January 1992

Who's Afraid of Multiple Realizability?: Functionalism, Reductionism, and Connectionism

Hinton, G., & Sejnowski, T. (1986). Learning and relearning in Boltzmann machines. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 282–317). Cambridge, MA: MIT Press.

Hopfield, J. J. (1982, April). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA, 79,* 2554–2558.

Hopfield, J. J. (1984, May). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Science, USA, 81,* 3088–3092.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica, 47* 263–291.

Kandel, E. R., & Schwartz, J. H. (1985). *Principles of neural science* (2nd. ed.). New York: Elsevier North Holland.

Lachter, J., & Bever, T. G. (1988). The relation between linguistic structure and associative theories of language learning: A constructive critique of some connectionist learning models. In S. Pinker & J. Mehler (Eds.), *Connections and symbols* (pp. 195–247). Cambridge, MA: MIT Press.

Machtey, M., & Young, P. (1978). *An introduction to the general theory of algorithms*. New York: North Holland.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 3–44). Cambridge, MA: MIT Press.

Nadel, L., Cooper, L. A., Culicover, P., & Harnish, R. M. (1989). (Eds.). *Neural connections, mental computation*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., and McClelland, J. L. (1986). A General Framework for Parallel Distributed Processing. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45–76). Cambridge, MA: MIT Press.

Pylyshyn, Z. (1984). *Computation and Cognition*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60–94.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 170–215). Cambridge, MA: MIT Press.

Rumelhart, D. E., & Norman, D. A. (1981). Introduction to G. E. Hinton, & J. A. Anderson (Eds.), *Parallel models of associative memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sejnowski, T. J. (1986). Open questions about computation in cerebral cortex. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 372–389). Cambridge, MA: MIT Press.

Shepherd, G. M. (1988). *Neurobiology* (2nd. ed.). New York: Oxford University Press.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1–74.

5 Who's Afraid of Multiple Realizability?: Functionalism, Reductionism, and Connectionism

Justin Schwartz
Ohio State University

1. INTRODUCTION

The current orthodoxy among cognitive scientists and philosophers of mind holds that reductive materialism has been decisively refuted by Fodor (1981) and Putnam (1975). Because mental states are in principle *multiply realizable* in a variety of physical ways, for example, in humans and computers, it is widely held that "If functionalism is true, physicalism is probably false" (Block 1980, p. 270). *Functionalism* is a disputed notion, often identified with multiple realizability—mistakenly, it is argued here. For our purposes it is the view in philosophy of mind that mental states are characterized by their causal or computational role in producing behavior. *Physicalism* here means the metaphysical thesis that every kind of mental state is identical with some kind of physical state. The corollary epistemological thesis is *reductionism*, the doctrine that the mental can in principle be systematically explained in terms of or reduced to the physical in virtue of these identities.[1] A similar multiple realizability argument dismisses the relevance to psychology of *connectionism*, sketched in the next section. This is a reductive materialist program if we identify connectionist networks with neural networks. According to the argument, even if cognition is implemented in the "brainlike" way connectionists urge, this has no explanatory bearing on psychology because it is only one possible implementation.

The multiple realizability of the mental, according to Fodor and Putnam,

[1]Smart (1971) and Armstrong (1968) are the classic statements of these positions. A different form of physicalism (constitutive materialism) which does not rely on identity is sketched in the body of the present chapter and further developed in Schwartz (1991).

shows that such reductionist claims are false because the mental and the physical unsystematically cross-classify each other. The same type of mental state (e.g., a belief that p) can be realized in a variety of physical states; thus, we must reject the classic identity thesis (strong or *type* physicalism). In addition, because the same mental state can have a variety of types of physical realizations which may have no scientifically interesting commonalities, the reductionist program of explaining mental properties and processes in physical terms must go by the board. The orthodoxy accepts a weaker identity thesis (weak or *token* physicalism), according to which each token or instance of any mental state is identical to some token of a physical state, but no type of mental state is identical to any type of physical state. The token identity is a merely metaphysical fact that does not underwrite any explanatory connection between the mental and the physical. Psychology is autonomous and irreducible. Call this view *autonomous functionalism*.[2] As Block's remark shows, some even hold that the expression autonomous functionalism is redundant: functionalism implies the autonomy of the mental.

I argue that this view is radically mistaken. If valid, the argument proves too much. A functional state of a system is one intrinsically characterized by its causal or computational role in mediating between inputs (e.g., sensory stimulation) and outputs (e.g., behavior), as well as by the internal relations among the functional states themselves. Multiple realizability is a property of any functionally characterized state, so the argument would establish the irreducibility of any theory that appeals essentially to such states, such as transmission genetics, which adverts to the states and properties of genes. *Gene* is a functional term. To say *x is a gene* is to characterize it in terms of its functional role in transmitting hereditary information. Genes are therefore multiply realizable. But if the multiple realizability argument for autonomy denies the reducibility of genetics, this is a reductio: transmission genetics is in part reduced to molecular genetics.

What has gone wrong? One diagnosis focuses on the empiricist notion of reduction invoked by the multiple realizability argument, which fails to capture many actual explanations of less basic by more basic sciences (see Schwartz 1991). A deeper diagnosis, though, holds that autonomous functionalism, while not perhaps bound to an untenable empiricist account of science and of reductive explanation, has an inappropriate conception of psychology as a priori universal, applying in the same way to all cognizers. This is no more plausible than a universal biology. I suggest that psychology, like biology, be *provincialized*, giving us different psychologies for different classes of cognizers. The potential universal applicability of psychology may be reconstrued as a consequence of its

[2]The autonomy is a matter of irreducibility. It has nothing to do with the psychological individualism presupposed by Stich's "autonomy principle," or Fodor's similar "methodological solipsism" (Fodor, 1981, pp. 225–256) according to which only what's inside the epidermis is admissible in an explanatory psychological theory (Stich, 1983, pp. 160–170).

idealized character. This move permits species-specific or, more precisely, provincial, reduction of psychologies.[3] A reduction of merely human psychology would count as satisfying reductive materialism, whether or not it reduced the psychology of any possible cognizer. Whether any such reductions take the form of establishing classical type-identities or some different relation of constitution (as suggested in sections 6 and 7) depends on the facts and on theoretical convenience. Whether any reduction of psychology at all is forthcoming is an empirical question, but I wish to remove the idea that multiple realizability is an impediment.

2. CONNECTIONISM AND GOFAI

The discussion bears on more than the fate of the identity theory and materialistic reductionism. Almost exactly the same issues emerge in the context of the current debate over the relation between the competing programs of GOFAI (Good Old-Fashioned Artificial Intelligence, in Haugeland's [1985] expression) and connectionism as accounts of human cognitive processes.[4] Very roughly, advocates of GOFAI maintain that such processes must invoke explicit rules operating on syntactically structured representations. Connectionists urge that these processes be understood instead in terms of the activation of nodes or patterns of nodes in neural nets that are not governed by explicit rules nor have any syntactic structure. Fodor and Pylyshyn (1988) argue that while GOFAI models may be implemented in a connectionist manner, they need not be, and therefore whatever the implementation, it is not explanatorily interesting in understanding cognition. The claim is essentially that GOFAI is multiply realizable in different implementations, and so autonomous with respect to them.

Here the issue is not directly the reduction of functionally characterized mental states to physical ones but of one functional model, GOFAI, purportedly of cognition, to another, connectionism, which Fodor and Pylyshyn maintain is not a model of cognition at all and the connectionists claim is the correct model of cognition. I do not attempt to adjudicate who, if either, is right. My point is more restricted, namely, that multiple realizability arguments against reduction of

[3]Species-specific is not the right term, as we can see in view of the biological analogy. All forms of terrestrial life appear to use DNA as the hereditary material (viruses are a primitive precursor of this system). So we have at least a planet-specific genetics shared by different species on Earth. (It may be in part because of this that the physiological basis for psychology also appears to be shared widely among different terrestrial species.) So I use the term *provincial*, explained more fully later, where that indicates any group of species or class of entities that shares a sort of structural basis for a higher-order functional system, biological or psychological.

