ISSN 0020–174X

INQUIRY

An Interdisciplinary Journal of Philosophy

Editor: Alastair Hannay

Vol. 29, No. 2, June 1986

Symposium on Patricia Smith Churchland's *Neurophilosophy*

Norwegian University Press

INQUIRY
An Interdisciplinary Journal of Philosophy

Founded by ARNE NAESS

Practice and theory in all areas depend on basic assumptions in the background of discussion. *Inquiry* gives precedence to articles and discussions which bring these assumptions into focus and which place questions of policy or theory in the wider context of traditional philosophical concerns. Preference is also given to papers composed in a style that renders their content accessible to non-specialists.
Papers will as a rule be published within one year of receipt.
Reviews are confined to extended discussions of books selected for their special interest. Books sent for review are listed in each issue.
Instructions for authors are given on the inside back-cover.

Typescripts for consideration and books for review should be addressed to:
The Editor, *Inquiry*
Institute of Philosophy
P.O. Box 1024 Blindern
0315 Oslo 3, Norway

Correspondence concerning subscriptions and advertising should be addressed (marked 'Journals – *Inquiry*') to:

Norwegian University Press
(Universitetsforlaget AS)
P.O. Box 2959 Tøyen
0608 Oslo 6, Norway

or

Norwegian University Press
(Universitetsforlaget AS)
c/o Publications Expediting Inc.
200 Meacham Ave.
Elmont, NY 11003, U.S.A.

Annual Subscription:
Institutions N.kr. 320,–, $53.00
Individuals N.kr. 165,–, $27.00

Subscribers notifying the Publishers of a change of address are asked to include their former address.

Published quarterly by Norwegian University Press (Universitetsforlaget AS), Oslo, Norway. Annual subscription US$53.00/27.00. Second Class Postage rate is paid at Jamaica, NY 11431. U.S. Postmasters: Send address changes to *Inquiry,* Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003. Air freight mailing in the U.S. by Publications Expediting Inc., 200 Meacham Ave., Elmont, NY 11003.

PAGE BROS Computer typeset and printed
offset litho by Page Bros (Norwich) Ltd. England

Contents

Symposium on Patricia Smith Churchland's *Neurophilosophy*

Inquiry, 29, 187–202

Symposium: Patricia Smith Churchland's *Neurophilosophy**

Explaining Behavior: Bringing the Brain Back in

Christine A. Skarda
University of California, Berkeley

What is needed today is a biologically grounded explanation of behavior, one that moves beyond the so-called mind–body problem. Yet no solution will be found by philosophers who refuse to learn about how brains and bodies work, or by neuroscientists pursuing experimental research based on outmoded or blatantly anti-biological theories. Churchland's book proposes a solution: to come by a unified theory of the mind–brain philosophers have to work together with neuroscientists. Yet Churchland's vision of a unified theory is based on an assumption that, while widely held, may not adequately reflect brain functioning in the production of behavior, namely, the assumption that brain processes *represent*. The present paper proposes an alternative view, suggesting that patterns of neural activity do not 'represent' anything, that brains do not 'read' or 'transform' representations, and that brains do not require representations to produce goal-directed behavior. Representations are replaced by self-organizing neural processes that achieve a certain end-state of interaction between the organism and its environment in a flexible and adaptive manner. Some of the implications of this view for neuroscientific research and the philosophy of mind are outlined.

I. Introduction

Today the mind and the body are as far apart as ever, and the prospects for their reconciliation given current conceptions of each are poor. On the one hand, philosophers of mind whose project it is to understand what an explanation is relevant at all to a philosophical theory. An extreme form relation, do not have any real idea of how to accommodate neuroscience into a genuine philosophical theory of the explanation of behavior. Indeed, some philosophers of mind doubt whether the neurophysiological level of explanation is relevant at all to a philosophical theory. An extreme form of this is expressed by functionalism, according to which the neurophysiological level of explanation is irrelevant to the explanation of behavior.[1] On the other hand, neurophysiologists do not know what to make of the philosophical explanation of behavior, and have reacted by implicitly accepting a mind–brain dichotomy while talking in materialistic terms. In a move similar, but opposite to functionalist philosophers they have banished concepts such as mind and purpose from their explanations. They claim that all they are, and need be, interested in is explaining how

