FODOR

A Historical Introduction to the Philosophy of Mind

Readings With Commentary

Peter A. Morton

broadview press

1997

Jerry A. Fodor

Selections from "Something on the State of the Art"

A census of the main problems in the philosophy of mind as they presented themselves in, say, the early 1960's would reveal quite a different population from the one with which philosophers are now primarily concerned. To begin with, the central preoccupations of the discipline were then largely ontological. It is not quite accurate to characterize inquiry as having been directed to the question: 'What are mental states and processes?'; philosophers of physicalist persuasion (i.e., adherents of the central-state identity theory; see below) took it that the nature of the mental was an empirical issue, hence not one for philosophers to solve. But it was widely held that philosophers ought to provide a survey of the conceptually coherent options, and that there are, in fact, fewer of these than might be supposed. It was, in particular, the rejection of Cartesian dualism and the consequent need to work out a philosophically acceptable version of materialistic monism that provided the main preoccupation of philosophers of mind at the beginning of the period these essays subtend.

The stumbling block for a dualist is the problem of mind/body interaction. If you think, as Descartes did, that minds are immaterial substances, you ought to be worried (as Descartes was) about how there can be mental causes of behavioral effects. Cynicism might suggest that the question how an immaterial substance could contribute to the etiology of physical events is not, after all, much more obscure than the question how material substances could interact causally with one another. But there is this difference: whereas we can produce lots of untendentious examples of the latter kind of interaction, there are no untendentious examples of the former kind. Physical causation we will have to live with; but non-physical causation might be an artifact of the immaterialist construal of the claim that there are minds.

Viewed this way, the issue is one of ontological parsimony. A Cartesian dualist is going to hold that there are species of causal relations over and above physical interactions. He is therefore in need of an argument why mind/body causation should not itself be viewed as an instance of physical interaction. Most philosophers now agree that no such argument has successfully been made. Even philosophers like Ryle, whose preferred style of anti-Cartesian argument was so very often epistemological, insisted on the force of such considerations: "there was from the beginning felt to be a major theoretical difficulty in explaining how minds can influence and be influenced by bodies. How can a mental process, such as willing, cause spatial movements like movements of the tongue? How can a physical change in the optic nerve have among its effects a mind's perception of a flash of light?" (1949, p. 19).

It is precisely the advantage of materialistic monism that it provides for the subsumption of mind/body interaction as a special case of physical interaction, thereby making problems like the ones that Ryle mentions go away. By the early 1960s, it was becoming clear that there are two quite different strategies for performing this reduction, one corresponding to the program of *logical behaviorism* and the other corresponding to the program of the *central state identity theory*. And it was also becoming clear that each of these strategies has its problems.

The essence of one kind of logical behaviorism is the idea that every truth-valuable ascription of a mental state or process to an organism is semantically equivalent to the ascription of a certain sort of dispositional property to that organism. ... In particular, in the interesting

cases, mental ascriptions were supposed to be semantically equivalent to ascriptions of *behavioral dispositions*. A behavioral disposition is one that an organism has if and only if (iff) it satisfies a certain indefinite (possibly infinite) set of *behavioral hypotheticals* which constitute the *analysis* of the disposition. A behavioral hypothetical is a (possibly counterfactual) statement whose antecedent is couched solely in terms of *stimulus parameters* and whose consequent is couched solely in terms of *response parameters*. Heaven only knows what stimulus and response parameters were supposed to be; but it was an Article of Faith that these notions could, in principle, be made clear. Perhaps stimulus and response parameters are species of physical parameters (parameters that would be acknowledged by an ideally completed physical theory), though not all logical behaviorists would have accepted that identification. In any event, precisely because the ontological impulse of behaviorism was reductionistic, success depended on the possibility of expressing stimulus and response parameters in a vocabulary which contained no mental terms.

What is attractive about logical behaviorism is that the proposed identification of mental properties with dispositional ones provides for a sort of construal of statements that attribute behavioral effects to mental causes—a construal that is, moreover, conformable to the requirements of materialistic monism. Roughly, mental causation is the manifestation of a behavioral disposition; it's what you get when an organism has such a disposition and the antecedent of a behavioral hypothetical that belongs to the analysis of the disposition happens to be true. It is, no doubt, a travesty to say that for a logical behaviorist "Smith is thirsty" means "if there were water around, Smith would drink it" and "Smith drank because he was thirsty" means "if there were water around Smith would drink it; and there was water around." But it is a travesty that comes close enough for our present purposes. It allows us to see how logical behaviorists proposed to assimilate worrisome etiologies that invoke mental causes to relatively untendentious etiologies like "it broke because it was fragile; it bent because it was pliable," etc.