[4]It is debatable how different the programs are and whether they really are in competition. See Bechtel (1988) for discussion. I believe but will assume without argument that they really do compete.

GOFAI or psychology to connectionism are as defective as those against reduction of the mental to the physical. Therefore, if GOFAI is the correct account of cognition and if GOFAI architecture has a connectionist implementation in humans, then connectionism provides a reduction of GOFAI in humans.[5]

The parallel between the mental–physical case and GOFAI-connectionism case may be closer than mere analogy. As Churchland (1989, pp. 153–196) observed, part of the attraction of connectionism is precisely its promise of offering a more realistic account than GOFAI does of neural processing in human brains. Connectionists hope that their networks will be revealed to be networks of neurons. If so, then, GOFAI aside, connectionism might provide a physical reduction of human cognition. It is still too early to say whether connectionism is more than merely promising as an account of the functional architecture of cognition, much less whether it is also the correct account of the neural architecture. Whether or not connectionism is the correct account of how cognition is implemented in any cognizer, however, my point is that if it is nonetheless the correct account for human beings, that is good enough for provincial reductive explanation of human cognition. And that would be good enough for reductive materialism.

3. ANTI-CHAUVINISM AND AUTONOMY

Underlying autonomous functionalism is the intuition that psychology, like arithmetic, is not tied down to the physical structures in which it may be instantiated. Just as the same mathematics applies whether we use electronic calculator, a wooden abacus, or a human brain,[6] so (it is said) psychological explanations must apply equally and in the same way to humans, thinking computers, or Martians. Our provincial physical constitution is highly contingent. Martian neurophysiology might even be silicon-based (like a computer) rather than carbon-based. But these differences cannot be that important when it comes to psychological explanation. To do psychology at all, we must abstract from the material constitution of beings whose behavior we seek to explain. To deny this is chauvinist, refusing to attribute mentality because of irrelevant differences in physical realization (Block 1980, p. 270).[7] This intuition may be expressed by the claim that beings who vary widely physically can have identical mental states, just as calculators that vary widely physically can have identical computational states. That is to say, psychological states, like arithmetical or computational ones, are capable of multiple physical realizations.

The multiple realizability argument presupposes that psychology has to fit any possible intelligent beings, even ones *ex hypothesi* different from us physically. Is this a reasonable demand? Humans are the only class of beings that we know has a full-blown intentional psychology, and who, therefore, could be objects of empirical psychological study as things now stand. We are interested in nonhuman intelligence, of course; and if we knew any thinking computers or intelligent aliens, we would be very interested in them, too. But so far as we know, there aren't any. Below I suggest some things we might say if there were. This chapter argues that functionalism and multiple realizability are independent theses.

The orthodoxy holds that multiple realizability, so construed, together with a number of supposedly uncontroversial claims about the nature of reductive explanation, shows that distinct disciplines will cross-classify their predicates so that there will be no clean correlations among their terms (Fodor, 1981, p. 145). The same belief that p would be realized by physical state P1 in humans, P2 in Martians, P3 in some type of computer, and so forth. Type-type correlations (implied by the classic identity thesis) are a necessary condition for reduction on the empiricist model invoked by Fodor, so nothing would count as a proper explanation of psychological phenomena in physical terms. Such phenomena will have physical descriptions, allowing us to a degree to save our materialist inclinations, but this materialism (token physicalism) will do no scientific work.

Fodor's presentation of the argument relies on the logical empiricist deductive-nomological (D-N) account of explanation as *nomic subsumption*.[8] To ex-

[5]Connectionism enthusiasts (I am a moderate fan of the program) tend to deny the two antecedent clauses. They think that GOFAI is wrong and connectionism right as an account of cognition. They typically hold that the correct relation between the two models is not that the latter reduces the former, but rather replaces it. This is a sort of eliminativism about GOFAI rather than any sort of reductionism about it. Many connectionists, however, think of connectionism as a program for reducing commonsense psychology if not GOFAI; although some, like Churchland, conceive of it (if *conceive* applies here) as a replacement for commonsense psychology, and therefore are full-fledged eliminativists about the mental, not just about GOFAI.

A different possibility is suggested later: even if GOFAI is wrong as an account of much of cognition, as I suspect, it may be construed as an idealization of connectionist mechanisms, which therefore (as I understand reduction) reduce it. At this point I do not think we know enough to say, although I suspect that this may be the best way to look at it. In Schwartz (in press) I urge that this sort of ideal type reductionism is the best way to understand propositional attitude psychology and our ordinary folk theory of mind.

[6]Cummins and Schwartz (1988) explore this analogy, which is common, but problematic. See Schwartz (in press, section 10.3) for more discussion.

[7]Liberalism, the opposite vice, is the attribution of mentality to beings which plausibly do not have it.

[8]Logical empiricism was the reigning philosophical account of the structure of science from the 1930s to the mid-1960s. See Nagel (1961) for a comprehensive account. It has succumbed to 40 years of withering criticism. Empiricism may be understood, roughly, as the doctrine that science systematizes into laws statements about observable regularities in experience, enabling us to predict the course of experience from statements of these lawlike regularities. Explanation on the empiricist account is exactly symmetrical to prediction. In contrast, scientific realism—the view of science I share with Fodor and (the old) Putnam—regards science as the description of the real nature of phenomena and their explanation by elucidation of the causal mechanisms that produce or (in cases of reduction) constitute them. Scientific realism accords no special status to observed phenomena and is not committed to seeking laws.

plain an event, a law, or a theory, according to this view, is to derive it logically from a more general law or theory. A reduced law or theory T_1 is fully explained by or reduced to a reducing law or theory T_2 just in case every theoretical term F of T_1 is linked by biconditional "bridge laws" to a term G of T_2 (so that [F iff G])—thus reducing the terms referring to entities or properties. With such links, the laws, and ideally the whole of T_1 would be deductively derivable from T_2. Fodor's argument denies that the terms and laws of psychology have the "right" kind of coextensions (lawlike ones) at the physical level.

Fodor (1981) illustrated the point with Gresham's Law—roughly, "Bad money drives out good"—an example of a purported law from economics, a nonbasic science. With respect to any physical explanation of such a generalization, he says:

> A physical description which covers all [monetary exchanges] must be wildly disjunctive. Some monetary exchanges involve strings of wampum. Some involve dollar bills. And some involve signing one's name to a check. What are the chances that a disjunction of physical predicates which covers all these events . . . expresses a physical kind? (p. 134)

To put it formally, at the level of psychology or economics, we have interesting explanatory laws of the idealized form

$$(x)(Fx \rightarrow Gx) \qquad (1)$$

where F and G stand for psychological or economic type terms, identifying real properties which are related as they are because of their intrinsic nature. These generalizations are real laws in part because their arguments are kind terms, reflecting the way the world is actually cut up. The properties to which F and G refer may be realized or implemented at the physical level in various ways, so that F and G correspond in some physical theory to a *disjunction* of complex physical terms; one has F just in case one has (U or V or W); one has G just in case one has (X or Y or Z), where U-Z are terms in physical theory, or at least are physical descriptions. These disjunctions are not (are not likely to be?) kinds. The disjunction (U or V or W)] is not a type term. There are then no type-type correlations among the terms of T_1 and T_2. The physical generalization corresponding to (1) is then

$$(x)([Ux \vee Vx \vee Wx] \rightarrow [Xx \vee Yx \vee Zx]). \qquad (2)$$

But nothing of this form, Fodor says, can be a law or an explanatory generalization just because the arguments of the conditional are not type terms.[9]

[9]Fodor's insistence that the arguments of a *proper law* must designate natural (or social?) kinds sits poorly with an empiricist account of reduction, given empiricist skepticism about kinds. Fodor insists on kinds because he is a scientific realist who thinks that things enter into causal and other relations with each other in virtue of their objective properties, including causal powers, but realists should not use the D-N model, which is motivated by an antirealistic empiricism. On the D-N model explanatory force is a linguistic feature of the theory (that the explanation is a valid argument) and not due to intrinsic characteristics of the phenomena, as a realist would like.