---

* Patricia Smith Churchland, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Bradford Books. Cambridge, Mass./London: MIT Press, 1986, xiv + 546 pp., $27.50.

the body and its brain works. How their work relates to mind is thought to be none of their business; such issues can safely be left to philosophers and psychologists. The upshot of this is that neither philosophers nor neuroscientists can do their job properly: philosophers argue about 'solutions' to the mind–brain problem knowing next to nothing of how brains actually work, while neuroscientists seek out the biological bases of hypothetical structures and processes posed by badly reasoned and sometimes blatantly anti-biological theories. Yet, there is really no point in looking for the biological basis of a psychological or behavioral process that neither exists nor makes conceptual sense, and a 'solution' to the mind–brain problem that seriously considers only the mind can, at best, claim only to be a solution to the 'mind problem', if there is one.

Given this 'shotgun divorce', Patricia Smith Churchland's book, *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, is welcome, as is its announced strategy of introducing philosophy and neuroscience to one another (p. 6). In fact, it is long overdue; the climate is ripe for cross-disciplinary attacks on the explanation of behavior. Ours is the age of cognitive science, of interdisciplinary efforts to explain behavior and cognition. In the present climate philosophers are less inclined than before to view their discipline as one concerned only with a priori issues. In part, this can be traced to the recent success of functionalism with its close ties to cognitive psychology. Philosophers of mind are now more receptive to considering empirical data gathered by psychologists when discussing such important issues as mental representation and mental images.[2] On the other hand, because of tremendous progress in the neurosciences, due in part to vastly improved technologies, neuroscientists have become optimistic that their discoveries may unravel some of the larger problems, including the biological bases of cognitive life. One prominent researcher, Eric Kandel, says, '[T]he excitement in neural science today resides in the conviction that the tools are at last in hand to explore the organ of the mind, and with this excitement comes the optimism that the biological basis of mental function will prove to be fully understandable'.[3]

Researchers in both fields have taken to looking over the fence that separates them in search of answers, as well as for the questions that will lead to fruitful new lines of inquiry. Of course it would be easier for both sides if the fence were simply removed, but there remains a great deal of resistance in both camps. The reasons for this are varied; much of it is due to ignorance and widespread misunderstanding of what the other side is doing, and can or cannot do. It is here that Churchland's book has one of its most important contributions to make. By introducing philosophers to neuroscience, by discussing and demolishing many of the philosopher's stock, and I think stale, objections to the explanatory relevance of neuroscientific data, and by presenting neuroscientists with a clear introduction to recent philosophy of mind and its historical roots in earlier theories, Churchland's book helps to remove the barriers of misunderstanding. Like Churchland, I am painfully aware of how badly such an effort is needed at present, and I cannot fail to admire her *tour de force* in this regard. What is wanted is a theory of the mind-brain that builds on a biologically grounded philosophy of mind and the best neuroscientific data currently available.

But Churchland's book is not only important as an interdisciplinary introduction to neuroscience and philosophy of mind. For those who know the basics, it is much more: it is a challenge to produce what Churchland refers to as a 'unified theory of the mind-brain' (p. 3), to come up with a genuine explanation of behavior. And from this perspective what I find most interesting about Churchland's book is the sort of unified theory she envisions.

Churchland claims that what we ultimately need is 'a theory of how the mind-brain represents whatever it represents, and of the nature of the computational processes underlying behavior' (p. 5). That the mind-brain *represents* and that by doing so it creates behavior, is never questioned; here is one of the central assumptions of Churchland's philosophy of mind. One of the principal aims of her book is to show how neuroscience in its present form is a constraint upon, and can be unified with, such a theory (p. 5). Throughout the book attention is focused on the question of how, rather than whether, brains represent; the assumption motivates Churchland's criticism of folk psychology and of functionalism, as well as her model of neural representation in chapter 10.

