THE JOURNAL OF PHILOSOPHY

VOLUME LXXVII, NO. 4, APRIL 1980

A PERSPECTIVE ON MIND-BRAIN RESEARCH *

THERE is a fairly widespread conviction among philosophers that direct study of the brain is not likely to be very fruitful in the endeavor to get a theory of how the mind-brain works. Surprisingly, this opinion concerning brain research appears to be shared among philosophers who otherwise have starkly incompatible views about the nature of the mind. On the one hand are the dualists and phenomenologists who take the mind to be a nonphysical entity and who believe that mainly by introspection can we come to understand such things as perception, memory, consciousness, and the like. Their reservations concerning the value of brain research are understandable, however lamentable. Less easy to understand are the reservations of some physicalists. Yet it is apparent that many physicalists, including not only philosophers,[1] but also psychologists[2] and artificial-intelligence researchers,[3] are frankly skeptical concerning what neuro-

* I am particularly indebted to Paul Churchland, Michael Stack, Larry Jordan, and Daniel Dennett for inspiring conversation.

[1] Cf. J. A. Fodor, *The Language of Thought* (New York: Crowell, 1975); Hilary Putnam, "The Nature of Mental States," in David M. Rosenthal, ed., *Materialism and the Mind-Body Problem* (Englewood Cliffs, N.J.: Prentice-Hall, 1971); Daniel C. Dennett, "Artificial Intelligence as Philosophy and as Psychology," in his *Brainstorms* (Montgomery, Vt.: Bradford, 1978), pp. 109–126.

[2] Cf. John R. Anderson, *Language, Memory and Thought* (Hillsdale, N.J.: Erlbaum, 1976). Anderson says:

> Obviously one would be foolish to claim that physiological data will never be of use to the cognitive psychologist. However, one would be equally foolish to expect that physiological data will help decide current issues about internal structure and process (14).

See also Zenon Pylyshyn, *Cognition and Computation* (in progress), and Julian Jaynes, *The Origin of Consciousness in the Breakdown of the Bicameral Mind* (Boston: Houghton Mifflin, 1976), p. 18.

[3] Cf. Joseph Weizenbaum, *Computer Power and Human Reason* (San Francisco: W. H. Freeman, 1976).

0022-362X/80/7704/0185$02.30

science can tell us about how the mind-brain works. Their reservations are prima facie startling, in light of the fact that the central tenet of physicalism is that mental states are physical states, and given that we humans have brains, our mental states are brain states. This skeptical attitude on the part of physicalists has frequently come to my notice insofar as my own studies in neuroscience have often been met with bemused incredulity, as though knowledge of neuroscience was probably no more nor less valuable to a philosopher of mind than, say, knowledge of cartography or soil science—it might provide entertaining and even useful examples, but it will not do more. I have therefore tried to inquire into the nature and justification of the physicalist's skepticism toward the value of neurological research. In so doing, I have discovered that this skepticism should really be distinguished into two basic kinds, one of which I shall call *principled skepticism* and the other *boggled skepticism*. As will be seen, sometimes principled skeptics share doubts with boggled skeptics, but the distinction is meant to correspond to two distinct rationales. Briefly, the principled skeptic believes, for reasons presently to be adumbrated, that the kind of explanation available through neurological research will always, given the nature of the case, be inappropriate to our understanding of such things as perception, memory, and consciousness. The boggled skeptic, on the other hand, tends to be awestruck by the 10^{12} neurons and the very considerable complexity of the nervous system. He allows that perhaps, in the far, far distant future, brain research might be revealing, but thinks that for the foreseeable future, it is only remotely possible that brain research will yield anything of interest to the psychologist or the philosopher. Both types of skeptism seem to me to be in error and to stand in need of dislodging. The guiding aim of this paper is to provide each with a bit of a jar.

Assuming that the cognitive activity of organisms is brain activity, why should brain-based theories of such activity be thought to be the wrong kind of theory—indeed, how could they fail to be the right sort of theory? The currently bruited answer that engenders spirited spurning of neuroscience proceeds in this vein: at best, neuroscience can provide only a *structural* theory, as opposed to the *functional* theory sought; at most, it will give us the engineering minutiae as opposed to the design configurations we want. It is in the nature of the case that neuroscientific theories will fall short, because cognitive states and processes are properly understood as functional states and processes, and the research program for psychology, precisely in contrast to neuroscience, is to determine the functional organization that accounts for cognitive activity. To be sure, cognitive states are, at least in ourselves, also brain states, but what is interesting about them qua cognitive states is their functional role, not their neuronal instantiation. This can be seen by reflecting on the following considerations: (1) there will be a one-many mapping of functional states onto physical realizations of those states, because functional states can be realized in a variety of material bases. Furthermore, within one species, or even within one organism across time, different batteries of neurons will be recruited for a type-identical task, say, computing the sum of 67 and 59. (2) Neuroscience will yield generalizations orthogonal to cognitive explananda. Simply put, it will give us the wrong generalizations.[4] The type-identical state consisting of the belief that hens molt may be produced in one person by his having read that hens molt, in another by his having been told, in another by his observation of molting hens, in another by deduction from other biological beliefs, and so on. As a result of this belief, together with the belief that molting hens do not lay and the common desire to maximize egg-laying in the hen-house, each of the persons may perform the same action, to wit, wringing the necks of the molting hens. But the neuronal, causal stories will be totally different, given the distinct acquisition of beliefs, and, consequently, the crucial cognitive similarity between the persons will be hopelessly unavailable from the neuronal story. Structurally based as it is, neuroscience will not have the resources to be sensitive to the important cognitive similarities. In contrast, cognitive psychology, faithful to the explananda by deliberate design, will provide the right sort of generalization, if anything can.

Having thus delineated the nature of the research problem, the functionalist naturally enough sees the appropriate research strategy as exclusively top down.[5] Briefly, it will consist in explaining

[4] Pylyshyn advances essentially this objection in *Cognition and Computation*.