As the informal presentations previously sketched suggest, the force of the argument does not depend on the details of the D-N model, so Fodor's use of it may be regarded as heuristic. Be that as it may, because the disjunctions will not be type terms,

> whether the physical descriptions of the events subsumed by [interesting psychological generalizations] have anything in common is, in an obvious sense, entirely irrelevant to the truth of the generalizations, or to their interestingness, or to their degree of confirmation, or indeed to any of their epistemologically important properties. (p. 133)

Disjunctive generalizations like formula (2) will not properly explain laws like formula (1), and therefore the more basic theories which describe the realization or implementation of less basic theories will not reduce those theories. The two levels are autonomous. We must accept "the disunity of science as a working hypothesis" (Fodor, 1981, p. 120).

Fodor and Pylyshyn's (1988) critique of connectionism as an account of cognition reveals the same structure of argument and may help to make its force more manifest. They argue that psychology qua cognition must be understood, in view of the evidence, as rule-governed manipulation of syntactically structured symbols (that parallel a semantic interpretation of those symbol strings). The classical computer architectures of GOFAI, whose algorithms perform just such manipulations, provide the best model of psychological theory. Connectionist models, which do not operate with rules or structured representations, may be regarded as implementing GOFAI models (insofar as they can do so). But the implementation is at best irrelevant to psychology and at worst misleading and impoverished. First, irrelevance:

> the *implementation*, and all properties associated with the particular realization of the algorithm . . . in a particular case, is irrelevant to the psychological theory; only the algorithm and the representations on which it operates are intended as a psychological hypothesis. (p. 65)

Tienson correctly reads this as invoking a multiple realizability argument. "If we found Martians who satisfied the GOFAI cognitive theory, but not the connectionist implementation, we would (and should) say that the Martians were cognitively, psychologically, like us" (Tienson, 1988, p. 13). Connectionism may or may not be right, but "it couldn't be *psychology*" (Fodor & Pylyshyn, 1988, p. 66).

Second, and quite differently, misleadingness and impoverishment. These objections have nothing to do with the concern about mere implementation. If, taken by the apparently closer analogy of connectionism to brain processes, we regard connectionism as a model of cognition and not a mere implementation of GOFAI programs, they say, we are driven back on a "largely discredited Asso-

ciationist psychology" (Fodor & Pylyshyn, 1988, p. 63), which cannot account for the most important features of cognition. These include: *productivity* (our unbounded capacity to represent indefinitely many propositions), *systematicity* (the way our ability to produce and understand some propositions is intrinsically connected, in virtue of the syntax, to our ability to produce and understand others); and *compositionality* (that each representation makes an approximately similar semantic contribution wherever it occurs). Connectionism, which views mental representation and processing as the activition of interconnected nodes of the unstructured connectionist net, suffers from the same weaknesses in explaining these phenomena as classical empiricist associationist psychology, which used "ideas" or mental images instead of nodes and relations of "similarity" among ideas instead of the connectionists' (presumably electrochemical) links activitating or inhibiting nodes (presumably neurons). "The current attempt to . . . 'take the brain seriously' . . . lead[s]," say Fodor and Pylyshyn, "to a psychology not readily distinguishable from the worst of Hume and Berkeley" (Fodor & Pylyshyn, 1988, p. 64). My concern is not whether connectionists can answer these latter objections, which are good if they can be made to stick,[10] but with the moral of the first objection from mere implementation: physical realization or functional implementation doesn't matter and isn't psychology any more than a calculator's computations are arithmetic.

4. PROVINCIALISM AND REDUCTIVE EXPLANATION

Four avenues of reply show that we could have reduction despite (or instead of) multiple realizability. Strictly they involve rejecting the multiple realizability of the mental as construed by autonomous functionalism.

1. One might doubt, with Kim (1980, pp. 234–236), that "vast heterogeneous disjunctions" would have to be as arbitrary as Fodor suggests. Here we reject Fodor's a priori strictures on admissible natural kinds: What kinds are admissible is a matter for scientists to decide.

2. We might deny that kinds or laws are necessary for reduction, so the heterogeneity of the disjunction would be irrelevant. It can be shown that the reduction base of a natural kind need not itself be a kind; nor that of a law itself a law. The thermodynamic kind term entropy, for example, reduces to a complex set of non-nomological and non-kindlike statistical mechanical facts about the probabilistic tendency of Bolzmann's H to decrease.

3. We might question whether, as a matter of fact, there are that many physical ways of making minds. The plausibility of the multiple realizability argument depends on certain science fiction suppositions that turn out to be questionable, for example, that silicon-based minds are nomologically possible. The usual analogy to silicon-based life suggests that they may not be (see Smart, 1968, p. 111).

These alternatives address whether psychology might have a physical reduction base of the appropriate sort. Avenue (2) turns on rejecting the D-N model of explanation; avenues (1) and (3) grant it for the sake of argument. In Schwartz (1991) I offer more extended criticisms of the D–N model as an account of reduction in particular. In what follows I focus on avenue (4), challenging the conception of psychology invoked by autonomous functionalism. I deny that psychology need be universal or structure-independent. The mental and its physical explanation may be *provincial*, bound to particular physical realizations. Reduction of mentalistic kinds to (kind or nonkind) multiple reduction bases, and even identification of the former with the latter, is legitimate. Generality is *only* a desideratum. It is destructive to make it constitutive of a realm of inquiry.

Because functional predicates pervade every science, the multiple realizability argument should establish the irreducibility of any science that appeals essentially to functionally characterized states. Indeed, it should establish the irreducibility of parts of physics widely regarded as in part already reduced to other, such as thermodynamics, with its functionally characterized notion of temperature, to statistical mechanics. So if such sciences are irreducible in virtue of multiple realizability, many accepted reductive explanations will fail to be either reductive or explanatory. The unity of science will come apart, not only between the social and natural sciences, but among and within the natural sciences.

5. A TALE OF TWO SPECIES

Here I pursue the analogy of psychology with genetics.[11] Watson and Crick's demonstration that the genetic material is a double helix composed of two strands of deoxyribonucleic acid (DNA) bound by complementary pairs of amino acids (bases) is widely thought to be a reduction-sketch of Mendelian genetics. As Hull (1974) said, "given our preanalytic intuitions about reduction, it *is* a case of

[10]Each side of the GOFAI-connectionist debate spends a great deal of time emphasizing the weaknesses of the other and playing up the rather different areas in which each is strong rather than trying to address the problems for their own accounts that emerge in that discussion. Consequently I am unclear on how well connectionists can answer charges like Fodor and Pylyshyn's—or for that matter, how well GOFAI advocates can answer connectionist objections. It seems to me early in the day to say that connectionist models cannot account for the sort of phenomena that concern Fodor and Pylyshyn; less so to say that GOFAI models, which have been around much longer, can deal with the sorts of problems that impel cognitive scientists to connectionism. I predict, though, that the correct account of cognitive architecture will be gerrymandered, with some models dealing better with some processes and other models with others.