Churchland's explanatory model is a 'cognitivist' theory. I disagree with cognitivism on biological as well as philosophical grounds. Yet Churchland will surely welcome the proposal of an alternative theory of the mind-brain. Like folk psychology, her cognitivist assumptions constitute a 'theoretical framework, and hence a framework whose adequacy can be questioned and assessed' (p. 311). The 'co-evolution of theories' that, Churchland suggests, will eventually lead to a unified theory, gives no theory, not even hers, epistemological immunity. Churchland's clear and forceful formulation of what I shall call a cognitivistic neuroscientific theory of the mind-brain challenges her readers to consider whether this or yet another form of unified theory can better advance our attempts to explain behavior.

In the present paper I present an alternative explanatory model. I argue that the patterns of neural activity responsible for behavior do not 'represent' anything, that brains do not 'read' them, and that 'neural representations' need not play a role in the production of behavior in animals. The notion of representation does not accurately describe the neural mechanisms that actually play a role in producing behavior. These claims are supported by a body of neurophysiological data that Churchland does not discuss. These data suggest a different direction for the unified theory that both Churchland and I seek, one that requires a reorientation of both philosophy and much of contemporary neuroscience.

II. The Tensor Network Theory as Explanation

Churchland's suggestion for a unified theory of the mind-brain is the 'tensor network theory' developed by Pellionisz and Llinas,[4] here exemplified by Roger the crab of chapter 10. According to this theory what the brain does is solve geometrical problems (p. 419) with arrays of neurons all working in parallel doing matrix multiplication (p. 417). This may come as a surprise to many of us who have trouble with any sort of mathematical problem, but the counterintuitive nature of this theory pleases Churchland: the geometrical (nonsentential) nature of its representations (p. 411) and its radical departure from folk psychology and common sense (p. 419) are among its virtues, she claims.

Roger is an extremely simple creature whose entire behavioral repertoire consists in reaching out and touching any object which he visually locates, yet Churchland claims that the principles on which he works 'could be applicable quite generally' (p. 443). Churchland's objective is to show us 'a new and powerful paradigm for understanding computational processes executed by the mind-brain and for understanding how the mind-brain represents at a variety of organizational levels' (p. 457); assuming all along, of course, that what is wanted is a theory of brain functioning that uses representations. The essence of the theory is this: Roger's sensorimotor behavior results from arrays of neurons performing vector-to-vector transformations on input (representations) from one coordinate system (e.g. visual) to another (e.g. motor). The two systems are separate, but systematically related. Neural representations on this model aren't sentences or images, instead they are vectors in phase space (the collective output of a neural assembly that represents a certain state of affairs in the world or in Roger's own body); the computations required by this cognitive theory are coordinate transformations between phase spaces (p. 426).

The details of the theory can be gained from the text and need not be repeated. What is important about the model are certain underlying assumptions about the nature of brain function in the production of behavior, and which give the model shape. Pellionisz and Llinas derived their model from observing *cerebellar* functioning and anatomy.[5] The cerebellum is chiefly responsible for adjusting or regulating posture and movements like those involved in the vestibulo-ocular reflex (VOR) discussed by Churchland in chapter 10. Experimental evidence indicates that the cerebellum does not create movements, and complete cerebellar ablation has no effect whatsoever on some movements, noticeably that of the jaws,[6] movements that Roger does not produce but that Churchland suggests could be produced using her model. Cerebellar dysfunction does not eliminate movements, but rather degrades or impairs them in various ways. The cerebellum is part of the cerebellar circuit, a system of interconnected parts of the brain that is said to 'transform and combine messages of intent and results into an optimal set of instructions for motor execution'.[7] The tensor network theory is designed to model the process of adjustment, transformation, and combination that gives rise to the 'fine tuning' of motor behaviors produced by the cerebellum. Churchland presumes that all behavior can be explained on the cerebellar model as 'a function of the input and the internal processing of the input' (p. 408), that is, as the result of internal *transformations* of input patterns to receptors. At best, this view is only partially correct.

