CHAPTER NINE

# Everybody Has Got It: A Defense of Non-Reductive Materialism

*Louise Antony*

*Miss Adelaide*: What is? Oh, the book. Yeah. The doctor gave it to me. He said it might help me get rid of my cold.
*Nathan Detroit*: With a book?
*Miss Adelaide*: He thinks that my cold might possibly be caused by psychology.
*Nathan Detroit*: How does he know you got psychology?
*Miss Adelaide*: Nathan! Everybody has got it.

J. L. Mankiewicz, 1955, *Guys and Dolls*, screenplay

It is a really striking fact about human beings that we think. Just this morning, for example, I deliberated about what to have for breakfast, wondered if I should let my husband sleep in, noticed the dogs were almost out of food, figured out where my favorite mug was, vowed to write a letter to the editor, imagined how nice a sweater I could make out of that lovely Australian wool, remembered I had to prepare for my seminar today, and wished I didn't have to prepare for my seminar today. That was all before 9 a.m.

Sometimes thinking is more spectacular. Human beings have done amazing things through thinking. They have written epic poetry, discovered laws of nature, navigated seas, composed symphonies, designed buildings, invented machines, and cured disease. Many (if not all) of these accomplishments involved not only thinking, but thinking about thinking, and thinking about what other people were thinking. Also involved was talking, which (at least when some people do it) seems to involve thinking, and, of course, understanding when other people were talking. The invention of a way of capturing talking in a less ephemeral form – writing – was a spectacular use of thinking, and led to even more opportunities for thinking, and for doing all the other things one can do through thinking.

We think that other people think. Thinking this works out really well. By attributing thoughts to other people, we are able to predict and explain their behavior in myriad and immensely useful ways. When I drive my car, for example, I think that the other people driving their cars know much the same things I do about the rules of the road and the basic properties of automobiles. I also think that they, like me,

wish to reach their destinations safely. Thinking all this, I drive sanguinely through an intersection marked by a green light. On the car ahead of me to my left, a signal light starts flashing; I take this to mean that the driver intends to pull into my lane ahead of me, and I decelerate slightly to accommodate her.

Of course, sometimes I am wrong about what I think others are thinking. I assume the person to my left knows that I have the right of way at the four-way stop, but when he begins to pull into the intersection out of turn, I revise my view, and come to believe that he believes that he has the right of way instead of me. His believing this would explain his behavior. I honk at him, trusting that he will understand from my honking that I think he is a jerk. He makes a small, conventional gesture with his right hand that I understand to mean that he wishes me to know that the feeling is mutual. And so it goes.

I hope you're finding this a bit boring. What I've been trying to do is hammer home the banality of the claim that human beings possess psychologies and that our psychologies are centrally involved in virtually everything we do, from our most sublime accomplishments to our most ridiculous gaffes. There is, however, no banality so banal that no philosopher will deny it, and many, many philosophers have denied that we have psychologies. Indeed, a cursory survey of the past century's work in the "philosophy of mind" might leave the impression that this is a discipline dedicated to the eradication of its own subject matter. There are two ways to deny that we have psychologies: one can either say "there are no minds," or one can say "there are no minds *as such*." The first group of naysayers are called "eliminativists," and the second are called "reductionists." It can be a little difficult to tell the difference.

Why would anyone deny the existence of thinking? That is indeed the question. What mind-deniers will tell you is that, one way or another, belief in mental things is incompatible with *materialism*. Materialism, for our purposes, is the doctrine that Descartes was wrong. Descartes, notoriously, argued that the mind, the *res cogitans*, the thing that thinks, was different in its essential nature from the body, and could even exist separately from it. Materialists deny all *that*. Now on the face of it, one should be able to reject a particular *account* of the mind – Descartes's – without having to give up the mind itself. I reject, after all, the view that the moon is made of green cheese, but I'm still pretty confident it's up there. Mind-deniers, however, think the distinction I have in mind cannot be made in this case – that any notion of the mental is bound, one way or another, to implicate us in some problematic form of dualism.

Eliminativists think that belief in mentality is incompatible with a robustly naturalistic view of the human organism. Eliminativists are unimpressed with either the ubiquity or the utility of psychological ascription. According to them, the informal psychologizing bruited above bespeaks what is essentially a pre-scientific "folk" theory, akin to vitalistic theories of life and supernatural theories of disease. Just as biology has obviated entelechies and witches, the maturing sciences of the brain will soon relieve us of the need for beliefs and desires, hopes or fears, pleasures or pains (Churchland, 1981, and EVOLVING FORTUNES OF ELIMINATIVE MATERIALISM). There is no hope, they'll argue, of folk psychology's simply being subsumed by a more precise science, as Newtonian mechanics was subsumed by relativistic physics, because the

taxonomy implicit in the folk theory is incommensurable with the taxonomies of the serious, well-established sciences (Stitch, 1983; Bickle, 2003). It may even be incoherent (Quine, 1960).[1]

Reductionists are a bit more charitable than eliminativists are toward the posits of common-sense psychology. Reductionists allow that psychological properties, states, processes, and entities exist, but think that these are all reducible to properties, states, processes, or entities of some other type. They deny, in other words, that there are distinctively psychological phenomena and regularities constituting a proprietary domain for a distinctive science of psychology: psychology can be, for these deniers, nothing more than a branch of human (or animal) biology.

The issues at stake here are more arcane than those that divide the non-reductive materialist from the eliminativist. The dispute here centers on a problem that is as old as Cartesian dualism – the problem of mental causation. For Descartes, as for contemporary dualists, the problem was explaining how two substantially different kinds of substance could interact causally. For contemporary materialists, the problem concerns the causal efficacy of mental *properties*. We begin with a principle generally accepted by materialists, *the causal closure of the physical*; it states that all physical events (that have causes at) all have nomologically sufficient physical causes. Mental events, we assume, sometimes cause physical events. If causal closure is true, then those mentally caused physical events must also have physical causes. But in that case, the putative mental causes look to be otiose. If they are not to be shaved off by Occam's razor, it looks as though they must be identified with the physical events that are doing the actual causal work. But if mental events just *are* physical events, then there are no specifically psychological properties at work, and no need for – indeed, no possibility of – a specifically psychological taxonomy or science (Kim, 1998).