[11]Richardson (1979) deploys the analogy to similar effect, using, however, real-life cases of multiple realizability in transmission genetics rather than my science-fiction case of Martian genetics. Moreover, he offers no diagnosis of the appeal of the multiple realizability argument.

reduction, a paradigm case" (p. 44). The double helix model accounts for the major features of Mendelian genetics in purely molecular terms. DNA replicates itself when the strands of the helix separate and each acts as a template for forming a new double helix, allowing for inheritance. Each strand can act as a template for forming other molecules, such as ribonucleic acid (RNA), carrying genetic instructions to cells in the metabolism of a given organism.

If reduction of a function means at least its explanation in terms of its physical structure, Mendelian genetics is in part reducible to physics and chemistry. We can construct, however, an autonomy argument for Mendelian genetics exactly parallel to Fodor's argument for the irreducibility of psychology or for the irrelevance of connectionist implementation. Genes, as noted, are functionally characterized by their causal role; genetic states are functional ones. Human genes are DNA molecules, but this is just a contingent fact reflecting special conditions on earth. Silicon-based Martians have a xeno-molecular genetics. Mendelian transmission genetics applies to them even though their genes are made of ABC. (ABC is chemically distinct but functionally the same as DNA.) So genes are multiply realizable as (DNA or ABC or . . .). Thus genetics cross-classifies physics and chemistry. Say that (DNA or ABC or . . .) is neither a kind nor displays any relevant nonkind de facto commonalities. If we need kinds or de facto commonalities at the physical level to have a reduction, Mendelian genetics is an autonomous discipline, irreducible to chemistry and physics.

If so, a physical explanation of heredity is impossible. Mendelian genetics is independent of physical realization beyond whatever degree of physical complexity is necessary for something to be an organism at all. It matters no more to Mendelian genetics what genes are made of and how they work than it matters to arithmetic or computer science what adding machines are made of and how they work. Genetics merely describes whatever commonalities obtain at some high functional level of abstraction appropriate to the taxonomy that fits the Mendelian laws. Genes may be token-identical to segments of DNA (or ABC or . . .) molecules, but their behavior cannot be explained in molecular terms. If we take this route, whatever it was they did, Watson and Crick did not discover anything of explanatory interest about "the molecular biology of the gene," to use the title of Watson's great textbook (Watson, 1977). Nor did they offer a program for investigating what is normally called the mechanics of heredity, for heredity is an autonomous functional notion in Mendelian genetics. Their work may be viewed as contribution to chemistry, perhaps, but not to biology.

The obverse of autonomy is elimination. If we accept irreducibility and reject autonomy, we might say that Mendelian genetics is not explained but eliminated in favor of molecular biology. What Watson and Crick showed was not that genes are segments of DNA molecules, but that there are no such things as genes. There are just segments of DNA molecules. Heredity, as an object of Mendelian genetics, simply disappears. We have elimination on the "better explanation" pattern articulated by Kemeny and Oppenheim (1970). If we go this way, though,

it is unclear why a similar story would not apply to the phenotypic characteristics Mendelian genetics explains, such as eye color or biological sex. Do we want to deny that there are eyes with particular colors or that there are organisms with particular sexual characteristics? Strictly the issue is whether these characteristics are phenotypes. We might not have to give up eye color or sex, but we would have to give them up as phenotypes. But this would be giving up quite a lot. It would give us at least a start on the full-fledged elimination of the manifest world of macroscopic objects advocated by Churchland (1979). We may end up with nothing but point masses in space-time.

6. CONSTITUTION AND IDENTITY IN GENETICS

Neither of these alternatives are attractive or accord with ordinary scientific judgment uncontaminated by philosophy. Some, like Churchland, may revel in elimination. Others bite the biological bullet on autonomy (e.g., Rosenberg, 1985). But more plausibly, molecular biology does reduce Mendelian genetics. If Watson and Crick did not show how to give physical explanations of heredity, what did they show? We need not leave the issue with a rhetorical question. It seems that the physical properties of DNA described by molecular biology explain features of transmission genetics, for instance, why inherited characteristics show up in the ratios they do in descendent generations; how it is that some characteristics are recessive and others dominant, and so forth. The realization or implementation of genetics in DNA is explanatory if anything is. If we accept that molecular biology provided a reduction of or a guide to producing reductions in transmission genetics, there seem to be two possibilities. Either

(a) reduction to different physical bases shows that we do not have the same functional state,

or

(b) whether we have a physical reduction of a kind of functional state cannot depend on whether it is always to the same physical basis.

In the first option we have parallel classic type-type identities holding between similar but distinct functional states and their various reduction bases. In the second we have reduction of what is in some sense the *same* functional state to multiple reduction bases.

Call the sort of explanation involved in cases like the genetic one *provincial*. Such explanation typically involves cases where the specific mechanisms that constitute the macrophenomena which interest us differ across classes of individuals. The claim that provincial explanation is explanation is at least that it is explanatory to appeal to the particular mechanisms that are sufficient but not

necessary for that phenomenon to occur.[12] To have an explanation, reductive or other, it is enough to state the mechanism by which some phenomenon is produced or constituted. It is not necessary that this be the only mechanism by which that phenomenon could be produced or constituted, or even that it be related in interesting ways to all the other such mechanisms. Reduction here is not bound to the empiricist sense of the term. It is rather a matter of explicating the mechanisms that produce or constitute some state, process, or entity, an enterprise that may or may not be deductive or nomological.[13]

Whether either of the two options, (a) or (b), hold in any case depends largely on the facts of the matter in the particular instance. Suppose Martian hereditary material is made of ABC, functionally identical but structurally distinct from DNA. If we insist that genes are *type-identical* to segments of DNA, having just the properties of DNA, Martians will not have genes. No matter: they will have *schmenes*, which are just as good as genes for the purpose (the explanation of inherited characteristics). Schmenes obey the laws of Mendelian genetics, except that when we wish to be pedantic we replace the expression *gene* with *schmene* throughout. Generally, the difference can be ignored, and we can say that Martians have genes which are governed by Mendelian genetics.

If, on the other hand, schmenes are triple-helixed EFG (a genetic material both physically and functionally distinct from DNA), giving different ratios of inherited characteristics in descendent populations, schmenetics will not be Mendelian: the transmission laws will be different. Here we cannot ignore the difference between genes and schmenes for practical purposes because it shows up on the functional level. (Note that there is a physical explanation for this fact.) In either case, holding the DNA-gene identity constant, biology will be provincial, in that we will have a set of class-specific biologies that vary, roughly, with the sort of conditions under which various kind or classes of organisms evolved.

Alternatively, we might give up the troublesome premise that genes are strictly identical to DNA in the sense of sharing all properties, including physical ones. Here, humans and Martians would have genes (the same sorts of things), and genes would be the same as DNA in humans but the same as (ABC or EFG) in Martians. Call the relation of sameness that applies here *constitution:* the same thing may be constituted or made up from different materials. Identity is a limit

[12]I am not offering an "account" of explanation as the provision of a sufficient but unnecessary condition, or indeed a general account of explanation at all, although were I to attempt such an account it would be a mechanical one; that is, I think that explanation is elucidation of causal mechanisms. Here I am offering a very partial (though I hope adequate for the purpose) characterization of the provincial aspect of a certain sort of reductive explanation.

[13]For this notion of explanation, applied to probabilistic rather than reductive contexts, see Railton (1981) and Salmon (1984, 1989). Railton attempts to revive a sophisticated version of the D-N model in this scientific realistic context; I think laws are nice if you can get them but their absence does not deprive an account of a mechanism that produces or constitutes some phenomenon of explanatory power.

case of constitution in which a thing or property can be constituted only one way. For other cases we give up Leibniz's law (that identical things share all properties) and allow that the same thing may have different microstructural properties. Which way it might turn out is an empirical question, and because all the genes we know are DNA, highly speculative.