[5] Thus Barbara von Eckardt Klein says:

> . . . in general, research must proceed from psychology to neurology. The reason is simple. We will not be in a position to discover how language-responsible cognitive structure is realized neurologically until we know what is so realized. In other words, evidence of the neurological realization of language-responsible cognitive structure cannot be properly evaluated except in the context of linguistic and psycholinguistic models of language-responsible cognitive structure (5).

in "What Is the Biology of Language?," in E. Walker, ed., *Explorations in the Biology of Language* (Montgomery, Vt.: Bradford, 1978). See also Pylyshyn, "Computation and Cognition," *The Behavioral and Brain Sciences*, III, 1, forthcoming.

such top-level, abstract, and familiar processes as intelligent choice and hypothesis testing as the outcome of less abstract and less complex processes on the part of a suitably organized collection of lower-level functional components. This is captured in Zenon Pylyshyn's claim that the appropriate type of explanation is explanation by functional decomposition.[6] Common-sense psychology, we are told, constitutes the correct first stage of cognitive theory, and essentially what is needed is an extension and development of common-sense psychological theory. From this point of view, the intelligent organism is a sentential automaton,[7] whose behavior is the outcome of a sequence of mental states (beliefs that *p*, desires that *p*, etc.), and the processing will be described in terms of the *semantic and syntactic* relations among the content-specifying sentences.[8]

The adequacy of the aforementioned complaints against the ambitions of neuroscience rests substantially on the adequacy of the central assumption, namely, that common-sense psychology is an essentially correct, if fragmentary, theory of how the mind-brain works. That common-sense psychology is indeed a theory, and as such might be seriously misdirected or significantly misconceived, is occasionally recognized as an empirical possibility, but it is nevertheless considered handily dismissible as a *mere* possibility.[9] Such resolute confidence notwithstanding, there are weaknesses in common-sense psychology, weaknesses which threaten to sully attempts to develop it into a scientific psychology. In what follows, I shall first provide a summary discussion of some of these weaknesses. I shall then turn to a discussion of a selection of examples of neuroscientific research which indicate the promise of a quaquaversal, or multi-directional, approach.

There is, to begin with, what I shall call the *infralinguistic catastrophe*. Nature abounds in infraverbal intelligent activity, the most spectacular of which is perhaps to be found in the aspiring infant hominid, lisping his way to linguistic fluency and worldly wisdom. Intelligent behavior is no more unique to humans than is a nervous system, and undoubtedly many nonverbal nonhumans display im-

[6] Pylyshyn, "On the Explanatory Adequacy of Cognitive Process Models," mimeo; parenthetical page references to Pylyshyn are to this paper.

[7] See my "Fodor on Language Learning," *Synthese*, XXXVIII, 1 (May 1978): 149–159.

[8] For a fuller description, see my "Language, Thought, and Information Processing," forthcoming in *Noûs*; and Fodor, "Computation and Reduction," in C. Wade Savage, ed., *Minnesota Studies in the Philosophy of Science*, vol. IX (Minneapolis: Univ. of Minnesota Press, 1978): 229–260.

[9] For example, Fodor, *The Language of Thought* (New York: Crowell, 1975), ch. 2.

pressive intelligence. The belled elephant who sneaked into the banana plantation after stuffing her bell with mud exhibits a capacity which deserves to be called cognitive. The question is, how best to model the infraverbal information processing underlying the behavior? As Jerry Fodor (*ibid.*) incisively points out, the only model available within the top-down approach is explicitly and unrepentantly linguistic. Thus, in order to explain infraverbal activity, a nonverbal innately understood language (Mentalese) is posited, and the infraverbal representations are accounted for as Mentalese representations. The infraverbal is not, on this view, *infralinguistic*; where there is cognition, there also is linguistic representation, no matter how far, phylogenetically or ontogenetically, the creature is from overt language. As to language learning, it is explained within the confines of this approach as nothing other than translation between Mentalese and the target language. As if all this were not bad enough, hard on the heels comes the consequence that there is no such thing as concept learning (*ibid.*). On this view, hard-wired in the organism are the conceptual "atoms," and the old war horse, logical atomism, is wheeled out to explain how we acquire—but do not learn—complex concepts, including, presumably, such concepts as meson, charmed quark, and force field. The analogy with the ultraviolet catastrophe in pre-quantum physics is regrettably apt.[10] In both cases, the consequences of the theory are impossible to digest. The recalcitrant result, alas, is no trifling oddity that can be trussed up and eased into the existing theory by minor adjustment. Rather, it presages the advent of a quite different theoretical paradigm.

The linguistic model for explaining infraverbal representation does have the advantage of being available, but the fact that it is inconsistent with genuine concept learning renders mere availability a diminishing virtue. True enough, language is the most powerful system for representing that we know anything much about. The system we do not yet understand is the system that is in fact used by the brain. Although it is natural enough initially to hypothesize that the system we do *not* know is similar in the relevant respects to the system we *do* know, the infralinguistic catastrophe reveals the perils attending that hypothesis.

Common-sense psychology is beleaguered with yet further woes. The companion concepts of belief and desire are to function as the coping stones of the envisaged scientific psychology, but it now ap-

[10] For an account of the ultraviolet catastrophe, see J. Andrade e Silva and G. Lochak, *Quanta* (New York: World University Library, 1969), ch. 2.

pears doubtful that they have the theoretical integrity to so function. There are difficulties with deductive closure, which, as Daniel Dennett [11] argues, are not readily solvable by positing a deducer mechanism to handle beliefs *in potentia*; there are difficulties in determining which beliefs and desires are causally in the picture, difficulties attending individuation of beliefs and specification of content (39 ff) and difficulties with determinateness when the content is specified sententially (39 ff). Stephen Stich [12] has produced several remarkable results which indicate a disastrous mismatch between scientific psychology's need to postulate causally efficacious cognitive states and common-sense psychology's ability to specify content for beliefs and desires. From a rather different standpoint, Paul Churchland [13] has adduced general considerations which indicate that the linguistic model is decidedly parochial and that how evolution solved the problems of information *processing* is probably quite different from how it solved the problem of information *exchange*. In a similar spirit, C. A. Hooker [14] stresses the communicative function of language, and argues that it is a relatively superficial phenomenon in the information-processing story. To argue these views here is neither appropriate nor necessary, but to cite them is important in dispelling the comfortable illusion that, apart from a superficial blemish or two, common-sense psychology will do nicely. Together with the infralinguistic catastrophe, these difficulties bespeak the need for conceptual innovation, and to suppose that the work of neuroscientists must be irrelevant here is self-defeating. If in search of a scientific psychology we confine ourselves to working top-down, it may be every bit as constraining as trying to determine explanations of chemical behavior from within alchemical theory, or the nature of life from within vitalism.[15]

Before considering the remaining arguments of the principled skeptic, I want to dramatize my point by showing how results from neuroscience can bear not just upon the details of top-down theory, but, more critically, upon the very top-down criteria employed to test the adequacy of top-down theories. In this way I hope to show

[11] Dennett, "Brain Writing and Mind Reading," in *Brainstorms, op. cit.*, pp. 39–50.