Both eliminativists and reductionists, therefore, deny the possibility of a non-reductive materialist theory of mind, each for their different reasons: eliminativists say it's because there can be no theory of mind, period, and reductionists say it's because a theory of mind must really be a theory of (some non-mental) something else. Eliminativists deny the doctrine of *psychological realism*; reductionists, the doctrine of *the autonomy of psychology*. I'm here to tell you that they are wrong. Both of these doctrines are correct; together they constitute the view of mind called *non-reductive materialism*. This is the view that says that (a) there are mental phenomena; (b) they are material in nature; and (c), notwithstanding (b), they form an autonomous domain.

My defense of this view will proceed as follows: first I'll review the arguments in favor of psychological realism, and defend it against the eliminativist challenge. In the course of doing this, I'll take a look back at the failure of the leading eliminativist program of the twentieth century, behaviorism. Reviewing the reasons for this failure will reinforce my prima facie case for the ineliminability of the psychological, but will also help address the reductionist, by illuminating the reasons why, from a scientific standpoint, psychological phenomena must be treated *as* psychological, and hence, as autonomous. Finally, I'll take up the "new" problem of mental causation, which, I'll argue, is an artifact of residual Cartesian thinking.

But first, two preliminary notes.

The debate about non-reductive materialism is, I acknowledged, esoteric – it is an in-house dispute among committed materialists. But I must warn the reader that there is an even more outré dispute on the horizon. Non-reductive materialists do not all agree with each other about exactly what it means to call the mind material. Some (the philosophers I think of as "Neumanians")[2] are willing to stop arguing once it has been shown that psychology is ineliminable, that the descriptions, predictions, and explanations of folk psychology must be taken at face value (Davidson, 1970; Baker, 1995; Burge, 1993). But others of us (and I am in this camp) think that a full defense of psychology requires more – an account of *how* psychology, with all its distinctive features, could be embodied in material beings. Such an account, we contend, requires providing a *reductive explanation* of psychological phenomena. Thus, I intend to defend a version of non-reductive materialism that insists on ontological autonomy for the entities and properties of psychology, while demanding at the same time an *account* of psychological phenomena in terms of non-psychological phenomena.

In what follows, I'll neglect the views of the Neumanians. I do so for two reasons: first, I think that an adequate answer to the eliminativist requires showing how mentality can be instantiated in a physical system, and how the posits of folk psychology can be integrated into a scientific account of the behavior of the human organism. The Neumanians insist that no such "vindication" is needed – that our ordinary experience suffices to establish the reliability of our folk psychological generalizations and explanations. This, to my mind, evinces a confusion between the *epistemic* ground of our acceptance of folk psychology, and the *ontological* constitution of the psychological realm. As I'll argue myself, the evidence for the truth of our psychological ascriptions is overwhelming: that doesn't obviate the scientific impulse that asks what it is about the world that *makes* them true. In any case, if a reductive explanatory account can be provided, as I think it can, then I can see no cogent argument against providing it.

That, however, brings me to my second reason for setting aside Neumanian non-reductive materialism. Neumanians, deep down, don't believe that it is *possible* to give a reductive explanatory account of the truths of psychology. And the reasons they offer come awfully close to the arguments offered by eliminativists against the possibility of a successor science to folk psychology. Neumanians, like many eliminativists, think that the taxonomy implicit in everyday psychologizing is bound to cross-classify with those of biology, mainly because of the intentionality of psychological types. Much of the inspiration for this line of thought comes from Donald Davidson, who argued that mentalistic ascriptions were "governed" by different "constitutive principles" than were claims about the physical world. In particular, Davidson thought that psychological ascriptions had to conform to *normative* demands that were alien to the physical realm. This disparate set of commitments meant, in his view, that there could be no *lawful* connection between the mental and the physical. In my view, Davidson was trying to ward off a certain possibility: "competition" between rational and non-rational evidence about the content of a mental state. For suppose that we had a well-confirmed theory that said that a person's being in brain state 67 is sufficient for that individual's thinking that Helena is the capital of Montana. Then it might happen that that person could be in brain state 67 without

satisfying the rational conditions we ordinarily require in order to make such an ascription. But to allow even the possibility of competition of this sort seemed to Davidson to jeopardize our self-conception. Hence, he wrote, "nomological slack between the mental and the physical is essential as long as we conceive of man [*sic*] as a rational animal" (Davidson, 1980, p. 223).

I think some similar desire to insulate folk psychological practice from certain kinds of empirical risk lies behind Neumanians' insistence that psychology have, as it were, autonomy with a vengeance. I have offered an extended critique of this line of thought in Davidson, and I won't rehearse it here (Antony, 1989, 1995). Suffice to say that we know, thanks to Turing, that it is possible for a physical device to reliably track rational relations. There is no reason, therefore, to think that predictions made from what Dennett (calls "the intentional stance," predictions that exploit rational relations among the presumed contents of mental states, will fail to cohere with predictions made from a lower-level "physical stance" (Dennett, 1971).[3] In any case, we needn't *modus tollens* when we can just as well *modus ponens*. The Neumanians are worried that if we accede to the demand for a reductive explanation of folk psychology, then the failure of such an explanation will jeopardize folk psychology, and they're mightily skeptical that there'll be a reductive explanation. But in my camp, we reason the other way around: given the abundant evidence for folk psychology, there *must* be a reductive explanation forthcoming.[4]