How we go in a particular case, once the facts are specified, depends on the theoretical advantages of choosing one rather than the other approach. Course (b), opting for constitution, has the advantage of ontological simplicity: we don't multiply kinds of hereditary material at the functional level. If hereditary material is functionally identical to DNA or ABC, this course seems attractive. On the other hand, constitution might seem a more obscure and less familiar notion than identity—although hardly less puzzling (see Hirsch, 1982). So if hereditary material is functionally distinct DNA or EFG, we might prefer to mark the double difference (functional and physical) by choosing option (a), insisting on identity. I'm not sure that a great deal rests on the choice once we know the facts. It would be up to the scientists to decide what to say.

Either way, the possibility that genes may be differently constituted does not block explanatory appeal to a particular constitution in a particular case, such as human eye color or biological sex. That is, we can say it explains why someone has blue eyes and is biologically female that they have the genes (or, technically, cistrons) for blue eyes and have XX chromosomes, even if that explanation doesn't work for blue-eyed *Martian* women. Insofar as such appeal counts as explanation of the phenomenon at a more basic level, we have a reduction or at least a reduction program. This is so even if (nonidentificatory) constitution blocks nice type-type connections or multiple parallel identities block a universal genetics.[14]

7. GENETICS AND PSYCHOLOGY

If "multiple realizability" does not show that genetics is independent of physical realization, it does not show that psychology is either. It might rather show that the functional character of mental states permits reduction, at least in the sense that genetics is reducible. Suppose there were good type-type connections be-

[14]A reduction program is a research program for finding reductive explanations based on the use of certain techniques (e.g., those of molecular biology or cognitive neuroscience) and based on exemplary explanations already achieved in the domain. All real examples of reduced sciences are reduction programs based on clusters of promising reductive explanations of certain phenomena. There is not—nor will there be—a single case in which we attain the positivist ideal of a complete reduced science, all of the laws and predicates of which are derived via bridge-laws from a more basic science, and the whole expressed in nice axiomatic form. To deny that a science is in fact reduced because what one has is a reduction program instead of a philosopher's fantasy is to legislate to science, which is not the philosopher's mandate.

ween human mental and physical states, and likewise with Martians, but no nice ross-species fit. Should we conclude psychology is irreducible or just that it has ifferent reduction bases? Should we search only for an abstract functional psyhology that describes whatever humans and Martians have in common, or do we also) look for psychological generalizations that apply to each species, and for he physical basis of these in the particular cases?

The classic type identity thesis can be preserved, then, at the cost of the claim iat beings of different physical constitutions could share identical mental states. his is option (a): physical difference is sufficient for denial of functional ameness. The mental state that correlates with human C-fiber firings is identical pain; that which correlates with Martian Z-fiber firings is identical to *schmain*, nd so forth, giving parallel reductions of the provincial psychologies of various orts of cognizers to the particular reduction base that realizes them. Mental ates would not be multiply realized at all. They would be uniquely realized in articular types of physical states. Neurophysiology and psychology would coary among classes of cognizers.

We would of course be interested in functional commonalities or similarities s corollary (not an alternative) to a reductive project. To say that Martians had a ifferent psychology is not to say that they have no psychology, any more than to ay that they have a different biology is to say that they have no biology. The roposal, then, is not chauvinist. On option (a) we need not even give up a niversal psychology to predict, explain, and otherwise interact with nonhuman ognizers. We may abstract from or ignore the differences for certain purposes—ay, moral ones. What Martians feel when they are burned or stabbed is techically schmain, not pain, but since schmain is aversive and horrible, we can eat schmain as the moral equivalent of pain, and say that it is wrong to cause eedless pain to people, including Martians.

As in the biological case the plausibility of this move depends in part on the cts. Option (a) is most plausible for the mental if Martian psychology is quite ifferent from ours functionally as well as physically, as with the EFG heredity se. If Martian biology differs radically from ours in the right ways, Martian sychology probably would as well. Imagine that Martian sensory modalities, nd thus the character of their experience, is quite alien. They directly perceive iagnetic fields or magnetic variations as we do light and are "visually" sensitive electromagnetic radiation only in the FM band. They have three sexes to go ong with their triple-helixed schmenes and so the generalizations that describe eir sexual impulses are very different from the ones that describe ours. Here we ight give up the idea that they have the same mental states as humans—thus eeping the classic identity thesis, although this might not matter for many rposes, for instance communication of scientific results, negotiation about olitics, arranging tours of Martian "artists" (as we would call them), and the ke.

If, however, Martian psychology was functionally just like ours (although

realized in silicon-based neural tissue), we might choose option (b): that functional sameness admits of reduction to different physical bases. This would involve giving up type-identities for relations of constitution, the classical identity thesis for *constitutive* materialism. On this view, differently constituted beings could share the same mental states in a nonidentificatory sense of *same*. This allows multiple realizability, in a sense, but it allows reductive explanation too. Types of mental states would not be type-identical to types of physical ones, but neither would the former be merely token-identical to the latter. Rather mental types would be constituted variously but systematically by different sorts of physical types, allowing us to explain a psychological state or process in physical terms within each class of beings, although not universally. In either case, psychology would be universal; reduction, provincial.

Either approach allows us to preserve reductionism for psychology in the same sense that we have it for genetics. If a Watson and Crick of the mind tomorrow publish the *psychological code*, revealing the kind of physical mechanisms that constitute most human thought, it would be a reduction even if there are large areas of thought and behavior to which the psychological code has no direct application, as there are large areas even of genetics (e.g., population genetics) where molecular genetics has no direct application, and even if most of the phenomena in the domain of the psychological code theory are too complex to produce explanations in its terms, as is the case with genetics.

The same considerations apply to connectionism as an implementation of GOFAI or indeed of commonsense psychology. At present it would be premature, to say the least, to claim that connectionism is the psychological code or to propose Rumelhart and McClelland as psychology's Watson and Crick. Unsurprisingly, given the recency of the revival of connectionist models, we do not have a connectionist theory nearly as well articulated as classical computational theory. What we have instead are exciting and suggestive connectionist models of specific processes, like Rumelhart and McClelland's (1986, Vol. 2, pp. 216–271) model for learning the past tense of English verbs, which appears to simulate the kinds of errors and progress humans make in this sort of syntactic learning. (Pinker & Prince, 1988, criticize the adequacy of the model.)

Suppose, though, that Fodor and Pylyshyn are right that such models merely implement, in human brains, cognitive processes correctly described by GOFAI models. That is no reason to deny that connectionism explains cognition in humans, or to assert that connectionism isn't psychology. It may be only human psychology, but what's wrong with that? In any particular case, say one involving nonconnectionist Martians, we would be faced with a choice structurally similar to the one previously discussed. We can say either that human psychology is connectionist but Martian psychology isn't (option [b]) or that humans have psychology, since our cognition is connectionist, but Martians have *schmychology*, since theirs isn't (option [a]). Which we might wish to say depends in part on whether the implementation makes a difference at the functional level. Similar

reasoning applies, *mutatis mutandis*, if radical connectionists (see Cummins & Schwarz, 1988) are right that connectionism should displace GOFAI as our account of cognition and be construed instead as a direct implementation of or explanation for commonsense psychology.

8. UNFAIR TO ALIENS?

The biological argument shows that there may be no universal transmission genetics. It will be objected that this conclusion is to be resisted for psychology. One can argue that because of its subject matter (intentional behavior), psychology is universal in a way that biology is not. Provincialism misses out just what it s we and the aliens have in common, namely our mental states. Fodor (1981) vrites:

> We could, if we liked, *require* the taxonomies of the special sciences to correspond to the taxonomy of physics by insisting on distinctions between the kinds posited by the former whenever they correspond to distinct kinds in the latter. That would *make* the laws of the special sciences exceptionless if the laws of the basic science are. But it would also likely lose us precisely the generalizations we want to express. (p. 143)

he generalizations of psychology apply universally because beings of various onstitutions can be in the same states governed by the same psychological eneralizations. Exceptions can be explained away nonpsychologically by appeal more basic sciences.

