

## THE IDENTITY THEORY AND FUNCTIONALISM

### The Identity Theory

U.T. Place and J.J.C. Smart wanted to avoid the bad consequences of Behaviorism, but without falling back into Cartesian Dualism. They did that ingeniously, by hypothesizing that mental states and events are indeed real occurrences and internal to us, but nonetheless physical: mental states and events are neurological states and events (literally) inside our brains. For example (Putnam's), perhaps *pain* is the firing of c-fibers. The model here is empirical scientific identifications: water with H<sub>2</sub>O, lightning with electrical discharge, and genes with segments of DNA molecules.

This Identity thesis avoids not only the usual objections to Dualism, but our three objections to Behaviorism as well. And it avoids them for the right reasons: Behaviorism went wrong precisely by denying that mental states and events are *internal causes of behavior*.

☺♥ Euphorical ♥☺

Even better, there are two direct arguments for the Identity Theory. The Argument from Correlations: Smart appeals to mind-brain correlations: It's reasonable to think that every mental state or event at least has a *corresponding* brain state or event. The best, because most parsimonious, explanation of those correlations is that the mental states/events *just are* the "corresponding" brain states/events. (In general: When Xs are invariably accompanied by Ys and you can find nothing to distinguish Xs from Ys, the best explanation is that Xs just are Ys.)

Smart is presupposing, epistemologically, that parsimony or simplicity is a reason for believing one hypothesis rather than another. That is controversial as philosophy of science. Science itself seems simply to assume it (the curve-fitting example). But some philosophers of science ridicule it: We should accept hypotheses on the basis of evidence, not on the basis of what are really only aesthetic properties, like simplicity. (What next? Prettiness? Cuteness? Being set to music?)

The Causal Argument (Armstrong, Lewis): Mental concepts are causal role concepts, and so they afford role-occupant identifications (cf. genes and segments of DNA molecules). E.g.:

1. Pain = Whatever state of a person plays role P (being typically caused by tissue damage, and in turn causing wincing, crying out, withdrawal, favoring, etc.) [We know this a priori; we've all got the concept of pain.]
  2. The occupant of role P = the firing of c-fibers (i.e., it is c-fiber firings that are typically caused by tissue damage, etc.). [Discovered empirically by neuroscientists.]
- 

∴ 3. Pain = the firing of c-fibers. QED

I'll leave the evaluation of this argument to you. Campbell criticizes premise (1), the Causal analysis of pain, by invoking his "imitation man" (a cousin of the Tinfoil Man).

Well, nothing's easy (not even in the 1960s)--even once we've set aside the "ineffectual" objections, Smart's 1, 2 and 7.

Objection 4: Let's grant to the objector that the after-image is not in physical space (even though it has a location in the subject's phenomenal field). Let's also grant the argument's validity. But (first reply:) Smart is right—no one has said that *the after-image*, the little shimmering spot, is identical with something in the brain. It's the *experience of having-an-after-image* or "after-imaging" that the Identity Theorists identified with an event or episode in the brain. Rejoinder: But what *about* the after-image itself? Smart's reply to that: "[T]here is, in a sense, no such thing as an after-image." After-images are not *things*. (I, WGL, will argue later in the course that after-images damn well are things and that that is a big problem for the materialist.)

Objection 6: I am infallible and incorrigible about my own current sensations. I am not infallible or incorrigible about c-fiber firings or about any other physical processes for that matter. Therefore, sensations are not physical processes. Possible reply: You can be mistaken about your own current sensations. (The fraternity pledge example.) Smart's reply: Infallibility and incorrigibility are *description-relative*. When we call something a "sensation," by the logic of the term the subject of a "sensation" can't be mistaken about it. That doesn't show that one can't be mistaken about it under another label. Different logics doesn't entail that the relevant terms refer to distinct entities: "I see the Morning Star" entails that it's morning while "I see the Evening Star" does not entail that, but the two terms refer to one and the same entity, the planet Venus.

WGL's closely related reply: This argument is bad in just the way Descartes' Doubt argument was bad; indeed it's a version of his argument.

We didn't talk about the famous Objection 3 due to Max Black, because it's complicated and difficult. I've posted the special freebie handout on it for those who are interested.