That's the first preliminary note; the second concerns *qualia*. Qualia are the qualitative aspects of certain, mainly sensory, mental states – the "what it's like" to smell a rose, taste a lemon, touch velvet, and so forth. There has been a resurgence of interest in states such as these, with some philosophers arguing that they represent an irremovable obstacle to a comprehensive materialism. Few of these philosophers are forthright substance dualists (Swinburne, 1997); most are "property dualists," arguing that the qualitative properties of such states fail to supervene metaphysically on the physical states with which they are lawfully correlated (Jackson, 1982; Chalmers, 1996; Nida-Rümelin, 2004). Others argue only that the apparent inexplicability of qualia within materialist constraints presents us with a serious epistemological challenge – how *could* materialism be true if there are qualia (Levine, 2001)? Other materialists, however, are persuaded that materialism can accommodate qualia, and advocate one or another of the following three strategies. One, eliminativism: explain the data about qualia without appealing to qualia themselves (Dennett, 1988; Rey, 1993). Two, functionalism: treat qualitative states as higher-order functional states, in one of the ways propositional attitudes are standardly treated in NRM[5] (Shoemaker, 1975; Lycan, 1987 and 1996; Loar, 1990; Levin, 1991; Dretske, 1995; Tye, 1995; Papineau, 2002; Jackson, 2006). Three, reductionism: identify qualitative states with their neuorophysiological correlates (Hill, 1991).

This is not a debate that I can enter into here – not that I want to, anyway. I bring it up only to point out that any one of these materialist options *regarding qualia*, including eliminativism and reductionism, is available to the non-reductive materialist. NRM is the position that at least *some* psychological states, events, or entities are extant and autonomous, not that *all* such states (or alleged states) are. A successful argument for eliminativism or reductionism about qualia, therefore, does not in itself touch non-reductive materialism about propositional attitude states. For that reason,

I'll be focusing in what follows on states of the second kind, and leave the partisans in the qualia debate to work it out among themselves. The dualists, as always, will be completely ignored.

I turn, then, to the arguments for psychological realism. As I've already indicated, it is folk psychology, that loose system of constructs and platitudes by which we explain and predict the behavior of our con-specifics (as well as many of our *non*-specifics), that provides the strongest prima facie case for psychological realism. So let me be a little more systematic, and draw up the kind of thing Georges Rey has called an "explanatory budget" (Rey, 1991) – a list of mundane features of our (ostensible) mental life that demand explanation, one way or another.

1 *Reasoning and deliberation*: Reflecting on what we want, together with the things we believe, we conceive of and determine on a course of action, which, frequently enough, we pursue. Also, reflecting on things we believe, we often come to believe new things. In both these cases, we seem able to exploit rational relations among propositions that express the states of affairs we want or believe to obtain.

2 *Intentional inexistence*: Wanting something, we imagine the thing that would satisfy us – we have the capacity to conceive of things that do not, or do not yet exist. Sometimes we imagine things just for the fun of it. Sometimes we take other people's imaginings seriously and come to believe in, and possibly even worship, things that don't exist.

3 *Opacity*: The particular actions we undertake appear to be a function of the way we *take* the world to be, rather than just the way the world *is*. When deliberate action is involved, the world's features affect what I do only insofar as I represent those features to myself.[6] The movie may actually begin at 7:25, but the time I leave the house will be determined (alas!) by my belief that it starts at 7:45.

4 *Predictive power*: knowing what people believe and want, we frequently can predict what they are going to do. Understanding what people say gives us a leg up, too, since people often tell us what they are going to do before they do it. "I'll be the one wearing the red carnation." Relying on attributions of mental states, we can often predict things we could never possibly have predicted otherwise. I construct a trivial multiple choice test and administer it to an auditorium full of undergraduates. On the assumption that they know the correct answers to the question, and want to do well on the test, I correctly predict the pattern of graphite marks that will appear on (almost all) of the optical-scan sheets I collect.[7]

Now suppose we simply take all these observations at face value, and ask, openly and naively, what could account for them? I suggest that the following picture emerges quite naturally. The creatures who exemplify these characteristics possess a capacity to generate, store, and manipulate *representations* – states that can carry information about the way the world is, but that can also simply express a way that the world might be. These states, in addition to these representational, or semantic, properties, have *causal* properties – they are affected by things that happen to the creature, and they cause the creature to act in its turn. The causal powers are somehow coordinated, in a law-like way, with the semantic properties.

A good naturalist would make this picture the starting point of scientific investigation – why not? The data are manifest; the picture offers an explanation. The first question to ask would be how to understand the notion of "representation" – what kinds of physical states and mechanisms could implement the information processing posited in the naive picture? Turing, of course, provided an answer, by demonstrating how, in principle, a completely physical and fully automatic representation-processing machine could be built. This would be a machine with structured internal elements that could be construed as symbols and internal states defined partly in relation to those symbols, built in such a way that the principles governing the causal interactions among the states (in conjunction with "inputs" and "outputs") mirror rational relations among the representational contents encoded in the symbols. It is important to the adequacy of Turing's model as a model of mind that the "mirroring" be quite strong, and it is – the physical features of the representational elements to which the machine's causal laws are sensitive are precisely the features that serve to encode the elements of the representational contents that are semantically relevant. The generality of the mirroring – the ability of the mechanism to track all the semantic relations that exist among the contents of the symbols – is due to the compositionality of the symbol system as a whole.

The application of Turing's theory of automatic computation to psychology yields a satisfying precisification of the naive conception of mind: Thinking is fundamentally a matter of the manipulation of symbols – physical items with representational properties. The logically relevant aspects of the representational properties of the symbols are encoded in their syntactic forms, and the compositional structure of the symbol system mirrors the semantic and logical relations in which the representational contents of the symbols participate. Mental states are functional relations to mental symbols, and mental processes are computational processes defined over the mental symbols. The hypothesis that minds are like this is the hypothesis that minds have a "classical" architecture. In the 1970s this hypothesis was first articulated and defended, as the "language of thought" theory, by Jerry Fodor, perhaps the world's foremost champion of intentional realism,[8] but it has received substantial development since then, notably by cognitive scientist Zenon Pylyshyn (Pylyshyn, 1986).