A good deal turns on what counts as *the same* mental state. Any functionalist ill individuate a mental state by its causal role in mediating sensory input and ehavioral output, such that same role, same state. If difference in reduction ases makes no difference in role, then human and alien mental states are the me in the relevant sense of *same*, that is, they play the same functional role. his is option (b), holding that functional sameness is enough for sameness, hatever implementation explains the operation of the function in a given case. *ce* Fodor, it explains why a particular cognizer has a given mental state to lvert to its parochial physical constitution, although such an explanation may t be to the point in a particular pragmatic explanatory context. But we retain hatever functionally expressed generalizations we want the special (here, intenonal) sciences to give us.

Suppose, though, that psychology initially fails to reflect functional differences that show up under physical description, but the physical theories that ver the behavior of the correlates of a state "shared" by humans and aliens

produce different predictions about or explanations of their respective behavior.[15] Fodor suggests that we treat the psychological failure as an anomaly to be physically explained: psychology is immune from revision in the face of physical evidence. This is not a contradiction: in explaining away the physically induced deviations, we cease to do psychology.[16] But if the "generalizations we want to express" fail to capture important differences in behavior or its causes, they are false, in which case we may not want to express them. The alternative—option (a)—is to say that we have different mental states because their implementation is physically different. For some purposes, strict falsity may not matter and we can treat the generalizations as idealizations. In that case we do not lose them.

9. SAMENESS OF CONTENT

The objection may be sharpened as follows. Provincialism, it might be said, requires us to abandon the intuitively plausible claim that beings of varying physical constitution can share intentional states *of the same content*. In 1974, Cornell astronomers at Arecibo Observatory in Puerto Rico sent a radio telescope message to star cluster M13, which says—has as its content—among other things, the proposition (call it p) that "there are about four billion intelligent creatures whose genetic basis is DNA on the third planet from a given star" (Sagan, 1979, p. 321). The point was to produce in the aliens in M13 the same belief that we had. On my account, though, if Martians are silicon-based, it appears that they cannot entertain our belief that p. But surely our intuition that they can share our belief that p is far stronger than our intuition that beliefs are identical to some type of neural structure! If one of the intuitions has to go, it is the reductionist one that backs provincialism.[17]

The relevant disanalogy with biology is in the reliance of psychology on the notion of *content*. This (it is said) has no biological parallel. The universality of

---

[15]It is not unusual for a reduced theory to differ from its original formulation. Classical thermodynamics was anomalous because unlike the rest of classical physics it was not time-reversal invariant. Under statistical mechanical reduction, the anomaly vanished. Finding such improvements, often inaccessible at the level of the special science, is a motivation for seeking reduction.

[16]Here Fodor sounds more like Davidson than he might like. Fodor insists that his irreducibility claim is empirical, while Davidson (1980) purports to have an in-principle argument for his anomalous monism. I suspect that Fodor's attachment to a universal psychology is as a priori as Davidson's to the universality of decision theory.

[17]This objection depends on accepting (as Fodor and I do) that psychological explanations may advert to content. I don't think, though, that Fodor can maintain both this and his methodological solipsism (see footnote 2). The objection is not open to Stich, who maintains that psychological explanations must be purely syntactic and content-free—a thesis, popular among cognitive scientists, sometimes called "the formality constraint."

psychology implies that if the aliens are intentional agents they must be able to instantiate the same intentional contents as humans. This a charitable reading of Fodor's (1981) claim, quoted above, that reduction "would lose us precisely the generalizations which we want the special sciences to express" (p. 143), that is, generalizations framed in terms of sameness of propositional content.[18]

One reply derives from a naturalized account of content like that defended by Dretske (1988). We can preserve identity of content by making content depend not on the particular internal physical state that realizes an intentional attitude, but on the complex sort of causal covariation that, if some suitably intricate causal theory of content is right, counts as representing that p. Two intentional states will have the same content if they are causally related in the right way to whatever they represent. Thus we and the aliens may be said to have the same belief that p in virtue of the fact that both of our internal states adjust in appropriately similar ways in response to a causal input from the same source. That the physical realizations of the functional states may differ qualitatively would be either irrelevant, if what matters is that they play the same functional role in adjusting in response to input, or no more relevant than the fact that each instantiate numerically different instances of the representation.

Another answer (explored in Schwartz, in press) presupposes no controversial theory of naturalized representation. It is controversial in another way. Here we bite the bullet and allow that humans and differently constituted aliens would not share content. This need not be a costly concession. For most purposes we could ignore the differences and proceed as if the content were the same, that is, ascribe the same contents *ideal typically*. The expression is due to Max Weber (1949), who proposed that rational actor explanations in sociology ought to be made in this way. People's motivations often deviate considerably from economic rationality, as Weber knew, but he suggested that for sociological purposes the best methodology was to *abstract* from the various heterogeneous and irrational motivations and, where possible, to use the ideal type of the rational actor as the basis for explanation of action.[19] I suggest that we may abstract similarly from fine differences in content.[20]

[18]The uncharitable reading is that the same intentional generalizations should be valid for all intentional agents.

[19]Weber (1949) writes: "An ideal type is formed by the *one-sided accentuation* of one or more points of view and by the synthesis of a great many diffuse, discrete, more or less present and occasionally absent *concrete individual phenomena*, which are arranged according to those one-sidedly emphasized viewpoints into a unified *analytical* construct. In its conceptual purity, this (analytical) construct cannot be found anywhere in reality" (p. 90).

[20]The idea is similar to (and inspired by) the Putnam-Burge thesis that "meanings ain't in the head," where that means at least that content ascription depends on factors external to the agent's internal states. See Schwartz (in press, section 10.4) for discussion and development of this thesis. Burge's (1979) variation on the idea shows, the proposal does not require, although it is not inconsistent with, either a causal or a naturalized theory of content.

This may be what we in fact do in ascribing propositional attitudes to humans. A growing body of evidence suggests that the propositionalist "language of thought" model that, as Fodor argues persuasively, is embedded in our commonsense psychology, is false as an account of human cognition.[21] According to propositionalism, thought is structured like a natural language, a notion that is articulated in GOFAI models of cognition. Connectionism is one way of providing a systematic alternative to the idea that cognition is a structure of propositions, seeking insight from neural structure in a way that Fodor (for one) thinks is impossible. Rather than crunching structured symbols according to rules, connectionists claim, in thinking we are "processing activition vectors through artfully weighted networks" (Churchland, 1989, p. 195). The case for connectionism—its superior performance on a number of dimensions and in a number of areas where GOFAI is weak—supports the anti-propositionalist orientation. The results of Kahneman, Slovik, and Tversky (1982) or Nisbet and Ross (1980) can be interpreted as showing that people do not think in logic or work with mental sentences at all. Johnson-Laird (1983) has proposed that people operate functionally not on propositions but on *mental models* which may be thought of as three dimensional quasi-images. It may be possible to integrate these cognitive ideas into the brainlike artificial intelligence models of connectionism.

Our actual mode of representation "in the head" probably diverges from the language of thought model on several dimensions (syntactic and inferential, as I argue in Schwartz (in press)). No matter. It is good enough for the purposes for which we use it to serve us well. Failing a Churchlandian conversion to self-ascription terms of neurophysiology (which might not serve our purposes), we abstract from whatever may actually be in the head and use our current and evolving psychology. As among ourselves, so with nonhuman cognizers. Communication, explanation of behavior, and so forth, is possible as long as we and they ascribe to each other propositional attitudes. We might learn each other's neurophysiology and try to communicate in those terms, but why make life difficult? Reference to propositions is pragmatically convenient, and good enough for most purposes—speaking with one's spouse, persuading the electorate, discussing mathematics with aliens.