[12] "Do Animals Have Beliefs?," *Australasian Journal of Philosophy* LVII, 1 (March 1979): 15–28; and "Autonomous Psychology and the Belief-Desire Thesis," *The Monist*, LXI, 4 (October 1978): 573–591.

[13] *Scientific Realism and the Plasticity of Mind* (New York: Cambridge, 1979).

[14] "The Information-processing Approach and Its Philosophical Ramifications," *Philosophy and Phenomenological Research*, XXXVI, 1 (September 1976): 1–15.

[15] For a critique of dominant top-down strategies in epistemology, see Paul M. Churchland, *op. cit.*

that the orderly transition from the top to deeper and deeper psychological levels may be threatened, and to demonstrate the risk in counting on the security of any scaffolding in the abstract structure that is our current and prevailing cognitive theory. In consequence, I hope to render more reasonable the idea that bottom-up research should be an important ingredient in top-down research.

In "Towards a Cognitive Theory of Consciousness," [16] Dennett succeeds in giving the flavor of functional organization at a level down from the level at which one standardly conceives of oneself when, for example, one thinks of oneself as a unified conscious being. Psychology at the *subpersonal* level, as Dennett puts it (153), explains activities of *persons* as the outcome of complexes of activities carried out by functional parts or faculties which conjointly are the person. A theory at the subpersonal level will posit a functional organization of modules each of which deals with information in certain specified ways: they may operate on information received, exchange information, and some will put out behavior commands. Admittedly, the painting is in bold strokes, but these are the sort of features borne by a cognitive theory in its nascent stages and proceeding top-down. Dennett's aim in the paper is to construct a functional organization such that it would be reasonable to say that anything that instantiates that particular functional organization is conscious. As such, it will be a theory about what it is to be conscious, and *eo ipso* about what consciousness is (TCC 149). It seems to me that Dennett's is a decidedly worthy theory—it is eminently plausible and highly sensible, at least viewed from the vantage point of the abstract heights of top-down research. What I wish to focus on here are two fundamental claims Dennett makes, drawn from our *personal*-level psychological theory, which he deems so safe as to be used to test the adequacy of his subpersonal-level theory. The first criterion is this: "if one can say something about [i.e., report observationally on] some current feature of the perceivable world, one has experienced it" (TCC 158). The second criterion is this: someone experiences something if and only if he is conscious of it (TCC 149).

Both claims seem as safe as houses, and it is fair to say that both are quite highly entrenched in common-sense beliefs about consciousness and conscious beings. From *within* the prevailing psychological paradigm, these propositions might well be graced with the

[16] *Brainstorms, op. cit.*, pp. 149–173; hereafter referred to as "TCC."

status of conceptual truths, but from without it can be seen that at least one is false. Here is why.

It is well known that in higher primates, including man, damage to the primary visual cortex results in blindness. However, recent tests conducted under rather special circumstances show that monkeys whose primary visual cortex had been destroyed nonetheless have a residual capacity for visual discrimination, and a rather impressive residual visual acuity. Since the monkeys could not report on their visual awareness or lack of same, it remained a possibility that they were not in fact blind.[17] As luck would have it, in 1974 L. Weiskrantz *et al.*[18] were able to garner information on "blind-sight" from a human subject. The case involved D.B., who for medical reasons had the major portion of the primary visual cortex on the medial surface of the right hemisphere removed. As a result, he was blind in his left visual field in both eyes (i.e., homonymous hemianopia). According to D.B., he had no visual perception in that field. When tested for visual discrimination in that field, he protested that he could see nothing in the left field, and his examiner therefore asked him to guess his answers. Astonishingly, it was found that D.B. was indeed able to make discriminations with respect to gross features of his environment, and could do so with a very considerable accuracy. So long as the presentations were not too small, and the duration of the presentation not shorter than about .0625 seconds,[19] he could correctly discriminate horizontal bars of light from vertical bars (accuracy of 30/30, 29/30), diagonal bars from vertical bars, X's from O's, red lights from green lights, and in addition he could point with great accuracy to where in the *blind visual field* a light was shining. It was incontestable that his "guessing" was no mere guessing and that he was making use of visual information to form a judgment, despite the fact that he did not have consciously accessible visual information. Weiskrantz does not yet have an explanation for this phenomenon, but a speculation worth pursuing is that, in addition to the well-known geniculostriate visual pathway, there is a phylogenetically older pathway involving the superior colliculus and the visual association cortex, and that perhaps this route carries the visual information that makes "blind-sight" possible.[20]

17 N. K. Humphrey and L. Weiskrantz, "Vision in Monkeys after Removal of Striate Cortex," *Nature*, CCXV (1967): 595–597.

18 "Visual Capacity in the Hemianopic Field following a Restricted Occipital Ablation," *Brain*, XCVII (1974): 709–728.

19 For a complete account of the data, see Weiskrantz, *op. cit.*, p. 716.

20 *Ibid.*, pp. 720–727. For recent work on "blind-sight" in other subjects, see

What is wondrously intriguing about the case of D.B., of course, is that he was providing visual reports on features of his environment, yet he was not doing so on the basis of visual experiences. Indeed, from the account Weiskrantz gives, evidently D.B. was not reporting on the basis of any experiential phenomena at all. Perhaps one might balk at my description here, and suggest that he was reporting on the basis of experiences, but that these experiences were not conscious. However, which description is preferred matters not at all to me. What is important is that on *either* description, one of the "safe" assumptions goes to the wall. Either we *can* make perceptual judgments about some current feature of the perceivable world even though we have not experienced it, or some of our experiences are such that we are not conscious of them. In the top-down approach, it is our intuitions, drawn from common-sense psychology, which play the prominent theoretical role, and the moral of the Weiskrantz story is that our intuitions, even at their best, are frail and vulnerable hypotheses.