The LOTT explains the central phenomena. The hypothesis that mental representations are syntactically structured explains how psychological processes can respect rational relations during deliberation. The hypothesis that agents' behavior is mediated by representations explains both intentional inexistence and opacity phenomena. And the hypothesis that representations are realized in physical structures whose forms strongly mirror syntactic structure explains how representations can have causal powers that track rational relations. Finally, the entire picture explains the projectibility of mentalistic discourse: it explains how beliefs, desires, and other mental states implicated in perception and action can constitute natural kinds, capable of grounding prediction and explanation.

Not only does computationalism provide a satisfying account of folk psychological data, it has proved immensely fertile when extended beyond the realm of conscious and deliberate thought. Beginning with Chomsky's pioneering approach to language acquisition, and continuing with David Marr's theory of visual processing (Marr, 1982), the computationalist model has offered promising explanations of largely

unconscious cognitive feats performed by human beings on a daily basis, such as face recognition. The idea that an innate "theory of mind" underlies our ability to quickly interpret the facial expressions of our con-specifics, and to give intentionalistic construals to characteristically human patterns of behavior, has gained wide acceptance among psychologists: there is serious evidence that absence of such a "psychology module" might be the central deficit in autism. Computationalism has also been extended to the cognitive achievements of infrahuman animals, such as birds' acquisition of their species' songs and insects' spatial navigation, by ethologists such as Peter Marler (1984) and C. R. Gallistel (1990) to account for animal cognition.

So here is the situation: we have available to us an intuitively appealing model of mind, one that explains the central phenomena of mentality and that has generated new and fruitful programs of research within the fields of human psychology and ethology. It is striking, then, that there is so much resistance to this model within philosophy. But what is more striking than the resistance itself is the fact that critics of this picture *have no alternative to offer.* As Georges Rey has pointed out, new paradigms are supposed to recommend themselves by addressing anomalies the old theory cannot explain, but also by "handl[ing] the old one's successes" (Rey, 1991, p. 1).

It is quite true that there are some outstanding problems associated with the computationalist account of mind, the largest of which concerns the notion of "representation." I said that Turing showed that there could be a mechanical device with states that could be *construed* as representations, and I also said that representational properties had to be *encoded* by the posited mental symbols. In the case of artificial minds – computers – we can make sense of all this because *we* create the encodings, and *we* do the construing. But the representational powers of the mind cannot arise in the same way. Partisans of the LOTT must therefore confront the problem of how to *naturalize intentionality*: they must eventually explain how non-intentional processes and relations can give rise to representational relations; how, without the prior activity of minds, some bit of physical reality can come to be "about" something else.

Now the difficulty of this problem may, on its own, frighten some philosophers into mind-denial – surely, they may reason, there's some way to account for human behavior without having to get embroiled in all *that.* But there really isn't. There has been only one serious attempt to develop a non-cognitivist psychology – that is, a psychology that did not posit representational states – and it failed spectacularly. The reasons for that failure are highly instructive. It behooves us, therefore, to remind ourselves why behaviorism didn't work.

Behaviorism was the view that the behavior of all organisms, including human beings, can be predicted and explained without any appeal to independent inner variables (Skinner, 1938, 1957). Behavior was held to be a function of two factors: one, the current stimulus situation of the organism, and two, the organism's past history of environmental interactions. It is not that behaviorists thought that none of the organism's endogenous states were causally relevant to the production of the organism's current behavior. Obviously the organism's history of reinforcement had to leave some kind of mark on the organism in order to affect behavior later on. Moreover, there needed to be a certain number of innate behavioral propensities – there had to be unconditioned reinforcers, and innate "similarity spaces" for example – in order for even classical conditioning to get off the ground. (It is lack of endogenous

states of the right kind that explains, in a sense, why houseplants cannot be trained. Cats, of course, are another story.) But these endogenous factors, it was held, could be safely ignored, because the relevant functional relationships were not affected by the nature of the intervening states of the organism. To put the idea somewhat perversely, we could say that behaviorists thought that the internal states and goings-on that (obviously) mediated the connection between stimulus and behavioral response didn't deserve to be called "mental" – they were neurophysiological substrates, not independent psychological variables. But of course, these endogenous states were the best candidates there were for mental states – so if *they* weren't mental, there might as well be no mental states at all!

The case against behaviorism had both conceptual and empirical dimensions. The philosophical version of behaviorism, developed and championed mainly by Gilbert Ryle (1949) was logical behaviorism, the thesis that "mental" states were nothing but patterns of behavior, and, correlatively, that mentalistic expressions could be analyzed in terms of, and hence eliminated in favor of, talk of dispositions to behave. The particularly fine-grained mentalistic ascriptions made to and by human beings could be accounted for, Ryle argued, by the particularly rich behavioral repertoire afforded to us by the capacity to speak. Thus, the "belief" that Helena is the capital of Montana is largely constituted, according to the logical behaviorist, by the disposition to make and to respond to verbal behavior involving sounds such as "What is the capital of Montana?" and "Helena." What *exactly* accounted for this complex dispositional structure – i.e., language – in human beings was presumed to be some purely *quantitative* difference in neurology between us and other reasonably smart primates. (It had to be simply a matter of degree, because the black box approach to the mind offered no other degrees of freedom. But in fact the matter never received serious attention.)