Historically, behaviors that could be explained in terms of transformations performed on input have been referred to as 'reflex', to distinguish them from 'voluntary' behaviors. According to this classic distinction, reflex behaviors are stimulus-dependent in the sense that there is no output except that which is triggered by input; what takes place within the organism is a transformation of the input pattern. Cerebellar neurons are marvelously suited as the physiological basis of such behavior. The cerebellum contains large populations of neurons that act primarily via feed-forward synaptic connections on one another, so that an incoming stimulus causes output by effecting each link in the neuronal chain in turn like a string of dominoes.[8] The model developed by Pellionisz and Llinas is designed to explain behavior produced in this way. As a result, their model is limited in its application. This limitation is indicated by experimental evidence that shows that cerebellar functioning only adjusts neural patterns of output originating elsewhere in the brain; 'intent', a term used to refer to the nonstimulus-dependent, creative feature of many behaviors, is not caused by cerebellar functioning.[9]

Attempts have been made to explain all behavior on the basis of the reflex model, but they encountered grave obstacles. So-called voluntary behavior resisted all attempts, e.g. that of the behaviorists, to build it up out of simpler, involuntary mechanisms. The process threatened to leave an entire class of behavior without a meaningful physiological basis: Descartes, for example, had to introduce the nonphysical, mental element of 'will' in the causal sequence leading to voluntary behavior in order to distinguish it from simple reflex behavior, a tactic that left philosophers and scientists to the present day anguishing over the so-called mind-brain problem. But reflex theory not only encountered difficulties when applied to voluntary behavior, it also proved unsatisfactory for explaining reflexes. The more physiologists learned about reflex behavior after the pioneering work of Sherrington[10] and Pavlov,[11] the more apparent it became that the machine neurology[12] of the classical reflex-arc was inadequate for explaining behavior. Critics of the reflex model argued that it was defective[13] or an oversimplification.[14] Sherrington, who developed the concept of the reflex arc as the basic unit of physiological functioning, cautioned that the simple reflex is a 'fiction';[15] and later in his research career Sherrington uncovered what he termed the 'central excitatory (inhibitory) state' (c.e.s.), a neural process taking place among a pool of neurons that *actively* coordinates and integrates reflexes that, in themselves, are produced passively, a neural process that acts on present stimulus input but is not caused by it.[16] Alternative explanatory frameworks were proposed early on, one of the most interesting by a philosopher, Maurice Merleau-Ponty. Merleau-Ponty claimed[17] that the brain's main function is to 'reorganize' behavior rather than to react to input, and that even reflex behavior depends upon the 'style of behavior' that the organism engages in and upon its goals and present circumstances. He argued further that if a merely transformational and passive process can be isolated in the artificial setting of the laboratory, this in no way guarantees that such a process operates in the intact, freely behaving animal to produce behavior. He suggested, instead, that in the nervous system the physiological processes underlying the reflex are permeated and entirely changed by the cooperative functioning of the 'higher centers' in the cerebral cortex, which use principles and constraints not expressed by classical reflex theories.

The structure of neural dynamics in other parts of the brain, notably in the cerebrum, bears out Merleau-Ponty's claim that something more than a transformation of incoming stimuli goes on in the brain. At the physiological level the cerebrum differs significantly from the cerebellum. Cerebral neurons are densely interconnected to form innumerable feedback loops. These neurons interact, rather than react, forming a hierarchy of distributed neural masses. Cerebral neural masses are not merely error-correcting as are the neurons in the tensor network model, they are creative; they generate spatial patterns of activity *de novo* within the brain, patterns that form the neural bases for adaptive behaviors or actions (see III below). I do not mean to suggest by this that we should oppose a 'cerebral theory' to a 'cerebellar' one; rather I want to point out that the story of behavior is more complex than Churchland's model indicates, and that behavior depends upon the *creative* neural processes taking place in the brain, processes unlike those taking place in the cerebellum alone or in Churchland's model.

The upshot of all this is that while Churchland's tensor network theory might be an adequate model for reflex behavior of machines and of animals subjected to the highly artificial conditions of the laboratory, it is doubtful whether, as Churchland claims, 'the sensorimotor problems faced by more realistic creatures can be understood as reducing at bottom to the same general type of problem that Roger faces – namely, the problem of making coordinate transformations between different phase spaces' (p. 426). Cerebellar functioning is directed in the intact nervous system by the creative neural processes taking place in the cerebral cortex, but Churchland's model only transforms, it does not create, neural patterns. In an imaginary world in which robots simply react to stimuli in the stereotypic way that Roger the crab does, this model may work, but in the real world where organisms search for food, pursue it, and make choices this model is bound to fail. Action, not simply reaction, is what behavior is all about. Roger only reacts; moreover, Roger reacts in a 'generic' way; his touching is generic touching, as opposed to genuine biological touching which is always 'touching in order to', e.g. push away, caress, examine, or warn, behaviors with quite different forms.