This suggests a different construal of the universality of psychology. Psychology is universal insofar as it is ideal-typical. As soon as we cash out (i.e., reduce) the ideal-typical ascriptions we make in terms of the actual mechanisms that produce behavior, it goes provincial. The sense in which we may hope for a universal psychology is just that for many purposes, the differences won't matter much. We could then safely abstract from them. So, although perhaps philosophically drastic, the revision would be practically minimal. For some scientific purposes we would have a better understanding of the psychologies of various

[21]It's controversial whether a language-of-thought thesis correctly captures or commonsense psychological ascriptions. Stich (1983) and Fodor (1975) offer strong cases that it does.

sorts of beings, and that would make about as much difference in our ordinary lives, or in scientific work that is peripheral to cognitive psychology, that deep scientific theory usually does—not necessarily very much.

10. NIHIL HUMANUM

The multiple realizability argument for autonomous functionalism owes most of its power not to the technical details of its mistaken empiricist model of reduction but to the antichauvinist appeal of the universality of psychology. The deep presupposition is that anything that deserves to be called psychology must apply in some sense in the same way to any cognizer whomsoever, however different physically such a being might be from us. A similar idea motivates the dismissal of connectionism as "mere" implementation of GOFAI as a universal model of cognition. We should resist the Siren call of such "antichauvinism." After all, it is human beings we are interested in when we do psychology and social science. We are especially interested in understanding, explaining, and predicting human behavior because it is our behavior, and what we do matters a great deal to us. If theories developed for these purposes happen to fit aliens or computers, that's all to the good. But we should not hobble our development of such theories, including their development through reduction, by insisting that they must do so.

Although the view urged here is not chauvinist in that it does not deny mentality to the differently-constituted, it is provincial in that it sets the primary task to hand of the intentional sciences to be the explanation of human behavior. This sort of provincialism does not "lose us precisely the generalizations which we want the special sciences to express" (Fodor, 1981, p. 143), I have pursued an analogy with biology; taking a cue from Fodor, I turn to an analogy with economics.

If an economic theory that applies to our own society fails to be adequate to all societies in all times and places, do we reject it as insufficiently general? If a good candidate for an absolutely general economic theory came along we would jump at it. But we do not insist that any economic theory meet such standards, or we'd be rather short on economic theory. Likewise, a good human psychology might not be all we want, but it would be nice to have one. If human psychology, as reduced to its provincial physical basis, turns out not to apply to Martians because their provincial physical basis is different, this no more casts doubt on the explanatory force of human psychology than the failure of neoclassical economics to describe feudal economies casts doubt on its explanatory force for capitalist ones.[22]

[22]Other things may cast doubt on the explanatory force of neoclassical economics: Institutionalist economists like Thorsten Veblen and Marxist political economists object that its abstractions leave out too much that is important and relevant to the understanding of economic phenomena, such as class relations and aspects of the organization of the production process. But this is a quite different objection from the one that economics may not be universal; rather the problem is supposed to be that it is bad in the particular case. Fodor and Pylyshyn's objection that connectionism is misleading and impoverished, failing to explain too much that matters, is an objection of this sort, quite distinct from their a priori and universalist claim that implementation is irrelevant. If the former objection can be made to stick (something I don't consider), connectionism is doomed—but because it is misleading and impoverished, not because it is merely an implementation.

Given this interest in human beings, the demand of absolute generality raises the following question. The antichauvinist intuition suggests that *intentional agent*, where that includes the class of entities that could have a psychology, is a kind larger than that of human beings. If we make their commonalities constitutive of psychology, could it turn out that psychology doesn't apply to homo sapiens? The point is not that the demand for absolute generality makes an autonomous psychology a priori, but that the a priori demand for generality makes it an issue about how well such a psychology will do for humans. Similarly, an absolutely general biology that fits any possible form of life (carbon- and silicon-based, etc.) might fail to describe much of what interests us about any particular biology.

On the economic analogy, the possibility that an abstract universal psychology might fail to fit humans is far from implausible. According to one view, held in different ways by von Mises (1979) and Friedman (1979), economics is the logical deduction from certain a priori axioms of conclusions about the behavior of highly abstractly described rational actors.[23] Now economics, so regarded, may not have much to do with what goes on in banks and factories. But if what goes on in banks and factories is what interests us, we should keep far from economics and study what goes on in banks and factories.[24] Or we might say that economics should study just this. Such a response turns on a different set of interests and a different—a scientific realist—conception of science from Friedman's or von Mises's. These concerns may be related. We may reject instrumentalism (Friedman, 1979) and a priorism (von Mises, 1979) because we think that it is part of the job of science to describe the actual mechanisms that produce or constitute the phenomena that interest us and because we think that more accurate knowledge of these mechanisms, among other things, will lead to better predictions.

[23]See von Mises (1979, p. 64); Friedman construes economic theory (and scientific theory generally) in an austerely instrumentalist way, as a "set of tautologies" to be evaluated by the criteria appropriate to judging a "filing system" (Friedman, p. 21). An important difference is that Friedman thinks that empirical adequacy matters to economics, while von Mises does not. Economic science is a "mental experiment . . . involv[ing] thinking through the implications of a proposition in the light of its compatibility with other propositions we accept as true [whether or not these] make reference to experience" (von Mises, p. 61).

[24]This is precisely what Marxists and Institutionalists say about neoclassical economics. Interestingly a common neoclassical economist's reaction is to reply that the research done by Marxists and Institutionalists isn't economics—at best it's sociology. "Real" economics, neoclassicists say, is mostly abstract mathematics describing the behavior of idealized rational actors under constraints.

If psychology, as understood by autonomous functionalists, isn't about humans, let's study what is about humans and see how it fits with physics, neurophysiology, and the other sciences. Or we might call the study of human behavior psychology and say that the abstract theory (GOFAI? propositional attitude psychology?) which applies to the broader kind—all cognizers—may deserve the name, but this enterprise, human psychology, certainly does. And here the implementation or particular realization of human psychology in physical brains and (perhaps) connectionist networks may matter a lot, both in suggesting new avenues for research and in solving puzzling problems that resist approaches at a higher level of abstraction. No matter if this emphasis on the particular undermines universality because it may not apply to Martians. Let's not take anti-chauvinism so far that we deny what is important to us on behalf of beings who may not even exist!

We may be able to do better than a provincial psychology. It is possible that there is a universal psychology, and perhaps it is a propositional attitude psychology best modeled by GOFAI. I have suggested that we could construe such universality as a function of idealization. It is even possible that there is, perhaps as a matter of physical fact, a unique realization or implementation for anything that might count as being a full-fledged psychology. These are empirical questions. My plea is a conditional one for the legitimacy of provincialism. Given our interests and general desiderata about what can count as a psychological explanation, or a physical reduction thereof, provincial explanations and reductions are fine if they are all we can get. Moreover: it is a constraint on any more general such explanations that they be reasonably close approximations to the provincial ones. What is human had better not be alien to us.

The prospect for reduction is open. Materialism need not be stripped down to the mere token identity theory. The truth of the classic identity theory or of constitutive materialism would depend on empirical results, but multiple realizability defeats neither of them. Neither does multiple realizability deprive connectionist models of explanatory interest with regard to GOFAI or commonsense psychology, supposing that connectionism can avoid the objections of misleadingness and impoverishment. The multiple realizability argument is the main case for autonomous functionalism and the irrelevance of implementation, so the orthodoxy should be rejected—the autonomy claim, that is, and not the functionalism. As a theory of mind and as research program in psychology, functionalism has nothing to do with multiple realizability.