So, logical behaviorism provides a construal of mental causation, and the glaring question is whether the construal it provides is adequately robust to do the jobs that need doing. In the long run this is the question whether the identification of mental properties with behavioral dispositions yields a notion of mental causation that is rich enough to reconstruct the etiologies propounded by our best psychological theories: the ones that achieve simplicity, explanatory power, and predictive success. However, we needn't wait for the long run, since there are plenty of cases of *pre*-theoretically plausible etiologies for which plausible behavioristic construals are not forthcoming, and these, surely, are straws in the wind.

Here is a quick way of making the point. Suppose "John took aspirin because he had a headache" is true iff conjunction *C* holds:

C: John was disposed to produce headache behaviors and being disposed to produce headache behaviors involves satisfying the hypothetical *if there are aspirin around, one takes some*, and there were aspirin around.

So, *C* gives us a construal of "John took aspirin because he had a headache." But consider that we are also in want of a construal of statements like "John was disposed to produce headache behaviors because he had a headache." Such statements *also* invoke mental causes and, pre-theoretically at least, we have no reason to doubt that many of them are true. Yet in these cases it seems unlikely that the putative mental causes can be traded for dispositions; or if they can, the line of analysis we've been pursuing doesn't show us how to do it. We cannot, for example, translate "John had a headache" as "John had a disposition to produce headache behaviors" in these cases, since, patently, "John was disposed to produce headache behaviors because he had a headache" doesn't mean the

same as "John was disposed to produce headache behaviors because he had a disposition to produce headache behaviors." Yet, "John had a headache" surely means the same in "John took aspirin because he had a headache" and in "John was disposed to produce headache behaviors because he had a headache," so if the behavioristic analysis is wrong for the second case, it must also be wrong for the first.

This is, no doubt, a relatively technical kind of difficulty, but it points in the direction of what is now widely viewed as a hopeless problem for logical behaviorism. Mental causes typically have their overt effects *in virtue of their interactions with one another*, and behaviorism provides no satisfactory analysis of statements that articulate such interactions. Statements that attribute behavioral dispositions to mental causes are of this species, but they are far from exhausting the kind. Consider that having a headache is sufficient to cause a disposition to take aspirin only if it is accompanied by a battery of other mental states: the desire to be rid of the headache; the belief that aspirin exists; the belief that taking aspirin leads to a reduction of headache; the belief that its side effects are not worse than headaches are; the belief that it is not grossly immoral to take aspirin when one has a headache; and so on. Moreover, such beliefs and utilities must be, as it were, *operative*; not only must one *have* them, but some of them must come into play as causal agents contributing to the production of the behavioral effect.

The consequences of this observation are twofold. First, it seems highly unlikely that mental causes can be identified with *behavioral* dispositions, since the antecedents of many putatively behavioral hypotheticals turn out to contain mentalistic vocabulary ineliminably. But, moreover, we shall have to find analyses for etiologies in which interactions among mental states involve possibly quite long and elaborate causal chains; and here we have a fertile source of counterexamples to the generality of *any* kind of dispositional construal of mental ascriptions, behavioral or otherwise. "It occurred to John to move his knight, but then he noticed that that would leave his king in trouble, and since he wanted to avoid check, he thought on balance that he'd better temporize and push a pawn. When, however, he had considered for a while which pawn to push, it all began to look boring and hopeless, and he thought: 'Oh, bother it,' and decided to resign. (Overt behavior finally ensues.)" Surely, the mental life is often like that; such cases occur everywhere where mental *processes* are at issue, and it is, perhaps, the basic problem with behaviorism that it can't reconstruct the notion *mental process*. But these cases seem to be most glaringly the norm in reasoning and problem solving, so it's not surprising that, among psychologists, it has been the cognition theorists who have led the recent antibehaviorist campaign. It seems perfectly obvious that what's needed to construe cognitive processes is precisely what behaviorists proposed to do without: causal sequences of mental episodes and a "mental mechanics" to articulate the generalizations that such sequences instantiate. The problem was—and remains—to accommodate these methodological requirements within the ontological framework of materialism.