The principled skeptic's objection that neuroscience cannot provide explanantia for the desired cognitive explananda is, therefore, enfeebled by the examination of whether the cognitive explananda are theoretically sound enough to stand as the unsullied end of scientific pursuit. If we cannot assume the adequacy of the cognitive explananda, then we cannot fault a strategy that seeks explanations which may fail to converge upon those explananda. However, should there really be a unity underlying cognitive explananda (e.g., the belief that hens molt), then neuroscience may well describe that unity, not of course in terms of the behavior of the peripheral sensory neurons, but in terms of informationally similar structures in the central system. Here it needs to be stressed that neuroscience is not limited to talking about structural minutiae, such as the spiking frequency of individual neurons, but can and does use network concepts (*neurofunctional* concepts), and aspires to the construction of higher-order network concepts. (For more on this, see below, pp. 204–207.) On the other hand, the ostensible unity implied by the cognitive description may in fact have no reality. In that event, what may happen is that neuroscience, initially making use of common-sense concepts, will make discoveries that transmogrify them. Neuroscience is as able to bootstrap as any other science.

---

also M. T. Perenin, "Visual Function within the Hemianopic Field following Early Cerebral Decortication in Man—II. Pattern Discrimination," *Neuropsychologia*, XVI, 6 (1978): 697–707.

Before going further, I want to make a clean breast of the fact that I have knowingly used 'information', 'representation', and their cognates in what must be described as an imprecise fashion. The imprecision, however, is owed not so much to my willful woolliness as to the state of the sciences concerned. The question of how to characterize information processed by the mind-brain and how to characterize the processing of information is a matter for *theory*, and it cannot yet be said that there is an anointed theory. In obvious but precarious contention is the theory of top-down advocates, which sees cognitive representation as inescapably sentential and which sees information processing as primarily a matter of inference within the frameworks of deductive and inductive logic. A rather different and more abstract conception is that of C. E. Shannon and W. Weaver [21] as embodied in the mathematical theory of communication, in which the "information" associated with an arriving signal is a measure of the reduction it effects in the uncertainty among a set of possible "messages." In this case, "information" processing can consist of transformations that bear no relation to those of deductive and inductive logic, e.g., amplification, filtering, averaging, integrating, etc. Further, a close and provocative relative of this notion of 'information' emerges naturally from statistical thermodynamics, wherein "information" is identical with negative entropy.[22] The identity is provocative because statistical thermodynamics provides the physical basis for understanding the evolution of semi-closed systems toward progressively greater order.

For *pre-theoretic* purposes, "information" may simply be conceived of in the following way: the state S of an object O contains the information that P just in case O would not be in S unless P. Neuroscientists for the most part are wililng to content themselves provisionally with some such vague sense of 'information', a sense sufficiently palpable only to underwrite the claim that information processing is what wants investigation and is what we need a theory of. This method is surely sensible, and certainly there is no reason for them to quit their laboratories until 'information' is well defined. On the contrary, it is the aim of neuroscience to *discover* how the mind-brain represents, how information about the world is filtered, stored, and transmitted, and, in so discovering, to pro-

21 *The Mathematical Theory of Communication* (Urbana: Univ. of Illinois Press, 1949).

22 L. Brillouin, *Science and Information Theory* (New York: Academic Press, 1956).

vide a theory characterizing "information" and "information processing." Consider, for suggestive example, the following characterization from A. Pellionisz and R. Llinas: [23]

> . . . the internal language of the brain is vectorial (330).
>
> Thus, the cerebellar cortex is represented as a system which receives vectorial inputs and emits vectorial outputs and where 'information processing' is indeed a transform of a mossy fiber input status vector (row vector of I-elements) into a Purkinje cell activity vector (row vector of I-elements) into a Purkinje cell activity vector (row vector of J-elements). The transform is determined by a network tensor which is, in this particular case, technically a matrix of I rows and J columns (333).

Later in the paper I shall discuss this work, but here my main point is that the desired definition of 'information' awaits theoretical development. A theory of how the mind-brain processes information and a definition of 'information' will emerge *en deux*.

The common-sense conception of our mind-brains and how they work does have an immensely powerful hold on the imagination, and a particularly tenacious conviction which fuels principled skepticism claims that, no matter how sophisticated explanations of human behavior become, no matter how "far down" our explanations go, if they are to succeed in explaining cognitive activities as successfully as our current explanations, they must, unavoidably, advert to representations (interpretations, meanings). Allegedly, no structural description can possibly do what a description in terms of representation can do. Pylyshyn, for example, argues that for nonpsychological things, the downward reduction terminates 'when the functional specification of a component can be mapped on to a physical law'. In contrast, for things which do have psychological processes, Pylyshyn says that "the functional reduction of cognitive functions should come to rest on cognitive laws," not, he insists, on neurophysiological laws (5). His reasons are interesting:

> In the psychological case the functions performed as well as the inputs and outputs of components must be interpreted in terms of what they *represent* in order for the description to address cognitive phenomena . . . the upshot of this feature of symbolic systems (including of course computers) is that although components function according to physical principles, the explanation of *what they are*

23 "Brain Modeling by Tensor Network Theory and Computer Simulation. The Cerebellum: Distributed Processor for Predictive Coordination," *Neuroscience*, IV (1979): 323–348.

> *doing* must make reference to what the symbols represent. Thus explanation by reduction cannot come to rest on physical laws (4).

The reservation Pylyshyn has here can perhaps be demonstrated by the following: Suppose someone says, "That chimney is on fire," and runs and telephones the fire department. According to the principled skeptic, no strictly physical explanation of what the person did will be as good as an intentional explanation, because, for starters, no physical account can do justice to the meaning of the utterance. The meaning of the utterance cannot be specified via a structural account of the equivalence class of utterances that have that meaning. The same message would have been contained whether it had been screamed, whispered, semaphored, given in American Sign Language, or sung to the strains of Pagliacci. Accordingly, explanations drawn from neurophysiology will always be defective in one decisive respect—they will have no truck with representation, sententially conceived.