ACKNOWLEDGMENTS

Thanks are due for helpful comments to John Dinsmore, Alan Gibbard, Don Herzog, David Hills, Peter King, Robert McCauley, Diana Raffman, Peter Railton, and Laurie Stowe.

REFERENCES

Armstrong, D. M. (1968). *A materialist theory of the mind*. London: Routledge & Kegan Paul.

Bechtel, W. (1988). Connectionism and the philosophy of mind: an overview. In T. Horgan and J. Tienson (eds.), Connectionism and the philosophy of mind: Spindel conference 1987 [Special issue]. *Southern Journal of Philosophy XXVI*, 17–42.

Block, N. (1980). Troubles With Functionalism. In N. Block (Ed.) *Readings in philosophy of psychology* (Vol. 1, pp. 268–305). Cambridge MA: Harvard University Press.

Burge, Tyler. (1979). Individualism and the mental. In P. A. French, T. Uehling, Jr., & H. Wettstein (Eds.), *Midwest studies in philosophy IV. Studies in metaphysics* (pp. 73–122). Minneapolis: University of Minnesota Press.

Churchland, P. M. (1979). *Scientific realism and the plasticity of mind*. Cambridge: Cambridge University Press.

Churchland, P. M. (1989). *A neurocomputational perspective*. Cambridge MA: MIT Press.

Cummins, R. & Schwarz, G. (1988). Radical connectionism. In T. Horgan & J. Tienson (Eds.), Connectionism and the philosophy of mind: Spindel conference 1987 [Special issue]. *Southern Journal of Philosophy XXVI*, 43–62.

Davidson, D. (1980). *Essays on actions and events*. Oxford: Clarendon Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge MA: MIT Press.

Fodor, J. (1975). *The language of thought*. New York: Thomas Crowell.

Fodor, J. (1981). *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge MA: MIT Press.

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. In S. Pinker & J. Meler (Eds.), *Connections and symbols* (pp. 3–72). Cambridge MA: MIT Press.

Friedman, M. (1979). The methodology of positive economics. In F. Hahn and M. Hollis (Eds.). *Philosophy and economic theory* (pp. 18–35). Oxford: Oxford University Press.

Haugeland, J. (1985). *Artificial intelligence: The very idea*. Cambridge MA: MIT Press.

Hirsch, E. (1982). *The concept of identity*. Oxford: Oxford University Press.

Hull, D. (1974). *Philosophy of biological science*. Englewood Cliffs, NJ: Prentice-Hall.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, UK: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kemeny, J. G. & Oppenheim, P. (1970). On reduction. In B. A. Brody (Ed.), *Readings in the philosophy of science*, (1st ed., pp. 307–318). Englewood Cliffs NJ: Prentice-Hall.

Kim, J. (1980). Physicalism and the multiple realizability of mental states. In N. Block (Ed.), *Readings in philosophy of psychology* (Vol. 1, pp. 234–236). Cambridge, MA: Harvard University Press.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace, & World.

Nisbett, R. & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. In S. Pinker & J. Meler (Eds.), *Connections and symbols* (pp. 73–194). Cambridge MA: MIT Press.

Putnam, H. (1975). *Mind, language, and reality: Philosophical papers* (Vol. 2). Cambridge, UK: Cambridge University Press.

Railton, P. (1981). Probability, explanation, and information. *Synthese, 48*, 233–256.

Richardson, R. (1979). Functionalism and reductionism. *Philosophy of Science, 46*, 533–558.

Rosenberg, A. (1985). *The structure of biological science*. Cambridge, UK: Cambridge University Press.

Rumelhart, D. E., McLelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. (Vols. 1 & 2) Cambridge MA: MIT Press.

Sagan, C. (1979). The quest for extraterrestrial intelligence. In *Broca's brain: Reflections on the romance of science* (pp. 314–325). New York: Ballentine Books.

Salmon, W. C. (1984). *Scientific explanation and the causal structure of the world.* Princeton, NJ: Princeton University Press.

Salmon, W. C. (1989). Four Decades of Scientific Explanation. *In* P. Kitcher & W. Salmon (Eds.) *Scientific explanation: Minnesota studies in the philosophy of science XIII* (pp. 3–219). Minneapolis: University of Minnesota Press.

Schwartz, J. (in press). Propositional attitude psychology as an ideal type. *Topoi.*

Schwartz, J. (1991). Reduction, elimination, and the mental. *Philosophy of Science, 58,* 203–220.

Smart, J. C. C. (1968). *Between science and philosophy: An introduction to the philosophy of science.* New York: Random House.

Smart, J. C. C. (1971). Sensations and brain processes. In D. M. Rosenthal (Ed.) *Materialism and the mind-body problem* (pp. 53–66). Englewood Cliffs, NJ: Prentice Hall.

Stich, S. (1983). *From folk psychology to cognitive science: The case against belief.* Cambridge, MA: MIT Press.

Tienson, J. (1988). Introduction to connectionism. In T. Horgan & J. Tienson (Eds.), Connectionism and the philosophy of mind: Spindel conference 1987 [Special issue]. *Southern Journal of Philosophy XXVI,* 1–16.

von Mises, L. (1979). The science of human action. In F. Hahn & M. Hollis (Eds.), *Philosophy and economic theory* (pp. 57–64). Oxford: Oxford University Press.

Watson, J. (1977). *The molecular biology of the gene* (3rd ed.). New York: Benjamin.

Weber, M. (1949). *The methodology of the social sciences* (E. A. Shils & H. A. Finch, Ed. & Trans.). New York: Free Press.

6 Exploring the Symbolic/Subsymbolic Continuum: A Case Study of RAAM

Douglas S. Blank
Lisa A. Meeden
James B. Marshall
Indiana University

1. INTRODUCTION

It is difficult to clearly define the symbolic and subsymbolic paradigms; each is usually described by its tendencies rather than any one definitive property. Symbolic processing is generally characterized by hard-coded, explicit rules operating on discrete, static tokens, whereas subsymbolic processing is associated with learned, fuzzy constraints affecting continuous, distributed representations. In addition, programming languages such as LISP and mechanisms such as Turing machines are typically associated with the symbolic paradigm, whereas connectionism is frequently associated with the subsymbolic paradigm. Debates contrasting the two paradigms sometimes center on these mechanisms, for example comparing the capabilities of Turing machines with those of connectionist networks (see Adams, Aizawa, & Fuller chap. 3 in this volume). However, connectionist networks can be proven to be computationally equivalent to the abstract notion of Turing machines (Franklin & Garzon, 1990). Therefore the computational mechanism is not the crucial issue in separating the symbolic and subsymbolic paradigms. What then is the crucial issue?

We believe there are three major issues that distinguish the symbolic paradigm from the subsymbolic paradigm: (a) the type of representations; (b) the style of composition; and (c) the functional characteristics. We have summarized the key elements of these differences between the two paradigms in Table 6.1. However, most cognitive science and classical Artificial Intelligence (AI) models cannot be completely characterized as either purely symbolic or purely subsymbolic using these criteria. Instead, most models fall somewhere in between the two extremes, or in the so-called *Gap*. For this reason, it seems appropriate to view the para-

The Cognitive Science Series: Technical Monographs and Edited Collection

Donald A. Norman, Andrew Ortony, and Roger Schank, Series Editors

---

Gentner/Stevens: Mental Models

Hinton/Anderson: Parallel Models of Associative Memory, Updated Edition

Norman: Perspectives on Cognitive Science

Dinsmore: The Symbolic and Connectionist Paradigms: Closing the Gap

The Symbolic and Connectionist Paradigms
Closing the Gap

Edited by

John Dinsmore

Washington University

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS

1992 Hillsdale, New Jersey Hove and London