Ryle's eliminativist/reductivist program foundered on the fact that human intentional behavior – the behavior for which we are most apt to give mentalistic explanations – is always a function of at least two independent mental variables: a belief and a desire. There is thus no range of behavioral responses proprietary to any individual mentalistic ascription, even given fixed circumstances. Any piece of behavior can evince any belief whatsoever, provided it is combined with an appropriate desire. So I may evince my belief that Helena is the capital of Montana by saying out loud within your hearing, "Helena is the capital of Montana," *if* I want to inform you about US geography. But if I want instead to mislead you, then I will evince that same belief by saying anything *but* those words. Similarly, one and the same behavioral response can evince contrary beliefs, depending on my other mental states. If I believe that Helena is *not* the capital of Montana, and wish to mislead you, I may say exactly the same words I'd say if I believed it was, but wanted to inform you.

The empirical case against behaviorism included, famously, Chomsky's detailed critique of Skinner's account of human verbal behavior (Chomsky, 1959). Chomsky first showed that the theoretical apparatus of operant conditioning theory was, in one way or another, inadequate for explaining the actual course of human linguistic development. If crucial concepts such as "operant" and "stimulus generalization" were given their strict, technical meanings, the theory failed on grounds of empirical inadequacy. If these concepts were "analogically extended," as Skinner said they must be, then, Chomsky argued, the theory lost empirical content. The second important

element of Chomsky's assault was the "poverty of the stimulus" argument, deployed against the associationist element of behaviorist theory. What Chomsky and others demonstrated, in the first instance, was that the amounts and kinds of data that the operant conditioning model predicted would be necessary for the acquisition of language simply did not match the data actually available to successful language learners. Chomsky posited an innate, domain-specific cognitive structure that encoded highly general information about the structure of human languages that sharply constrained the range of grammars a child could hypothesize in response to the linguistic data provided by other speakers.

All this is familiar enough. But there is another line of empirical criticism that may be less well-known, and that is, for my current purposes, more interesting. Although many cognitive scientists and many philosophers were completely convinced, on the basis of the critiques outlined above, that behaviorist theory could not account for *complex* human behavior, such as the acquisition and deployment of language, they were not inclined to doubt that classical and operant conditioning explained at least *some* elements of our behavior. It still seemed reasonable to suppose that the behaviorist story made sense for those relatively unconscious and elementary bits of learning that were investigated in standard behaviorist learning experiments.

William Brewer, however, saw reason to challenge even this bromide (Brewer, 1974). It is important, he argued, to distinguish the phenomenon of conditioning itself from the non-mentalistic *explanation of* conditioning offered by behaviorists. It is one thing to condition a subject to exhibit a high galvanic skin response (GSR) at the sight of a particular apparatus; it is quite another to show that this conditioning is accomplished *automatically*, without any cognitive mediation. It is possible, after all, that the *way* conditioning occurs is entirely cognitive: that the subject *learns that* a certain apparatus is capable of producing a shock, and accordingly becomes fearful in anticipation of receiving the shock. Indeed, as I've been arguing, this is the explanation that common sense suggests. The crucial question for behaviorists, then, is this: What makes the non-mentalistic explanation preferable to the more natural mentalistic one?

What one would need to choose between the non-cognitive and the cognitive explanations of the conditioning effect are experiments that controlled for the putative mental states of the subjects. Permitting ourselves to speak, provisionally, with the vulgar, we would want to ask what would happen, for example, if the subject in a GSR experiment were given reason to think that she would not, in this instance, receive a shock – perhaps by being shown the apparatus being unplugged? Well, as it turns out, Brewer reports, the answer to this very question – together with a wealth of other highly pertinent data – was in fact present in the literature generated by behaviorists themselves. The data clearly supported the mentalistic interpretation. In the case described above, if the subject comes to believe that the conditioned stimulus (the sight of the machine) has been "dissociated" from the unconditioned stimulus (the electric shock), the conditioned behavior is almost instantly extinguished, contrary to the predictions of the behaviorist model. Brewer's extensive review of *behaviorist* experiments involving a variety of such "dissociations" (CS from US, operant from reinforcer, etc.), led him to conclude that "all the results of traditional conditioning literature are due to the operation of higher mental processes, as assumed in

cognitive theory, and that there is not and never has been any convincing evidence for unconscious, automatic mechanisms in the conditioning of adult human beings" (Brewer, 1974, p. 27).[9]

I've been arguing that from the point of view of common sense, behaviorism is an extremely radical thesis: it after all denies what appears to be the most salient feature of our psychological lives, namely that it is psychological. Given that, one would have thought it could only have been accepted under irresistible empirical pressure. And yet, as it turns out, the theory was grossly and evidently inadequate in accounting for even the paradigm phenomena in its purported domain. There was, in short, no empirical reason whatsoever to prefer behaviorism to common-sense mentalism. Why did it then flourish for as long as it did?

One reason, surely, was the positivist Zeitgeist of the early twentieth century, which was hostile to theoretical posits. Watson, in his 1915 Presidential Address to the American Philosophical Association (Watson, 1916), made clear that he thought the main argument for behaviorism was the epistemological advantage it afforded over introspectionist psychology. Mental states, per se, were not interpersonally observable, and hence were not fit objects of scientific investigation. No doubt the lingering association of mentalism with dualism – perhaps it was inconceivable to many that one could be a mentalist without being a dualist – inhibited the thought that mental states might earn their way into the realm of genuine science in just the way photons and electrons had, by being ineliminable elements in serious scientific explanations for ordinary observable phenomena.

Quine (1960) offered philosophical reasons for thinking that there could be no science of the mind. But, notoriously, his argument for this conclusion begged the question against the psychological realist. Quine wanted to establish that there were no objective facts about either the meanings of words or the contents of thoughts, by showing that such putative facts would be necessarily underdetermined by the physical facts. Quine began with the explicitly behaviorist premise that human children had to have acquired language on the basis of conditioning to the contingencies of verbal use. On theoretical – that is, empiricist – grounds, he excluded from the child's data set information about general grammatical or semantic structure, as well as any psychological information that (one might have thought, pre-theoretically) would help the child in forming theories of language and language use (information such as: "When Daddy points to something and says a word, he's trying to tell me what that kind of thing is called.")