Having said this, however, I have not addressed Churchland's central assumption: that the mind-brain represents. The next sections present neurophysiological data that contradict this claim, making neural representations inconsistent with as well as unnecessary for behavior that is adaptive and goal-directed.

III. Observations

Cognitive theories in neuroscience postulate that neural representations are the basic units used by brains in the computational tasks responsible for behavior. According to some theorists, neural representations can be identified with the activity patterns of individual neurons,[18] while more recently the trend has been to identify them with the electrical and chemical activity of large populations of neurons.[19] The question now is whether the evidence supports such a hypothesis. This question is important to working scientists because as they observe neural activity patterns in the brain they require structured hypotheses about what they are looking at and on what to look for.[20]

A recent series of experiments has disconfirmed the cognitivist hypothesis.[21] These experiments studied the processing of incoming olfactory data in rabbits with electroencephalogram recordings (EEGs) from arrays of 64 electrodes chronically implanted over the lateral surface of the olfactory bulb. Early experiments disconfirmed the reflex-based models proposed by earlier theories and showed no odor-specific dependence of EEG spatial pattern on novel odorants presented to naïve animals. Nonetheless, each animal was found to have its own distinctive EEG spatial pattern that was as characteristic as a hand-written signature.[22] Under classical conditioning, however, the data told a different story. The EEGs of rabbits taught to respond differentially to two odors, showed three spatial patterns of amplitude, one for each conditioned stimulus (CS) and one for the background odor complex. In subsequent recording sessions a new stable EEG pattern persisted whether or not the odor-conditioned stimulus was present, although the rabbits continued to respond preferentially with each odor presentation. When conditioned to respond to new odors, new EEG patterns emerged and stabilized. When a previously learned odor was reintroduced a new spatial pattern emerged, not that recorded in earlier training sessions. The data indicated that a change occurs with learning, but there was no evidence that the EEG pattern represented the presence of a particular odor.

These data ruled out passively generated neural representations completely constrained by receptor input, but left open the possibility that representations in some other form played a role in odor recognition. Investigating this hypothesis, Freeman and his colleagues asked whether the EEG patterns they recorded manifested states of 'expectancy' rather than responses to present stimuli.[23] They hypothesized that the activity of neural assemblies might constitute a 'search image',[24] and that pattern changes reflect updating of this search image under conditioning.[25] This hypothesis says that changes in bulbar EEG burst-amplitude spatial patterns are related to changes in expectancy: when the animal expects a particular odor to which it has been trained to respond preferentially, a neural representation in the form of an activity pattern should exist in the bulb. But this hypothesis, too, was disproved by further experimentation. Freeman[26] found that differences corresponding to a specific learned odor are not detectable prior to odor presentation, but only when the odor is present and after the animal has learned to respond to odors during the training process. Changes in EEG spatial patterns, thus, depend on learning and a state of expectation, but the correlations between odor and EEG spatial activity patterns required to identify the latter as a representation of the odor do not exist until after the expected odor is actually presented to the animal.

This implies that while olfaction is essentially a goal-directed process, rather than simply a reactive one (i.e. capable of being modeled using a reflex-based theory), it does not require a prior active neural representation of its goal. The findings show that the cognitivist explanatory model is inadequate,[27] but they also rule out reflex-based models. The question faced by Freeman and his associates was what sort of explanatory model might accommodate these new findings? The interpretation they developed requires some understanding of bulbar neural mechanisms.

IV. Self-organizing Neural Mechanisms

The parallels that exist between the olfactory bulb and other cortical areas with respect to dynamics of the neural masses that comprise them and the type of neural activity they generate are significant. While it is true that the olfactory bulb has special features not found elsewhere in the brain, the dynamics responsible for its essential features are the same as in other sensory systems.[28] This section contains a brief survey of the neural dynamics that function in the olfactory bulb to produce odor discrimination and expectation.

