensory qualia are therefore an inevitable concomitant of any system with the kind of functional organization at issue. It may be difficult or impossible to 'see' the qualia in an alien system, but it is equally difficult to 'see' them even when looking into a human brain.

I leave it to the reader to judge the adequacy of these responses. If they are adequate, then, given its other virtues, functionalism must be conceded a very strong position among the competing contemporary theories of mind. It is interesting, however, that the defense offered in the last paragraph found it necessary to take a leaf from the identity theorist's book (types of quale are reduced to or identified with types of physical state), since the final objection we shall consider also tends to blur the distinction between functionalism and reductive materialism.

Consider the property of *temperature*, runs the objection. Here we have a paradigm of a physical property, one that has also been cited as the paradigm of a successfully *reduced* property, as expressed in the intertheoretic identity

"temperature = mean kinetic energy of constituent molecules".

Strictly speaking, however, this identity is true only for the temperature of a gas, where simple particles are free to move in ballistic fashion. In a *solid*, temperature is realized differently, since the interconnected molecules are confined to a variety of vibrational motions. In a *plasma*, temperature is something else again, since a plasma has no constituent molecules; they, and their constituent atoms, have been ripped to pieces. And even a *vacuum* has a so-called 'blackbody' temperature—in the distribution of electromagnetic waves coursing through it. Here temperature has nothing to do with the kinetic energy of particles.

It is plain that the physical property of temperature enjoys 'multiple instantiations' no less than do psychological properties. Does this mean that thermodynamics (the theory of heat and temperature) is an 'autonomous science', separable from the rest of physics, with its own irreducible laws and its own abstract nonphysical subject matter?

Presumably not. What it means, concludes the objection, is that *reductions are domain-specific*:

temperature-in-a-gas = the mean kinetic energy of the gas's molecules,

whereas

temperature-in-a-vacuum = the blackbody distribution of the vacuum's transient radiation.

Similarly, perhaps

joy-in-a-human = resonances in the lateral hypothalamus,

whereas

joy-in-a-Martian = something else entirely.

This means that we may expect some type/type reductions of mental states to physical states after all, though they will be much narrower than was first suggested. Furthermore, it means that functionalist claims concerning the radical autonomy of psychology cannot be sustained. And last, it suggests that functionalism is not so profoundly different from the identity theory as was first made out.

As with the defense of functionalism outlined earlier, I leave the evaluation of this criticism to the reader. We shall have occasion for further discussion of functionalism in later chapters. At this point, let us turn to the final materialist theory of mind, for functionalism is not the only major reaction against the identity theory.