At the same time, however, Quine liberally idealized *away* the very constraints on amount and type of evidence that in fact make the behavioristic story he told impossible as an account of the acquisition of human language. It matters little whether a child *could* acquire, say, the meanings of common verbs through operant conditioning over many trials with explicit reinforcement, since the child acquires them over few trials and without any explicit reinforcement – because, in other words, the child actually acquires such meanings *without* the kinds and number of experiences that would be required for the behaviorist model to apply. That Quine, the archetypal naturalized epistemologist, could so blithely ignore the real world conditions in which language emerges is testimony to the power of a priori hostility to the mind. The irony is rich – it is Quine himself who taught us that it is bad method to disregard

or trade off explanations of known psychological facts for putative gains of a more theoretical – or ideological – nature.[10]

I take it as settled, then, that a good, naturalistic materialist ought to be a psychological realist. But now the reductionist challenge must be faced: Is there an autonomous psychology, or are psychological kinds simply biological kinds spoken of mentalistically? Again, a little history is in order.

In the middle of the twentieth century, U. T. Place (1956) and J. J. C. Smart (1959) advanced and defended the "mind–body identity theory," the view that mental states could be identified with neurological states. Because Place and Smart argued that every *type* of mental state could be identified with a *type* of neurological state, their version of reductionism became known as "type-reductionism." Place and Smart took their view to be simply the natural expression of materialism about the mind, and did not consider the possibility that there might yet be an autonomous, yet materialist, science of the mind. They focused on consciousness and sensations, phenomena that were most amenable to this sort of treatment. It is plausible, at any rate, that a sensation type, such as pain, might at least be reliably correlated with a single type of neurological event or process, such as the firing of C-fibers. What Place and Smart failed to consider, however, were propositional attitude states, such as beliefs and desires. It seemed highly unlikely that these states would be marked by some distinctive kind of neurological cell type or process, and much more likely that they would involve states of great neurological complexity, states that might vary in their details from person to person, or even from one time to another in the same person.

Other materialist philosophers who did pay central attention to these phenomena, notably Putnam, Armstrong, and Lewis, argued that such states needed to be understood in terms of their typical causal profiles, or functional roles (Putnam, 1965 and 1967; Armstrong, 1968; Lewis, 1972). These early functionalists appreciated what the behaviorists did not, namely that behavior crucially implicated complex internal states. But their treatment of mental states did incorporate one important insight of Ryle's, and that was that psychology was a more abstract level of description than was biology. Materialism demanded that every psychological state have *some* physical realization, but because psychological states were functional, the details of that physical realization didn't matter to their identity conditions. Putnam made this view explicit in arguing that mental properties were "higher-order" properties. A higher-order property is the property of having some other ("lower-order") property that meets a certain causal/functional specification. The lower-order properties are called "realizer" properties. The view that mental properties are higher-order, functional properties that could be realized in a variety of distinct lower-order properties became known as the thesis of *multiple realizability* (MR).

The view that mental states are multiply realized yielded a different picture of the relation between mentalistic types and biological types from the one presumed by the identity theorists. On their view, recall, all psychological types would be correlated and henceforth reduced to types in a lower-level science, presumably biology. Bridge laws, expressing the relations among these types, could then be used to reduce all the generalizations of psychology to biology. According to MR, however, the requisite biconditional bridge laws are unobtainable. Psychological states supervene on the biological: lower-order biological states would be nomologically sufficient for, and

hence would necessitate, higher-order psychological states, but the converse does not hold. Generalizations describing regularities at higher orders are thus irreducible to laws at lower orders. They are autonomous.

The multiple realizability view not only captures the abstractness of psychological description relative to the biological, but it appears to explain how mental events can be causally efficacious in the physical realm. Although they reject the view that mental *types* are reducible to physical *types*, defenders of MR are, by and large, materialists, and take it as part and parcel of their materialism that mental events, like all concreta, have physical properties. (Some advocates of MR accept the stronger view that every instance of a mental type is identical with an instance of some physical type, the view known as *token*-reductionism.) In any given causal interaction involving mental events, it will be the mental event's lower-order physical realizer properties that account for the particulars of that causal transaction. Mental events can therefore "inherit" the causal powers of their physical realizers. Different instantiations of the same mental type must all display, at an appropriate level of abstraction, the same causal profiles; but the explanation for how those profiles are maintained can vary widely from case to case.

Jaegwon Kim has recently charged that this apparent selling point of the MR view is in fact a fatal liability (Kim, 1993 and 1998). To concede that mental events do not form a causally homogeneous group, he argues, is to concede that mentalistic groupings are not genuine natural kinds, and hence, are unfit for science. But MR supporters must concede this: to do otherwise would be to posit new "emergent" causal powers that would compete with the lower-order physical realizer properties, in a way that should be unacceptable to a materialist. The only way out of this dilemma, Kim argues, is to give up on multiple realization, and return to type reductionism. If mental phenomena are to be integrated into the physical world, they must enter as biological, not psychological phenomena.

But what about considerations of multiple realizability? If mental states can indeed be realized in a variety of distinct physical properties, then how can the identification with lower-order properties go through? The phenomenon of multiple realizability, Kim responds, is so far merely theoretical. Our only actual extant examples of minds involve creatures with brains, and our extant psychological theories are designed to describe them. It is fanciful to expect that there would be nontrivial generalizations subsuming any *possible* mind, and it is only this possibility that sustains the argument against identifying (familiar terrestrial) minds with brains.[11] As for the presumed neurological inter- and intra-personal diversity of the biological realizer states, this is not the kind of diversity that warrants the description "***multiple*** realizability." Better to say that these states are multiply ***instantiable***. While there will be variety in the microstates that instantiate a given psychological state, all these states will be sufficiently similar in their causal powers to be subsumable under the same causal laws.