The way to begin to defeat this objection is to reflect on the over-arching aims and methods of neurophysiology; for it is easy to be dumb-founded by papers with such titles as "Command Neurons in Pleurobranchaea Receive Synaptic Feedback from the Motor Network They Excite," [24] and to wonder how such studies could possibly come to connect with the problems we face in figuring out how we perceive and how we can use language to represent the world. A neurophysiological account of how an organism processes information will include hypotheses about what really *is* the information contained in certain neuronal states at various levels from the periphery, what information is filtered in and what filtered out, and how information is integrated. Given the nature of the case, these hypotheses will be sensitive on the one hand to our understanding of the sort of behavior the states are found to lead to, as well as to our physical and biochemical theory which we shall draw upon in determining what brings about a change in neuronal state. So long as the route from the peripheral input to the behavioral output is pretty direct, the hypotheses concerning what information is borne will be more or less straightforward. Alas, in true life the route is seldom simple and direct, though, when it is, we call the behavior a reflex, and, not needing to ascribe representations to so simple a system, we cheerfully characterize the information contained by appealing to biochemical theory and the overt details of the physical circumstance. The complexity enters when the input

[24] R. Gillette *et al.*, *Science*, CIC, 4330 (1978): 798–801.

to, say, a motoneuron, comes from an extensive array of sources, some of which themselves contain highly integrated information from a variety of sources and some of which contain "internal" information. Evidently when the internal information transmitted to a neuron is very rich, it becomes profoundly difficult to specify precisely what information is contained in its various states. To the extent that it is difficult, we may tend, provisionally, to talk not about the information in the neuronal states, but about the animal's representations. The methodological point is that a demonstrably useful way to track down and converge upon what information is contained is to find out where the neuron set gets its information (e.g., from the cochlea, from the hypothalamus etc.), to find out what kinds of information the neuron is sensitive to, to find out what happens when you give it certain kinds of information, to find out what happens to other parts of the system when it transmits information or when it is prevented from transmitting information, and then to fit this in with whatever else one knows about the nervous system. Nobody thinks this child's play, but these are the things standardly done in the neurosciences to render the complexity tractable. In this fashion then does neuroscience put the squeeze on representation explanations of behavior. The increase in complexity in the nervous system as one moves up the phylogenetic scale is admittedly breath-taking, but there is no reason to suppose that the increased complexity is anything more than increased complexity or to suppose that representing is an emergent property absolutely inexplicable in terms of the underlying physical structure. Representation explanations are explanations we employ either out of ignorance or out of convenience; they are not explanations we employ by dint of the *sui generis* nature of intelligence. The explanation of how the human brain represents will turn out to depend on a theory of what information is contained at various levels and of how, by virtue of physical changes wrought in the system, information is transmitted. Though we surely cannot now say what this theory will look like, progress in the task has been spectacular in the last twenty years, and it does suggest that a solution is not in principle beyond our grasp.

The story of how language represents the world will undoubtedly be a late-comer, for it will be dependent upon our account of how the brain represents generally. There is no reason at this stage of inquiry to believe that the representation typical of linguistic behavior should constitute a unique and inexplicable case of representation. To be sure, social considerations will enter the story, for, as Quine has put it, language is a social art, and so that will add a dimension of complexity. The problem of the "equivalence class" in linguistic behavior (see above, p. 196) is not, I suspect, the bugbear for neuroscience that the principled skeptic imagines. Naturally the neuroscientist bent on getting a theory of how the brain represents would be foolish to try to specify the meaning of an utterance in terms of the physical properties of the utterances themselves or of the external stimuli to which they are related. Rather, the relevant specification will surely be in terms of informationally similar brain structures which figure in the etiology of the behavior. There is certainly no need for the neuroscientist to be red in the face about not *now* being able to provide the description, for it would be astonishing if we had more than common-room guesses at this stage of the endeavor. The equivalence-class problem can be deprived of such steam as it has by noticing that neither can the equivalence class of much *nonlinguistic* stimuli for *simple* animals be specified *in the way held out by the principled skeptic.* That is, if we consider a simple creature with respect to whose processing we do not feel compelled to invoke representation, we still have an equivalence-class problem so far as specifying what evokes flight behavior, feeding behavior and so forth, if we look to the purely intrinsic properties of the relevant conditions. The weakness with the equivalence-class objection is that it rests on a withered and parochial conception of how we might specify the meaning or the telos of behavior.

Though it is certainly a fool's errand to try to summarize in short compass the recent progress in brain research, in a spirit of *faute de mieux* I offer several examples which I think give some indication of how brain research puts the squeeze on representation explanations and contributes to our knowledge of how the brain processes information. My presentation will be ruthlessly synoptic, and it should be borne in mind that a proper appreciation of the significance of these examples does of course require a much more detailed discussion than I can hope to provide here.

Aplysia Californica is a humble sea slug with an approachable nervous system of about ten thousand neurons. This number is, relatively speaking, a mere thimbleful, and this fact, in conjunction with the relative accessibility of its nervous system and a plentiful supply of specimens, makes *Aplysia* a highly prized preparation. Indeed, *Aplysia* may be for neuroscience what *Drosophila* was for genetics. The strategy with *Aplysia* has been to try to get the full story on its entire neuronal organization, and thus to have, in at least one simple case, a complete account of what information its nervous system is sensitive to, how that information is filtered and transmitted, what information it stores and exactly how it stores it, how its behavioral repertoire is produced and exactly what changes are responsible for such behavioral modification as it displays. Naturally it is to be expected that there will be differences between vertebrates and invertebrates, but certainly there are similarities, some of which are well known (e.g., the formation of connections in the visual system). By exploiting the similarities, the job in the complex case can be made much less daunting, and thus the full story in the relatively simple case may well provide the basis from which to make gains in the complex case.