Fodor and Putnam's objection: The Identity Theory entails that for every type of mental state/event, there is a corresponding physiological type of state/event (with which it is identical). But this leads to a sort of species chauvinism. If pain is identical with, is nothing but, the firing of c-fibers, then in order for any creature to be in pain, the creature must have c-fibers. But there is no reason to think that anything so physiologically specific as c-fibers is required for any creature whatever—mongoose, mollusc, or Martian—to feel pain. The creature need only have some physiological structure of its own that does in it what our c-fibers do in us.

### Functionalism

Seizing on that last idea, that of the *job done* in us by physiological activity in the brain, the Functionalist identifies types of mental state/event with *functional roles*, the psychological roles characteristically played in us by the physiological structures. The Functionalist agrees with the Identity Theorist that every individual mental state/event (every mental "token," as we say) is identical with some physiological state/event token; but s/he denies the Identity Theorist's implication that for every type of mental state/event, there is a corresponding *physiological type* of state/event. Rather, mental types are to be identified with functional roles. Such roles, like mental states themselves, are *multiply realizable*; they can be played by different physiological structures in different creatures.

(Armstrong and Lewis too think in terms of roles, of neurophysiological states being mental states in virtue of occupying the relevant causal roles. But there are two important differences between the Causal version of the Identity Theory and Functionalism. First, the roles appealed to by the Causal theorists are characterized entirely in commonsense terms, while the Functionalist imagines them as being described in some technical vocabulary following the appropriate research. Second, in the same vein, the Causal analysis of the mental was a priori, a *conceptual* analysis of the meanings of mental terms; Functionalism, like the Identity Theory, is an empirical hypothesis or bet.)

**TERMINOLOGICAL WARNING** (to be repeated at more length in the handout, "Types and Tokens"): In next week's reading, Ned Block uses the label

“Functionalism” much more broadly than I do here. He uses it (rightly) to include Putnam’s theory, which he calls “machine functionalism,” and Psychofunctionalism, but also to include the Armstrong- Lewis Causal Theory, which he calls “a priori functionalism” and “Functionalism” with capital F. People also sometimes call the Armstrong-Lewis theory “Analytical Functionalism,” which is similarly misleading because the theory has nothing to do with *functions* in either sense of the term.

Fodor and Putnam thought of the (psycho-)functional roles *computationally*, and bought into the **computer model of the mind**: Mental states are like the computational (“logical,” “functional”) states of a computer. Software, not hardware. (Cf. the information-processing models simultaneously emerging in cognitive psychology, and AI.) Call this version of Functionalism **Machine Functionalism**.

The computer model accommodates multiple realizability as a matter of course; a single given computer program can be run on any number of different hardware configurations.

(Strictly speaking, Functionalism is not *per se* a physicalist or materialist view; it is compatible with Cartesian Dualism, though one wonders why anyone would want to be a Functionalist Dualist.)

Of course Machine Functionalism has its problems. Objection 1: If mental states are just computational states, it follows that a sufficiently big and powerful computer would have mental states. But that’s silly; computers can’t have genuine mental states, even though we sometimes apply mental talk metaphorically to them. In particular, a mere machine could not be conscious. For one thing, they’re not alive. (Only biologic or living organisms can be conscious.)

Objection 2: What is functional realization? It cannot be just a matter of one-one correspondence, because one-one correspondence is cheap. (Hinckfuss’ pail; Block’s Chinese giant.) Block thinks the Machine Functionalist faces a dilemma regarding the inputs and the outputs.

Objection 3: Machine Functionalism’s two-levelled picture, “the hardware” vs. “the software,” is unbiological in the extreme (and not even true of computers themselves). Reply: All right, just lose the two-level assumption and allow the computational model to be richly hierarchical. (That will also undercut Block’s dilemma regarding chauvinism and liberalism.)

