Chapter 7

Adaptational Role

Exposition

In an important series of publications, Ruth Millikan has offered a subtle and complex account of representation in terms of the adaptational role of symbols and the adaptational roles of the mechanisms that produce them and respond to ("interpret") them. (See especially Millikan 1984, 1986; see also Papineau 1984.) Millikan's treatment resists easy summary: I cannot, in the course of a single chapter, hope to do full justice to the theory. Instead, I will concentrate on the basics, illustrated in connection with what Millikan takes to be the most fundamental type of case. My purpose, as in my chapters on Fodor and Dretske, is the limited one of giving enough of the flavor of the theory to determine whether it is suitable as an explication of the concept of representation appealed to in the CTC.[1]

Here is the fundamental formulation:

> (M1) *C* is a truth condition for *r* in $S =_{df}$ *C*'s obtaining is a *basic factor* in a Normal Case for the performance of the Proper Functions of *r*- interpreters.[2]

Something *x* performs a Proper Function in a system *S* when it does the sort of thing the doing of which has been, historically, responsible for the replication of things of *x*'s type.[3] Thus, circulation of the blood is a Proper Function of hearts because it is the fact that hearts contribute to blood circulation in the way they do that has been, historically, responsible for the replication of hearts, and hence for the historical persistence of that type of organ.[4]

A basic factor[5] for the performance of a Proper Function is a factor that tracks the selective importance of the Normal Case of that function fulfilling its Proper Function.[6] A Normal Case is one in which the function is performed *successfully*—i.e., a case in which the item in question does its stuff and everything else conspires to produce the kind of result that is responsible for that item's replication.

Millikan's suggestive illustration is the bee dance. When a bee dance is performed properly, it has a certain *orientation*. A consequence of a properly performed bee dance is that spectator bees fly off in a direction that corresponds to the orientation of the dance. But the whole business is a *success* (i.e., comes off in the way that accounts for the replication of the relevant mechanisms) only if the bees find flowers in that direction (hence find pollen, hence produce nectar, hence honey, hence food, hence the means of survival). So a Normal Case—which need not be the *usual* case at all—is a case in which spectators respond to a dance by flying off in the direction that corresponds to the dance's orientation and find flowers as a result. Flowers' being in the relevant direction is, therefore, a *basic factor* for the dance interpreter's functioning Properly in response to a dance: no flowers, no pollen; no pollen, no food; no food, no reinforcement (via natural selection) for the mechanisms that respond to bee dances.

At first glance, it might seem that the Normal Case is loaded with *basic factors*: There has to be enough light, the wind can't be blowing too hard, the flowers can't be deadly to bees, there can't be bee predators or clouds of DDT in the relevant direction, and so on. What, one might wonder, makes it the case that the content of the bee dance is *flowers in the o-direction* (i.e., flowers in the direction corresponding to the orientation of the dance) rather than, say, *no predators in the o-direction*?

The answer is that the o-direction's being a predator-free zone is not part of the content of a bee dance because it is not basic to the Normal Case; rather, it is part of the background or boundary conditions of such cases.[7] In Normal Cases there are no predators, no nuclear wars, no hurricanes, no poisonous flowers, and so forth, and there is plenty of light and air, the earth moves in the

usual way relative to the sun, etc. But although these are all necessary conditions for Normal interpreter performance, they are not basic factors; flower presence in the o-direction is the (or a) basic factor.

What makes the presence of flowers in the o-direction a basic factor, and the presence of light and the absence of predators a background condition? The idea is that it is the presence of flowers, and not the absence of predators, that tracks the selective significance of Normal Cases of bee-dance-interpreter function. Although this is clear enough intuitively, the notion of *basic factors* operative here has application outside of a selectionist framework. I will digress briefly to indicate what this application is because it has some independent interest, and because seeing it at work in a nonselectionist context helps to clarify what is going on in the selectionist contexts that are central to Millikan's account.

Basic Factors

Consider the simple pendulum law:

$$2\pi\sqrt{l/g} = T$$

This is an idealization; it captures only the contribution of length and gravity to period, ignoring friction and air resistance. It works because of two facts. First, length and gravity are independent of friction and air resistance; g and l don't change as a function of friction or air resistance. The idealization represented by equation 7.1 wouldn't work if increasing the friction shortened the length. Second, length and gravity are "basic factors" in this situation: One can begin with equation 7.1 and superimpose the effects of friction and air resistance, but one cannot begin with the contribution of friction and air resistance to period and then superimpose the contribution of length and gravity. *There is no such thing as the way a pendulum would behave were length not a factor, whereas there is such a thing as the way a pendulum would behave were air resistance and friction not factors.* Experimentally, the point is that it makes sense to reduce friction and air resistance and note whether the period beings to approximate

$2\pi\sqrt{l/g}$; we cannot manipulate length and note whether the period begins to approximate what it would be if friction, air resistance, and gravity were the only factors, for there is nothing to approximate.