So far the strategy must be counted as having had brilliant success. The simple behaviors, such as gill-withdrawal, inking, locomotion, and egg-laying, are now largely understood in terms of the biophysical properties of identified cells and their invariant connections with each other and with effector organs.[25] This in itself is an extraordinary accomplishment; but perhaps more impressive are the results pertaining to behavioral plasticity, and on this matter I shall now focus.

Habituation is a gradual decrease in the amplitude or in the probability of a response to a repeated presentation of a particular stimulus, and is commonplace in organisms generally. It is an important component in acquiring familiarity with the environment and in learning on which stimuli not to bother expending energy. So, for example, a dog may initially be startled by the sound of gunfire, but in time loses his gun-shyness as he comes to appreciate that the stimulus is not harmful to him. As Eric Kandel explains, "the elimination of responses that fail to serve useful functions is as important as the development of new ones. . . . As a result of habituation to common, innocuous stimuli, an animal can put a large number of stimuli that do not affect its survival beyond its attention. The animal can focus on stimuli that are novel or that become associated with either satisfying or alarming consequences" (539). Ceasing to pay attention to the drone of one's air conditioner, the normal beat of one's heart, and the barking of a neighbor's dog are familiar examples of habituation in humans. In short, habituation serves to cull out responses not keyed to reward and satisfaction, and permits the expression of responses that are more

[25] For a very thorough and readable discussion of recent *Aplysia* research, see Eric R. Kandel, *The Cellular Basis of Behavior* (San Francisco: W. H. Freeman, 1976); parenthetical page references to Kandel are to this book.

useful. *Aplysia* too is blessed with the capacity for habituation, and the attempt to determine the nature of that capacity has been undertaken.

When *Aplysia* is squirted with a gentle spurt of sea water on its siphon skin, the gill muscle contracts and the gill is withdrawn, an obvious protective maneuver. After several repetitions of this mild and innocuous stimulus, the response habituates, and the habituation lasts for some thirty minutes (532 ff). Knowing in detail the circuitry for gill withdrawal, the investigators compiled a list of the nine possible means by which the habituation could take place, and each of the nine was systematically tested. It has now been demonstrated that the habituation was due not to fatigue of the gill muscle, not to sensory adaptation of the mechanoreceptors on the siphon skin, not to decreased sensitivity of the motor neurons, and so on. The changes that do account for habituation of the gill-withdrawal response have been traced directly to changes in the amount of excitatory neurotransmitter passed from the sensory neuron synapse to the motor neuron receptor sites (557–575). Moreover, long-term habituation, *circa* three weeks, has also been traced to synaptic depression,[26] and the neuronal story for dishabituation by a strong stimulus has also been painstakingly uncovered (Kandel, *op. cit.*, 575 ff).

The attempt to map the whole *Aplysia* is only one route among many others into the manifold mysteries of nervous systems, and it needs to be seen in the context of a broad spectrum of activities in the business of contributing answers. These activities range from research on the chemical affairs of cells, to the determination of tracts and pathways in the nervous systems of a wide array of organisms, to the localization of functions in the human brain by measuring differences in glycogen and oxygen uptake during the performance of different tasks.[27] The results from the *Aplysia* studies on habituation are exciting because, in giving us the precise locus of habituation in the nervous system of *Aplysia*, they give us a tremendous lead into a theory of simply learning. Not, indeed, a complete theory of simple learning, for there are many remaining questions. The cellular changes underlying alterations in the volume of neurotransmitter are not yet understood, and the sensory dimension of *Aplysia*'s life is on the whole less well understood

[26] V. F. Castellucci, T. J. Carew, and E. R. Kandel, "Cellular Analysis of Long-term Habituation of the Gill-withdrawal Reflex of *Aplysia Californica*," *Science*, CCII, 4374 (1978): 1306–1308.

[27] This work is beautiful, and I should have discussed it but for lack of space. See Niels A. Lassen, David H. Ingvar, and Eric Skinhøj, "Brain Function and Blood Flow," *Scientific American*, CCXXXIX, 4 (1978): 62–71.

than the behavioral. The question of the generality of the results is pertinent, and some there are who find it positively insulting to suppose that human learning or even human habituation might have anything at all in common with habituation in a vulgar, low-born slug. Insulting or not, it may turn out to be fact, much as it turned out that the exalted and humble alike share in DNA and that what ultimately distinguishes the two are merely different articulations of the same basic structural elements. The serious generality question is this: Is there reason to think that the story on gill-withdrawal habituation in *Aplysia* applies elsewhere? There is evidence, though less conclusive than for *Aplysia*, that the same style of underlying neuronal change is to be found in crayfish (Kandel, 649), and preliminary analysis of cellular changes in vertebrates (cats and frogs), though not yet conclusive, points to the same sort of mechanism (600). It is still too early to say very much here, because the habituation analysis in *Aplysia* is very recent. What should be underscored is how important the knowledge of the simple case is in approaching more complex systems: if there is more to vertebrate habituation, it may now be possible to track down what that more is; if there are differences, having one theory in hand will the better allow us to converge upon those differences; if habituation does figure in more complex learning, knowing the neuronal story for habituation will be invaluable.

In "Why Not the Whole Iguana?," [28] Dennett wistfully hankers after a complete model of a simple organism, but decides that, since what a cognitivist wants anyhow are the abstract, functional principles of information processing, it would be preferable to ignore on-the-hoof information processors, and instead create a whole cognitive beast, guided in one's manufacture by top-down theory:

> . . . one does not want to get bogged down with technical problems in modeling cognitive eccentricities of turtles if the point of the exercise is to uncover *very* general, *very* abstract principles that will apply as well to the cognitive organization of the most sophisticated human beings. So why not then *make up* a whole cognitive creature, a Martian three-wheeled iguana, and an environmental niche for it to cope with? I think such a project *could* teach us a great deal about the deep principles of human cognitive psychology, but if it could not, I am quite sure that most of the current A.I. modeling of familiar human mini-tasks could not either (103/4).