## Teleofunctionalism

One solution to the problem of liberalism is to impose a teleological requirement on functional realization. Thus, a pun: Functions, not in the computational, mathematical sense, but in the biological sense. The trouble with China, the pail et al. is that they are not *organisms* or even coherent, teleologically organized systems. An organism has organs, characterized in terms of their functions. Let us then tighten our notion of functional “realization”: A physiological state will not be counted as *realizing* a program state unless its function is to do so, i.e., its job is to play the program role it does; the organism must have genuine organic integrity and the physiological state plays its functional role properly *for* the organism, in the teleological sense of “for” and in the teleological sense of “function.” The state must do what it does as a matter of, so to speak, its biological purpose.<sup>1</sup>

A further advantage of Teleofunctionalism is that it suggests a better picture of psychological explanation, one that sits well with the actual explanatory programs of cognitive science. The Machine Functionalist still conceived psychological *explanation* in the Positivists' terms of subsumption of data under wider and wider universal generalizations. But Fodor (1968), Dennett (1978) and Cummins (1983) then defended a competing picture of psychological explanation, according to which behavioral data are to be seen as manifestations of subjects' psychological capacities, and those capacities are to be explained by understanding the subjects as systems of interconnected components. Each component is a “homunculus,” in that it is identified by reference to the function it performs, and the various homuncular components cooperate with each other in such a way as to produce overall behavioral responses to stimuli. The “homunculi” are themselves broken down into subcomponents whose functions and interactions are similarly used to explain the capacities of the subsystems they compose, and so again and again until the sub-sub- . . . components are seen to be neurophysiological structures. (An automobile works—locomotes—by having a fuel reservoir, a fuel line, a carburetor (in 1955 anyway), a combustion chamber, an ignition system, a transmission, and wheels that turn. If one wants to know how the carburetor works, one will be told what its parts are and how they work together to infuse oxygen into fuel; and so on.) Thus biologic and mechanical systems alike are

---

<sup>1</sup> Moreover. Ruth Millikan, Bob Van Gulick, Jerry Fodor, Fred Dretske and others have argued powerfully that teleology must enter into any adequate analysis of the intentionality or aboutness of mental states such as beliefs and desires. According to the teleological theorists, a neurophysiological state should count as *a belief that broccoli will kill you*, and in particular as *about broccoli*, only if that state has the indicating of broccoli as in some sense one of its psychobiological functions. More on this in due course, when we get to intentionality proper.

hierarchically organized, on the principle of what computer scientists used to call “hierarchical control.” This particular version of the teleological turn is called “Homuncular Functionalism.”

All this talk of teleology and biological function seems to presuppose that biological and other “structural” states of physical systems really have functions in the teleological sense. The latter claim is controversial, though not nearly as much so as it was throughout most of the 20<sup>th</sup> century. Some philosophers might still dismiss it as a superstitious relic of primitive animism or Panglossian theism or at best the vitalism of the 19<sup>th</sup> century; others would tolerate it but only as a useful metaphor; still others take teleological characterizations to be literally but only interest-relatively true, true *modulo* a convenient classificatory or interpretive scheme (Cummins 1975). This may seem to embarrass teleologized functionalist theories of mind.

Yes and no. Yes, because if a Homuncular and/or Teleological Functionalist type-identifies mental items with teleologically characterized items, and teleological characterizations are not literally true, then mental ascriptions cannot be literally true either. Equivalently, if people really do have mental states and events, on their own and not merely in virtue of anyone's superstitious or subjective interpretation of them, but their physical states do not have objectively teleological functions, then mental states cannot be type-identified with teleological states.

Fortunately for the Teleological Functionalist there is now a small but vigorous industry whose purpose is to explicate biological teleology in naturalistic terms, typically (though not always!) in terms of etiology. For example, a trait may be said to have the function of doing F in virtue of its having been selected for because it did F; a heart's function is to pump blood because hearts' pumping blood in the past has given them a selection advantage and so led to the survival of more animals with hearts. Philosophers of biology have continued to refine the early accounts and to make them into adequate naturalistic analyses of genuine function (Neander 1991, Godfrey-Smith 1994).

It should be noted that the correctness of type-identifying mental items with teleological items does not strictly depend on the objectivity or even the truth of teleological descriptions. For corresponding to each metaphysical view of teleology, including deflationary and flatly derisive ones, there is a tenable view of mind. Just as teleology may be a matter of interest-relative interpretation, so, after all, may mental ascriptions be. For that matter, just as teleology may be only metaphorical, fictional, or illusory, so may mental ascriptions be; some philosophers now hold that mental ascriptions are in the end false. But we shall consider those possibilities in due course.