The Normal Case for the functioning of the bee-dance interpreter is supposed to be an idealization in the same sense in which the simple pendulum law is an idealization: It focuses on the basic factors—the factors responsible for the basic phenomenon—and it ignores factors not constitutive of that basic phenomenon. In the Normal Case, the spectators react to the dance by flying off in the direction of the dance orientation. They find flowers. They gather pollen. They return to the hive and make honey. We don't mention the fact that clouds of DDT are not encountered, the fact that the wind doesn't blow them off course, that night doesn't fall, that the distance isn't so great that they drop from exhaustion, and so on, and so on. These are all relevant factors, but they aren't basic; they don't *ground* the phenomenon. There is an asymmetry between the role played by the flowers and that played by the absence of DDT. Although both are necessary conditions for success, the dance interpreter doesn't get replicated because of the cases in which spectators fail to encounter DDT, for there were presumably lots of such cases in which there were no flowers. Hence, "Spectators respond to the dance by flying off in the direction of the dance orientation, where they find no DDT and hence survive" does not describe a Normal Case in a way that specifies what makes Normal Cases of dance-interpreter function *Normal*.[8]

It is interesting to examine how Millikan's theory treats the case of the magnetotaxic bacteria discussed by Dretske. The problem, recall, is how to arbitrate between the liberal and conservative construals of the representational significance of magnetosome orientation. Liberals hold that the function of magnetosome orientation is to indicate the direction of safe water; conservatives hold that its function is to indicate the orientation of the magnetic field.

Millikan's theory comes down squarely on the side of liberalism, for the Normal Case is surely the case in which the locomo-

tion mechanisms respond to magnetosome movement by propelling the organism away from the surface to safe (oxygen-free) water. Hence, the relevant *basic factor* is the presence of safe water in the o-direction (*viz.* down, toward geomagnetic north), for it is the role played by the interpreter—the locomotion mechanism, in this case—in getting the organism to safe water that is responsible for the replication of the interpreter (the locomotion mechanisms).

But why not say "The magnetosome indicates the direction of magnetic north, and the interpreter acts on the rule *Safe water in the direction of magnetic north*"? Because magnetic north's being in the o-direction is not a basic factor in the interpreter's (the locomotion system's) performing its function of getting the organism to safe water.[9] The interpreter was replicated because, in the Normal Case, it responded to magnetosome orientation in a way that led the organism down to safe water, not because it led the organism in the direction of magnetic north. From the point of view of selection, the fundamental connection is the one between magnetosome orientation and the direction of safe water, and it is the interpreter's role in mediating that connection that has led to to its replication.[10]

Evaluation

Despite the considerable merits of Millikan's account, it will not do as an account of the concept of representation required by the CTC. The reason is simple: The CTC is committed to an ahistorical notion of representation, and Millikan's notion is essentially historical.

Duplicates

Imagine a kind of duplicating machine that duplicates organisms—not by cloning, or by any other biochemical process that uses the information coded in the organism's DNA, but just as a copy machine duplicates a printed page without understanding it. The machine I have in mind produces a perfect physical duplicate of an organism without "understanding it." The "transporter" familiar to fans of *Star Trek* is presumably such a device,

although it destroys the original in the process of producing a duplicate in another place.

Will a duplicate of an organism have the same representations as the original? The intuition is fairly widespread that it will. Indeed, the assumption behind the *Star Trek* transporter is that the duplicate is the same person who entered the transporter. There seems little doubt that, for the purposes of a psychology experiment, a molecule-by-molecule duplicate of a person would do as well as the original. To deny this seems to be to deny physicalism. (See the example attributed to Block by Fodor [1987, p. 40].)

The CTC goes farther than simple physicalism; it asserts that, in order to preserve the identity of a cognitive system (if not a whole mind or person), it suffices to produce a computational duplicate. Two systems running the same program on the same data structures are, according to the CTC, cognitively indistinguishable. Since physical duplicates are bound to be computational duplicates, the idea that physical duplicates must share cognitive states is a consequence of CTC. Computationalism holds that cognitive systems are automatic formal systems. It follows that any two automations of the same form system are the same cognitive system. However, a perfect physical duplicate of *D* will automate a formal system *F* if and only if *D* automates *F*.

An exactly parallel argument will work for any theory that holds that a cognitive system can be specified in a way that abstracts from its physical realization (e.g., connectionist theories), for to abstract from physical realization is to abstract from the history of those realizations. Current ahistorical state, according to this approach, determines current cognitive capacities, and hence must determine current representational content. The CTC abstracts from history for the same reason it abstracts from the actual items in a system's current environment: For better or worse, the CTC seeks a theory of cognitive capacities of the sort that might be brought to bear on radically different environments (with differing success, no doubt), and that might be realized in radically different stuff.