Which is why the physicalist reading of materialistic monism seemed so very attractive an alternative to the behavioristic version. Suppose we assume that mental particulars (events, states, processes, dispositions, etc.) are identical with physical particulars, and also that the property of being in a certain mental state (such as having a headache or believing that it will rain) is identical with the property of being in a certain physical (e.g. neural) state. Then we have a guaranty that our notion of mental causation will be as robust as our notion of physical causation, the former having turned out to be a special case of the latter. In particular, we will have no difficulty in making sense of the claim that behavioral effects may sometimes be

the consequence of elaborate chains of mental causes; or, indeed, that mental causes may interact elaborately *without* eventuating in behavioral effects. If mental processes are, in fact, physical processes, they must enjoy whatever ontological autonomy physical processes are conceded. More than that we surely should not ask for.

This is bracing stuff from the point of view of the psychologist. Endorsing the central state identity theory is tantamount to accepting a Realist interpretation of explanations in which appeal to mental causes figure: if mental particulars are physical particulars, then the singular terms in (true) psychological theories denote. Indeed, they denote precisely the same sorts of things that the singular terms in (true) physical theories do. Moreover, since physicalism is not a semantic thesis, it is immune to many of the kinds of arguments that make trouble for behaviorists. It is, for example, uninteresting from the physicalist's point of view that "John has a headache" doesn't mean the same as "John is in such and such a brain state." The physicalist's claim is not that such sentences are synonymous—that the second provides a linguistically or epistemologically adequate construal of the first—but only that they are rendered true (or false) by the same states of affairs.

I remarked that the identity theory can be held either as a doctrine about mental *particulars* (John's current headache, Bill's fear of cats) or as a doctrine about the nature of mental *properties* (universals like having a pain or being afraid of animals). These two variants—known as "token physicalism" and "type physicalism" respectively—differ in both strength and plausibility. For, while the former claims only that all the mental particulars that there happen to be are neurological, the latter makes that claim about all the mental particulars that there *could* be. Token physicalism is thus compatible with the logical—perhaps even the nomological—possibility of unincarnate bearers of mental properties (e.g. angels) and of bearers of mental properties that are incarnate but not in flesh (e.g. machines). Whereas, type physicalism does not recognize such possibilities, since, if the property of having a pain is the same property as that of being in a certain neural state, nothing can have the former property that does not have the latter.

Type physicalism is, on balance, not a plausible doctrine about the nature of mental properties—not even if token physicalism is a plausible doctrine about the nature of mental particulars; not even if token physicalism is a *true* doctrine about the nature of mental particulars. The argument against type physicalism is often made in a way that does it less than justice: "Why shouldn't (e.g.) silicon-based Martians suffer pains? And why shouldn't machines have beliefs? But if it is conceivable that mental properties should have such bearers, then it cannot be that mental properties *are* neural properties, however much the two may prove to be *de facto* coextensive. And neither can it be that they are *physical* properties if what you mean by a physical property is one that can be expressed by a projectible predicate in (say) an ideally completed physics. What silicon-based Martians and IBM machines and you and I are likely to have in common by way of the physical constitution of our nervous systems simply isn't worth discussing." ...

But that's really not the point. The real point is that, if we want a science of mental phenomena at all, we are required to so identify mental properties that the kinds they subsume are natural from the point of view of psychological theory construction. If, for example, we identify mental properties with neural properties, then we are in effect claiming that domains consisting of creatures with a certain sort of neural organization constitute natural kinds for the psychologist's purposes. (Compare: if we claim that the property of being a fish is the property of living in the water, then we are in effect claiming that a domain consisting of creatures that live in the

water constitutes a natural kind for the purposes of the marine biologist. Either that or we are claiming that "is a fish" is not a projectible predicate in marine biology. The essays that follow are neutral on fish, but they do assume the projectibility of properties like those expressed by typical mental predicates.)

Now, there is a level of abstraction at which the generalizations of psychology are most naturally pitched and, as things appear to be turning out, that level of abstraction collapses across the differences between physically quite different kinds of systems. Given the sorts of things we need to say about having pains and believing Ps, it seems to be at best just accidental, and at worst just false, that pains and beliefs are proprietary to creatures like us; if we *wanted* to restrict the domains of our psychological theories to just us, we would have to do so by ad hoc conditions upon their generalizations. Whereas, what does seem to provide a natural domain for psychological theorizing, at least in cognitive psychology, is something like the set of (real and possible) information processing systems. The point being, of course, that there are possible—and, for all we know, real—information processing systems which share our psychology (instantiate its generalizations) but do not share our physical organization.

It would be hard to overemphasize this point, but I shall do my best: Philosophical theories about the nature of mental properties carry empirical commitments about the appropriate domains for psychological generalizations. It is therefore an argument against such a theory if it carves things up in ways that prohibit stating such psychological generalizations as there are to state. And it looks as though type physicalism does carve things up in the wrong ways, assuming that the sort of psychological theories that are now being developed are even close to being true.