Many objectors to Kim's argument charge that his "causal exclusion" argument proves too much; that if successful, it would apply to all non-fundamental domains and sciences, delegitimating geology and biology along with psychology. Kim's initial response to this, the "generalization" objection, involved a distinction between "higher-*order*" from "higher-*level*" properties, and the domains defined in terms of them. A higher-level property is a property that applies to an object at a given level of

aggregation, and to no proper part of that object, whereas higher-order properties, by definition, apply to precisely the same objects as do their lower-order realizer properties. But higher-level objects, Kim argues, are associated with genuinely novel causal powers – there are things that can be done by masses of 10 grams that cannot be done by anything smaller. There is therefore no question of causal competition between higher- and lower-level properties; no property of a part of a higher-level object has the same causal potential as the properties of the whole. Because chemistry, biology, and geology are all higher-*level* relative to fundamental physics, they are not subject to the causal exclusion argument.

This reply does not work, for two reasons. The first is that psychology is not the only "special science" to make use of functional properties: in particular, such properties are ubiquitous and ineliminable in biology. Moreover, the functional properties appealed to by biology – properties such as "cell" and "gene" – are not just possibly but actually and manifestly multiply realizable. So if there is a problem about the nomic status of psychological generalizations, then there is an equally grave problem about the laws of genetics, and in that case, who cares about nomic status?

The second reason is this: Multiple instantiability is as good as multiple realizability for generating a causal exclusion problem. Consider a particular instance of biological causation, say an immune cell attacking a virus. The biological properties involved will all supervene on specific "microbased" properties, where a microbased property is the property of having such-and-such a microparticulate structure. But then these microbased properties will be higher-order with respect to the lower-order chemical or physical properties possessed by the biological entities' proper parts, and will be available to causally compete with the biological properties. Kim acknowledges this, but thinks that there is, in such cases, no difficulty in identifying the biological properties with the microbased properties. But there is – it is the same problem that confronted classical strong reductionism in the face of (at least the prospect of) multiple realizability: many different microbased properties can instantiate the same biological property; hence the biological property cannot be identified with any of them.

In general, Kim and other reductionists need to show that there is a compelling difference between biology and psychology, such that we can rest content with a biology that is autonomous from chemistry, but not a psychology that is autonomous from biology. I submit that no such difference will be – or can be – found. Biological theories earn their keep by providing fertile and explanatorily satisfying accounts of the phenomena we pick out under biological description. No one frets about how such theories will be "integrated" into the non-biological realm (although I understand that there have been such worries in the past), for it is presumed that the truth cannot be an enemy to the truth; that if biological phenomena are, as they certainly appear to be, part of the natural material world, that their existence is compatible with their being composed of chemical and ultimately physical stuff. Why cannot the same attitude be taken toward psychological phenomena? It is only if one assumes going into the game that "the mental" is somehow defined in contradistinction to the physical that there can even *appear* to be a problem about "locating" the mind in a physical world. Nonreductive materialists are thoroughgoing naturalists: we want only the same consideration for the psychological data as are according the data in any other domain.

Think about it.

## Notes

1 Oddly enough, this is a view also held by many philosophers who regard themselves as realists about the mind, e.g., Davidson, Burge, and Baker. More on this below.

2 In honor of Jaegwon Kim's likening their attitude to the insouciance of the *Mad* magazine cover boy, Alfred E. Neuman. There's a picture here: www.leconcombre.com/alfred/img2/alfred_e_neuman_1.jpg

3 I should not leave the impression that Dennett is a psychological realist in my sense. He probably should be placed in the Neumanian camp, since he is an instrumentalist about mentality. He believes that the display of sufficiently robust rational "patterns" in an entity's behavior is ontologically sufficient for attributing mentality to it.

4 For fuller discussion, see Antony and Levine (1997) and Antony (2001).

5 I am including in this group *representationalists* about qualia – theorists who believe that the qualitative character of qualitative states is determined by their representational content.

6 If you won't take my word for it, how about an economist's? "When making choices in the marketplace, 'People are not responding to the *actual objects* they are choosing between,' says Eric Wanner of the Russell Sage Foundation. 'There is no direct relation of stimulus and response. Neoclassical economics posits a direct relationship between the object and the choice made. But in behavioral economics, the choice depends on *how the decision-maker describes the objects to himself.* Any psychologist knows this, but it is revolutionary when imported into economics'" (Lambert, 2006, pp. 94–5). Any psychologist, maybe, but not any philosopher.

Mind-denial appears to have been rampant in the field of economics, delaying by decades the official recognition, not to mention the theoretical exploitation, of the mundane fact recorded above. In the same article, author Lambert quotes Eric Wanner, who worked together with Alfred P. Sloan to legitimize and develop the field of behavioral economics: "'The field is misnamed – it should have been called cognitive economics,' says Wanner. 'We weren't brave enough'" (Lambert, 2006, p. 52).

7 This example is due to Georges Rey (1997, pp. 88–94).

8 See Fodor (1975 and 1978a) for the canonical arguments, and Fodor (1978b, 1987, and 1990) for responses to some objections. In all of this, my intellectual debt to Fodor is profound, and, I trust, evident.

9 Georges Rey (1997, pp. 99–103) surveys and describes four types of "anomalies" – findings inconsistent with behaviorist predictions, but fully expected on a cognitivist model: latent learning, passive learning, spontaneous alteration, and improvisation.

10 For further discussion, see Antony (2000).

11 See also Millikan (1986).

## References

Antony, L. (1989). Anomalous monism and the problem of explanatory force. *Philosophical Review*, 98, 153–87.

— (1995). Law and order in psychology. *Philosophical Perspectives*, 9, 1–19.