A conspicuous feature of Millikan's account is that perfect

physical duplicates need not have the same representations. This is an immediate consequence of the fact that representational content is essentially a matter of *history* on Millikan's view. A perfect physical duplicate of me will have no evolutionary history or learning history at all. It will make no sense to ask of any of the mechanisms or subsystems of such a duplicate what factors were responsible for their replication: they weren't replicated in the sense required by Millikan's account.

It follows immediately that Millikan's account cannot be imported to explicate representation as that concept figures as an explanatory construct in computationalist theories of cognition. To put it crudely: According to computationalist accounts, history is an accidental property of a cognitive mechanism. According to computationalism, cognitive systems are individuated by their computational properties, and these are independent of history.

Someone with Millikanite leanings might be tempted to say that duplicate is cognitively equivalent to its original on the ground that it duplicates something with the right history. But this, in fact, amounts to abandoning the Millikan approach at a very fundamental level, for it amounts to conceding that it is really current state, ahistorically conceived, that underwrites cognitive content. What duplication produces, after all, is something that satisfies the same molecular blueprint as the original. If that is enough to ensure cognitive equivalence, then the significance of the original's history can only be that it resulted in an organism with the right molecular blueprint. History ceases to be of the essence, lapsing into the role of the only available technology for producing, without a prototype to copy, a system with the right molecular structure.

Another desperate attempt to somehow capitalize on Millikan's theory in order to give an account of representation that is consistent with computationalism is to think of the relevant history as somehow imminent in synchronically specified computational states. According to this idea, only when we think of the current state of a system as the product of a certain possible history does it make sense to characterize that system representationally. Although I think this idea has some merit of its own,[11] it is certainly no variation on Millikan, for it will surely be possible to describe various different histories that might have led to the same molecular blueprint or computational architecture—histories that will license quite different content attributions.[12] On Millikan's view, the point of appealing to actual history is to eliminate this kind of ambiguity (or indeterminacy, if you like); *actual* history determines the real content from among the possible ones.

A slightly more interesting possibility is that a physical duplicate would share the original's representations on the ground that selectionist/adaptationist explanations of the mechanisms in the original do carry over to the duplicate. The idea would be to claim, for example, that my duplicate's heart has the same Proper Functions as mine on the grounds that (i) the duplicate system is the way it is because *mine* is the way it is and (ii) my heart *does* have blood circulation as a Proper Function.

There is a good deal to be said for this suggestion (which I owe to Clayton Lewis). After all, genetic replication mechanisms are no more sensitive to selection history than the *Star Trek* transporter. Both mechanisms simply take things as they find them and replicate them, not on the basis of functional properties, but on the basis of physical properties on which functional properties supervene. It is hard to see, therefore, how to tell a story that will allow genetic replication to preserve Proper Functions but will not allow transporter replication to preserve them as well.[13]

Plausible as this line is, it won't help prop up a Millikanite reading of representation in the CTC. To repeat, the crucial issue is whether representation is grounded in the current state of the system, regardless of the history of that system. And on that issue the CTC is absolutely unambiguous: Computationally equivalent states are representationally equivalent. Yet computationally equivalent states need share no historical properties at all. From a computational perspective, historical properties are accidental properties. This is why computationalists can coherently hope to bypass the learning process. The CTC entails that if we *give* a system the same data structures that a natural system must

acquire by learning, then (barring differences in computational architecture) we have a cognitively equivalent system. No computationalist can consistently suppose that a system that knows English must have *learned* it. And what goes for learning obviously goes for evolution as well. It is which data structures you have, not how you got them, that counts. Without this assumption, AI makes no sense at all.[14] No account that (like Millikan's) takes the history of a data structure seriously can be right for the CTC.

History and Belief

It is worth digressing a moment to point out that beliefs *are* individuated in a way that is sensitive to history. To borrow an example from Stich (1983), the belief my duplicate expresses with the words "I sold my car for a thousand dollars" is false because he didn't own the car in question, whereas the belief I express with the same words is true. We therefore have different beliefs, though by hypothesis we are computationally (indeed physically) equivalent. The problem is that my duplicate never acquired title to the car in question; I did. Hence, history matters to belief contents. If you are interested in belief contents, then you will do well to formulate an account that is sensitive to historical properties.