This is a state of affairs which cries out for a *relational* treatment of mental properties, one which identifies them in ways that abstract from the physiology of their bearers. Indeed, there is a sense in which behaviorists had the right end of the stick in hand, despite the sorts of objections we reviewed above. Behaviorists, after all, *did* offer a relational construal of mental properties: to have a belief or a pain was to be disposed to exhibit a certain pattern of relations between the responses that one produces and the stimuli that one encounters. So there seemed, ten or fifteen years ago, to be a nasty dilemma facing the materialist program in the philosophy of mind: What central state physicalists seemed to have got right—contra behaviorists—was the ontological autonomy of mental particulars and, of a piece with this, the causal character of mind/body interactions. Whereas, what the behaviorists seemed to have got right—contra the identity theory—was the relational character of mental properties. Functionalism, grounded in the machine analogy, seemed to be able to get both right at once. It was in the cheerful light of that promised synthesis that the now dominant approaches to the philosophy of mind first began to emerge.

It's implicit in the preceding that there is a way of understanding the new functionalism that makes it simply an extension of the logical behaviorist's program. Having a belief (say) was still to be construed as having a relational property, except that, whereas behaviorists had permitted only references to stimuli and responses to appear essentially in specifications of the relata, functionalists allowed reference to *other mental states* as well. A functionalist could thus concede many platitudes that behaviorists are hard put even to construe—as, for example, that it is plausibly a necessary condition for having certain beliefs that one is disposed to draw certain inferences (no belief that P&Q without a disposition to infer that P and a disposition to infer that Q); that it is plausibly a necessary condition for performing certain acts that one be in certain states of mind (no prevarication without the causal involvement of an intent to deceive); and so forth.

But, in fact, reading functionalism this way —as a liberated form of behaviorism—is more than mildly perverse; for *all* that functionalism and behaviorism have in common is the relational construal of mental properties. If one considers the ontological issues (to say nothing of the epistemological issues), striking differences between the doctrines emerge at once. Unlike behaviorism, functionalism is not a reductionist thesis; it does not envision—even in principle—the elimination of mentalistic concepts from the explanatory apparatus of psychological theories. In consequence, functionalism is compatible with a physicalist (hence Realist) account of mental particulars. It thus tolerates the ascription of causal roles to mental particulars and can take literally theories which represent mental particulars as interacting causally, with whatever degree of complexity, in the course of mental processes. That, as we have seen, was the primary advantage the central state identity theory enjoyed. So it is possible for a functionalist to hold (1) that mental kinds are typically relationally defined; and (2) that the relations that mental particulars exhibit, insofar as they constitute the domains of mental processes, are typically *causal* relations: to stimuli, to responses, and to one another. Of these claims, a logical behaviorist can endorse only the first with an absolutely clear conscience, and a type-physicalist only the second.

So functionalism appeared to capture the best features of the previously available alternatives to dualism while avoiding all the major embarrassments. Viewed from one perspective, it offered behaviorism without reductionism; viewed from another, it offered physicalism without parochialism. The idea that mental particulars are physical, the idea that mental kinds are relational, the idea that mental processes are causal, and the idea that there could, at least in logical principle, be nonbiological bearers of mental properties were all harmonized. And the *seriousness* of the psychologist's undertaking was vindicated by providing for a Realistic interpretation of his theoretical constructs. Bliss!

REFERENCES

Ryle, G. (1949). *The Concept of Mind*, New York, Barnes and Noble.

Jerry A. Fodor and Ned Block

Selections from "What Psychological States Are Not"

The following argument seems to us to show that the psychological states of organisms cannot be placed in one-to-one correspondence with the machine table states of organisms.

The set of states that constitute the machine table of a probabilistic automaton is, by definition, a list. But the set of mental states of at least some organisms (namely, persons) is, in point of empirical fact, productive. In particular, abstracting from theoretically irrelevant limitations imposed by memory and mortality, there are infinitely many type-distinct, nomologically possible psychological states of any given person. The simplest demonstration that this is true is that, on the assumption that there are infinitely many non-equivalent declarative sentences, one can generate definite descriptions of such states by replacing *S* with sentences in the schemata *A*:

A: the belief (thought, desire, hope, and so forth) that *S*.

In short, while the set of machine table states of a Turing machine can, by definition, be exhaustively specified by listing them, the set of