I am less troubled by the suggestion that such construction might teach us something than I am with the corresponding presumption

[28] *The Behavioral and Brain Sciences*, I, 1 (1978): 103–104.

that the neuronal story of a simple organism will *not* teach us anything much. As Paul Churchland has argued,[29] this is analogous to supposing it is preferable to determine the nature of life by forsaking the investigation of the microstructure of actual living things and favoring exclusively the construction of functional models satisfying common-sense constraints about what living things do. This program for biology might conceivably have been useful, and who knows, Dennett's suggested program might possibly be useful. But to deny that the complete neuronal story of an organism will be useful is to deprive oneself to no purpose. True, humans do not exhibit gill-withdrawal, for want to gills to withdraw. However, habituation we certainly do exhibit, of complex behavior as well as of spinal reflexes. Dennett's desire for a complete story in at least one case is widely shared by neuroscientists, though they favor study of actual information processors, and *Aplysia* seems to be the front runner. Undoubtedly, *Aplysia* habituation, *Aplysia* information storage, *Aplysia* information transfer, etc., will inform the investigation of fancier nervous systems.

My second example is drawn from vertebrate research, and is chosen to illustrate the kind of progress that is being made on a problem which, *a priori*, might have seemed too "cognitive" or too high-level to be tackled by neuroscience. In a recent paper, Vernon Mountcastle [30] is concerned to investigate the brain mechanisms for directed attention, and, since the results of lesion studies clearly implicate structures in the parietal lobe, Mountcastle *et al.* set out to see what the neurons of the parietal cortex do when the animal is visually attending to an object of interest. The major behavioral elements in visual attention are saccadic movement, fixation, and visual tracking. The investigators implanted recording electrodes in the parietal cortex of monkeys, and found a rather remarkable neuronal organization. The results of the recordings revealed quite distinct functional units of neurons: (1) light-sensitive neurons, (2) saccade neurons, (3) visual-fixation neurons, and (4) visual-tracking neurons (active during and only during tracking, inhibited during saccades). Moreover, given the data, the orchestration of activity in these neurons appears to be this:

(1) activation of the light-sensitive cells by the appearance of an eccentrically placed target, with a latency of about 80 milliseconds.

[29] Paul M. Churchland, "Is *Thinker* a Natural Kind?," mimeo.

[30] "Brain Mechanisms for Directed Attention," *Journal of the Royal Society of Medicine*, LXXI (1978): 14–28.

(2) conditional discharge of the saccade neurons which follows the onset of activity in the light-sensitive neurons by about 50 msec.

Mountcastle suggests that:

> . . . during that 50 msec. a matching function is executed between the neural signals of the nature of the visual target and those of the internal needs of the organism. I cannot specify this mechanism, but recall the heavy reciprocal connections between the parietal association cortex and the limbic lobe (24).

(3) Provided that the saccade cells discharge, the tracking begins after a delay of 73 msec.

> Once the saccadic movements foveate the visual target, the powerful fixation mechanism comes into play, with its tracking appendage. These commands drive a tenacious visual grasp of objects of interest—the essence of selective or directed visual attention. The activity of fixation/tracking neurons is broken only by the suppression which precedes a subsequent saccadic movement to a new target (24).

Since it is well known that brain-stem structures have a role in eye movement and position, Mountcastle then proceeds to relate the function of the brain-stem oculomotor mechanism to that of the parietal cortex, aided by additional anatomical information about the projections of the parietal lobe neurons to brain-stem oculomotor neurons. His hypotheses concerning command function with respect to visually directed attention are worth quoting;

> The development in primates of a foveal vision with great resolving powers is accompanied by that of an elegant brain-stem apparatus for controlling oculomotor operations. It is my hypothesis that there has developed congruently in the parietal lobe and its distant connections a neural apparatus for the integration and governance of these events. This command source is itself sentient to a continually updated image of the position of the body, head, and eyes relative to the immediate surround and the gravitational field; it is linked to neural signals of the internal drive state of the organism; and from time to time generates commands for the fixation of gaze upon objects of interest—the first step in directed visual attention. I emphasize the general nature of this idea, for there must exist in the brain many sources of command for the direction of gaze (26).

Having plucked for display several pieces of research from the massive mosaic of which they are a part, I can hardly expect immediate and hearty concurrence in my view that they are exemplary of a useful way to proceed. Nevertheless, I do believe the results to be sufficiently striking to give the principled skeptic pause. I suppose reasons might be found for looking askance at these sorts of study, but I do not see how it can be seriously argued that such study is not likely to be useful in revealing the mysteries of mind-brain operation.

With these remarks I shall put aside for a moment my case against the principled skeptic, and I turn my efforts to the task of nudging the boggled skeptic who, recall, doubts that much of significance will be revealed by bottom-up research in the foreseeable future. Something should of course be conceded the boggled skeptic, for indeed the number, size, interconnectedness, and plasticity of neurons in the human nervous system means that neurological research *is* very difficult. Additionally, for many questions concerning the nature and function of normal human living brains, we cannot, for ethical reasons, study normal human living brains. What must be conceded to the skeptic is, however, a good deal shy of blank despair. Progress in the neurosciences, especially since the fifties, has been nothing short of awesome, partly because technological progress has made possible certain kinds of probing hitherto undreamt of. The electron microscope, micropipettes, microelectrodes, freeze-fracturing techniques, radioimmunoassay, the isolation of such substances as acetylcholine and dopamine (to give a handful of examples), have been of the utmost importance to the nascent science. I do not suppose any sober neuroscientist wants to claim that the teeming mysteries of the brain are about to disclose themselves or that the full story of how the human brain works is close at hand. The progress to date in neuroscience does not warrant that kind of unbridled enthusiasm. Yet, on the other hand, the progress demonstrably is sufficiently impressive to make outright skepticism equally inappropriate. A particularly striking example of the fruitfulness of bottom-up research concerns work done on the cortex of the cerebellum, and I shall provide a short discussion of this work in hopes of winning the skeptic over to a position of cautious optimism. Because my discussion of the structure of the cerebellum is all too brief, I refer the reader to the brilliant and eminently readable paper, "The Cortex of the Cerebellum" by Rodolfo Llinas.[31]

The cerebellum is the deeply wrinkled lump tucked under the end of the cerebrum, which consists of immense numbers of neurons and whose size is progressively larger the higher the brain on

[31] In *Scientific American*, CCXXXII, 1 (1975): 56–71.

the phylogenetic scale. A good deal is known about the sort of behavioral deficits caused by cerebellar lesions and about the sort of information that goes into this brain structure, which permits the hypothesis that the cerebellum is a central control point for the organization and orchestration of movement. It is to the neural structure of the cerebellar cortex that we must look in order to determine more exactly what the cerebellum does and *how* it does what it does.