— (2000). Naturalizing radical translation. In A. Orenstein and P. Kotatko (eds.), *Knowledge, Language, and Logic*. Boston Studies in the Philosophy of Science. Dordrecht: Kluwer Academic.

— (2001). Brain states, with attitude. In A. Meijers (ed.), *Explaining Beliefs: Lynne Rudder Baker and Her Critics*. Chicago: University of Chicago Press, CSLI.
Antony, L. and Levine, J. (1997). Reduction with autonomy. *Philosophical Perspectives*, 11, 83–105.
Armstrong, D. (1968). *A Materialist Theory of Mind*. London: Routledge.
Baker, L. R. (1995). *Explaining Belief: A Practical Approach to the Mind*. Cambridge: Cambridge University Press.
Bickle, J. (2003). *Philosophy and Neuroscience: A Ruthlessly Reductive Account*. Dordrecht: Kluwer Academic.
Brewer, W. F. (1974). There is no convincing evidence for classical conditioning in adult humans. In W. B. Weiner and D. S. Palermo (eds.), *Cognition and the Symbolic Processes*. Hillsdale, NJ: Erlbaum.
Burge, T. (1993). Mind–body causation and explanatory practice. In J. Heil and A. Mele (eds.), *Mental Causation*. Oxford: Oxford University Press.
Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.
Chomsky, N. (1959). Review of B. F. Skinner's *Verbal Behavior*. *Language*, 35, 26–58.
Churchland, P. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, 78, 67–90.
Davidson, D. (1980). Mental events. *Essays on Actions and Events*. Oxford: Oxford University Press. Originally published in L. Foster and J. W. Swanson (eds.), *Experience and Theory*. London: Duckworth, 1970, 79–91.
Dennett, D. (1971). Intentional systems. *Journal of Philosophy*, 68, 87–106.
— (1988). Quining qualia. In A. Marcel and E. Bisiach (eds.), *Consciousness in Contemporary Science*. Oxford: Clarendon.
Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.
Fodor, J. (1975). *The Language of Thought*. Cambridge, MA: Harvard University Press.
— (1978a). Propositional attitudes. *The Monist*, 61, 501–23.
— (1978b). Three cheers for propositional attitudes. In E. Cooper and E. Walker (eds.), *Sentence Processing*. Hillsdale, NJ: Erlbaum.
— (1987). The persistence of the attitudes. *Psychosemantics*. Cambridge, MA: Bradford/MIT Press.
— (1990). Why there still has to be a language of thought. In D. Partridge and Y. Wilks (eds.), *The Foundations of Artificial Intelligence: A Sourcebook*. Cambridge: Cambridge University Press.
Gallistel, C. R. (1990). *The Organization of Learning*. Cambridge, MA: MIT Press.
Hill, C. S. (1991). *Sensations: A Defense of Type Materialism*. Cambridge: Cambridge University Press.
Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–36.
— (2006). The knowledge argument, diaphanousness, representationalism. In T. Alter and S. Walter (eds.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford: Oxford University Press.
Kim, J. (1993). Multiple realization and the metaphysics of reduction. *Supervenience and Mind*. Cambridge: Cambridge University Press.
— (1998). *Mind in a Physical World*. Cambridge, MA: MIT Press.
Lambert, C. (2006). The marketplace of perceptions. *Harvard Magazine*, March–April 2006, pp. 50–7, 93–5.
Levin, J. (1991). Analytic functionalism and the reduction of phenomenal states. *Philosophical Studies*, 61, 211–38.
Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford: Oxford University Press.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249–58.
Loar, B. (1990). Phenomenal states. *Philosophical Perspectives*, 4, 81–108.
Lycan, W. G. (1987). *Consciousness*. Cambridge, MA: MIT Press.
— (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.
Marler, P. (1984). Song learning: innate species differences in the learning process. In P. Marler and H. S. Terrace (eds.), *The Biology of Learning*. Berlin: Springer.
Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman.
Millikan, R. (1986). Thoughts without laws. *Philosophical Review*, 95, 47–80. Reprinted in R. Millikan, *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press, 1993.
Nida-Rümelin, M. (2004). Phenomenal essentialism: a problem for identity theorists. *Philosophy and Phenomenological Research*, 58, 51–73.
Papineau, D. (2002). *Thinking about Consciousness*. Oxford: Oxford University Press.
Place, U. T. (1956). Is consciousness a brain process? *British Journal of Psychology*, 47, 44–50.
Putnam, H. (1965). Brains and behavior. In R. Butler (ed.), *Analytical Philosophy*, 1–20.
— (1967). The mental life of some machines. In S. Hook (ed.), *Dimensions of Mind*. New York: Collier.
Pylyshyn, Z. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
Quine, W. (1960). *Word and Object*. Cambridge, MA: Technology Press of MIT.
Rey, G. (1991). An explanatory budget for connectionism and eliminativism. In J. Tienson and T. Horgan (eds.), *Connectionism and the Philosophy of Mind*. Dordrecht: Kluwer.
— (1993). Sensational sentences. In M. Davies and G. W. Humphreys (eds.), *Consciousness: Philosophical and Psychological Essays*. Oxford: Blackwell.
— (1997). *Contemporary Philosophy of Mind*. Oxford: Blackwell.
Ryle, G. (1949). *The Concept of Mind*. London: Hutcheson.
Shoemaker, S. (1975). Functionalism and qualia. *Philosophical Studies*, 27, 291–315.
Skinner, B. F. (1938). *The Behavior of Organisms*. New York: Appleton-Century-Crofts.
— (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.
Smart, J. J. C. (1959). Sensations and brain processes. *Philosophical Review*, 68, 141–56.
Stich, S. (1983). *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
Swinburne, R. (1997). *The Evolution of the Soul*. Oxford: Oxford University Press.
Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: Bradford Books of MIT Press.
Watson, J. B. (1916). The place of the conditioned-reflex in psychology. *Psychological Review*, 23, 89–116.