It doesn't follow from this, of course, that *representation* is sensitive to historical properties. In fact, nearly the opposite follows: Since data structures aren't sensitive to historical properties, it follows that beliefs aren't data structures. Moreover, it follows that beliefs don't inherit their contents from constituent data structures, as the RTI claims. This is one reason why philosophers who are interested in the semantics of belief (and of 'believes') are bound to misrepresent representation in the CTC: Data structures are insensitive to all sorts of things—such as historical properties—to which beliefs (and the other propositional attitudes) are exquisitely tuned. There is a good reason for this. As Stich points out, the CTC doesn't want to explain why my duplicate can't sell my car. Or, to put it as Ned Block does (in the example cited by Fodor, quoted above), some differences in belief are not legitimate sources of psychological variance.

Conclusion

Millikan's theory is sophisticated and complex. I have had to ignore most of the sophistication and complexity, for it lies aside from my main concern here, which is to discover what representation must be if the CTC is to be true and explanatory. The CTC entails that history is irrelevant to content, and Millikan's theory says that history is the very essence of the thing. This is no knock on Millikan; she was after intentionality (belief, etc.), not representation in the CTC.

The Evidential Value of Adaptational Role

Before going on to other possibilities, however, we should take note of a final point. When I introduced covariance theories in chapter 1, I suggested that one might be moved to ground representation in covariance because one discovers that a certain neural structure is an edge detector by noticing that it fires when and only when there is an edge in the organism's visual field. Obviously, an adequate treatment of representation should account for the evidential role of covariance. Should we also insist that an adequate treatment of representation account for the evidential role of selective history?

The problem with this suggestion is that we seldom *know* the selective history of anything. Or, rather, we seldom have knowledge of selective history that is epistemologically prior to the very facts about function and representation we are interested in establishing. We don't begin with the selective history of the bee-dance mechanisms and then infer the content of bee dances; we exploit covariance to infer a probable content and then *reconstruct* the selective history. Of course, the availability of a plausible reconstruction in a case of this kind is powerful corroboration. The discovery that bees find flowers when they fly off in response to bee dances, together with our knowledge that they *need* to find flowers, leaves us feeling satisfied that the content of the dance is *flowers in the direction of the dance orientation*. Nevertheless, if it always happened that spectator bees flew off in the direction of the dance orientation, found a pile of rocks, milled around, then went home, we would, I think, be justified in attributing the

content *rocks in the direction of the dance orientation*, even though we would be mystified about the evolutionary significance of the whole business. It seems pretty clear that speculation about the evolutionary history (and even learning history) of central cognitive mechanisms will be possible only after we have a pretty good idea what representations are actually required. I don't see how we can hope to understand the adaptational significance of the abstract functional architecture of the brain until we know what cognitive capacities it underwrites. But to know that, we must traffic heavily in mental representation. Epistemologically, then, representation is prior to adaptational role. Moreover, and much more important, the explanatory order follows the epistemological order in this case. We can't explain the adaptational role of a cognitive capacity without presupposing mental representation, for a capacity is *cognitive* (and the particular cognitive capacity it is) only in virtue of its semantic characterization.

Those who hope to explain representation in terms of adaptational role, then, face a dilemma reminiscent of that faced by the Lockean covariance theorist. The adaptational significance of brain mechanisms is surely tracked by their cognitive significance to a large extent. If one wants to explain the adaptational significance of the brain, then, one must be in a position to specify the cognitive architecture of the brain. The adaptational significance of the brain presupposes its cognitive capacities, and (according to the CTC, at least) cognitive capacities rest on representational capacities. This order of explanation is undermined if representation is explained (defined) in terms of adaptational role. Hence the following dilemma: If one wants to explain (define) representation in terms of adaptational role, one cannot also explain the adaptational role of brain mechanisms in terms of its representational capacities, as the CTC proposes to do.

Chapter 8

Interpretational Semantics

Summary and Advertisement

A central insight of the seventeenth century was that mental meaning cannot be understood in terms of resemblance. If the semantic relations between mind and world cannot be understood on the hypothesis that the mind is *like* the world, literally sharing properties with the things it represents, how can it be understood? In the hands of Locke and his successors, covariance replaced resemblance. But whatever advantages this had for Locke, mental representation cannot be understood in terms of covariation by those who want to follow the CTC in supposing that mental representation *explains* how systems manage to get into states that covary with the states of the world. And the attempt to understand mental representation in terms of adaptational roles also appears to reverse the explanatory order central to the CTC, and to be inconsistent with the thesis that cognition (and hence representation) supervenes on abstract formal structure that need not be historically specified.

What is next? The demise of similarity, idealized covariance, and adaptationist theories leaves us with only one candidate: functional role. The approach I will sketch in this chapter—"Interpretational Semantics"—turns out to be ontologically equivalent to a kind of functional-role semantics. However, Interpretational Semantics has, to my intellectual palate, a very different flavor than functional-role semantics; it is motivated in a very different way than typical functional-role theories, it directs our attention to quite different issues, and it generates a