The cerebellar cortex is a spendidly organized part of the brain. To begin with, there are just seven types of neurons, and their respective positions, interconnections, and electrical properties are known in great detail. For example, the Purkinje cell is the sole output cell; its output is inhibitory; it is always oriented in a particular manner; it is associated with precisely one climbing fiber; a fine row of dendrites of a set of Purkinje cells receives input from parallel fibers, etc. The neurons of the cerebellar cortex are arranged in a highly orderly and regular fashion which has been well documented, and, accordingly, *it has been possible to hypothesize a wiring diagram describing in a general way the function of the cerebellar cortex, based on the detailed information concerning the structure of its components.*[32]

That neuroscience has come so far in understanding this part of the brain is obviously enormously exciting, and recently Llinas *et al.* have gone much further in an attempt to characterize the details of the processing in the cerebellar cortex and to determine second-order concepts that will describe network properties.[32] They have "grown" a frog cerebellar cortex in a computer, based on functional and morphological facts about cortical neurons. The computer model consists of 8285 Purkinje cells, 1.68 million granule cells and 16,820 mossy fibers. The simulation is a remarkable tool because it makes the information concerning precisely what happens to input in the cerebellar cortex infinitely more accessible, and of course it renders manageable the "unmanageable numbers." By using the computer model, Llinas has made discoveries concerning network properties which had not been expected and which he subsequently confirmed in tests on actual cerebellar cortices. I emphasize these accomplishments because they are highly interesting not only for what they tell us about the cerebellum, but also for the challenge they represent to the boggled skeptic. Llinas's

[32] A. Pellionisz and R. Llinas, "A Computer Model of Cerebellar Purkinje Cells," *Neuroscience*, II, 1 (1977): 37–48; and A. Pellionisz, R. Llinas, and D. H. Perkel, "A Computer Model of the Cerebellar Cortex of the Frog," *ibid.*: 19–36.

work is an example of the sort of means by which function can be winnowed out of facts concerning appallingly complex structure.

These remarks should be addressed to the principled skeptic as well, and a more recent paper by Pellionisz and Llinas [33] illustrates the manner in which a knowledge of the micro-architecture of the neuropil can suggest general, mathematical models to characterize the functions being discharged. For example, citing the connection in parallel of small groups of Purkinje cells in the cerebellar cortex, cells whose output (firing frequencies) converge upon a nuclear neuron, Pellionisz and Llinas give a most convincing reconstruction of the activity of such a battery in terms of the approximation of some specific function with a Taylor series. Given their common input, the output firing frequencies of the respective Purkinje cells correspond to the values of the first several elements in a Taylor series, and their summation at the nuclear neuron provides the relevant functional approximation.

Turning to more extensive neural networks, and inspired by the patently vectorial nature (specific firing frequencies along specific weighted paths) of their manifold input and output, Pellionisz and Llinas propose tensor analysis as the relevant mode of representation. The brain, they affirm, is a tensorial entity, and as such is subject to universal tensorial laws. Aside from its explanatory power with regard to coordination of behavior, the authors point out that on this hypothesis:

> . . . the genetic code would provide "only" the ontogenetic guidelines for building a tensor in a general sense, leaving the particular selection of the frame of reference and the establishment of the corresponding numerical values of connections to the individual epigenetic development, i.e. determined by "local" thermodynamic processes. As a result, by encoding reference-invariant tensors, not particular matrices, genetic specification is relieved from the awesome task of determining each and every neural connection in the brain (330).

It is not possible to evaluate these claims properly here, though it can be conjectured that the work may constitute a major breakthrough in the attempt to describe neuro-functional concepts.[34] The point I wish to make is that it should be plain, even to the most principled of skeptics, that the search for adequate *func-*

[33] "Brain Modeling by Tensor Network Theory and Computer Simulation. The Cerebellum: Distributed Processor for Predictive Coordination," *Neuroscience*, IV (1979): 323–348.

[34] See the comments by Theodore Melnechuck, "Network Notes," *Trends in Neuroscience*, II, 4 (April 1979): 6–8.

tional characterizations can proceed very profitably indeed from a bottom-up direction. And surely it is better to be guided here by the empirical *facts* about brain activity, as discovered by neuroscience, than to be guided by the ancient and ramshackle *theory* of cognitive activity embedded in the intentional idiom of common sense.

I shall conclude with two observations. First, it will be obvious from the diversity of techniques, problems, and subject matter in the examples that "bottom-up" does not distinguish a unique approach and that what constitutes the "bottom" is loosely delimited. For purists, the real bottom will of course belong not to neuroscience but to physics. It is only by contrast with the top-down character of cognitive psychology and conceptual analysis that the neuroscientific approaches are called "bottom-up," and these approaches are more accurately called "quaquaversal," to reflect the variety of routes into the problem of how nervous systems do what they do.

Secondly, what finally motivates my thesis is the insight, owed to Quine, Sellars, and others, that our conception of ourselves is, after all, *concept-mediated*, and, insofar, it is questionable whether those concepts permit us an adequate understanding of the facts. Once the question is asked, and once it is seen that common-sense psychology must be evaluated as a theory, not taken for granted because it is "dead obvious," the certitude of the exclusively top-down approach pales. I think this is on the whole a most salutory thing, for common sense may have no better a line on mind-brain function than it did on the nature of motion or heat. What we are and how our mind-brains work remains to be discovered, and those discoveries may depend on our willingness to see ourselves in the light of evolutionary dynamics and nervous-system kinematics; in short, to see ourselves as part of the natural order. In the words of Aldous Huxley,[35] psychology has no more right to be anthropomorphic than any other science.

PATRICIA SMITH CHURCHLAND

University of Manitoba

[35] *Eyeless in Gaza* (New York: Penguin, 1936), p. 19.