The olfactory bulb is comprised of densely interconnected excitatory (mitral and tufted) cells and inhibitory (granule) cells. A process of negative feedback between the local members of these two groups[29] constitutes a neural oscillator; excitation is followed by inhibition and re-excitation. The excitatory projection neurons receive input from receptors and transmit output from the bulb downstream centripetally to the next stage of the limbic system. Segregation of input by glomeruli establishes about 2,000 'oscillators' in each bulb. The oscillators are coupled by mutually excitatory synapses.

What happens when a waking animal is given an odor? First, the odor excites the set of receptor neurons sensitive to that odor, which in turn excites a subset of the excitatory neurons in the bulb. Provided that the odor is paired with an unconditional stimulus,[30] the synapses between each pair of coactivated neurons in the subset are strengthened.[31] It is reasonable to hypothesize that while each inhalation only activates a fraction of the receptors sensitive to a particular odor and the excitatory neurons to which they project, in a training session that involves several hundred inhalations of the test odor, a nerve cell assembly of mutually excitatory cells is formed. From then on, if any subset of receptors in the training set is excited by the learned odor, the entire assembly is stereotypically activated via mutual excitation.

Each inhalation brings with it a surge of receptor activity that excites bulbar neurons to re-excite each other particularly along pathways strengthened by training. Each neuron can amplify its input such that the amplified output can grow explosively in what is known as a 'nonlinear' fashion. This is an essential feature of bulbar dynamics. When the mutually excitatory connections have been strengthened through learning, input can cause the neural mechanism to break into oscillation in a sustained, self-enjoined manner that lasts well after the input ceases.

EEG evidence indicates that this process is a self-organized process.[32] This is no accident; the olfactory bulb, like the rest of the cerebral cortex, possesses the physical characteristics required for self-organization: it is inherently nonlinear, it consists of large numbers of densely interconnected, semi-autonomous elements, and it exhibits spatial and temporal integration that allows each element to spread its influence throughout the entire assembly.

This self-organizing property of brain dynamics is significant because self-organizing systems require an explanatory framework alien to that used by other models. Self-organizing phenomena such as weather patterns, fluid dynamics, and embryonic development have resisted attempts to explain them in terms of earlier explanatory models. Traditional explanations attempt to explain properties of the system, e.g. the emergence of turbulent behavior, in terms of the properties of parts of the system, as if the elements of explanation must mirror the compositional structure of the system. Yet, explanations of self-organizing phenomena can only be given in terms of qualitative forms of behavior of the system as a whole, in terms of system properties that resist analysis in terms of the properties of the parts that comprise the system or in terms of properties of the input to the system. As far as self-organizing systems are concerned, '[f]rom the point of view of explanation there is a relative independence from the nature of the substrate',[33] i.e. microreduction, the aim of traditional explanations, does not work.

Consider a simple example of self-organization in the physical world: the emergence of convection cells in a uniformly heated liquid. When heated slowly from below, the process of heat transport at first takes place without a qualitative change in the form of behavior exhibited by the system. But after a certain temperature-threshold is attained, heat transport is no longer facilitated by this uniform state and convection cells emerge within the liquid. These cells, known as Bénard cells, constitute a more structured state, a lattice-work of octagonal cells. The emergence of this more structured state cannot be explained only in terms of molecular properties, properties that put some molecules in charge of others and that determine the end-state. Instead, the system collectively organizes itself to produce a qualitatively new form of behavior when input to the system exceeds certain parameters. Input to the system does not determine what happens, it simply sets up the conditions such that the system will adopt a new form of behavior to deal with the constraints imposed by the new circumstances more effectively.

EEG evidence gathered by Freeman and his colleagues has indicated that the olfactory bulb exhibits at least five distinctive, stable, self-organized states: a low-level background state, a carrier state, an equilibrial state under deep anesthesia, a blatantly chaotic state under strong stimuli, and a drug-induced oscillatory state that may persist for hours. Freeman postulates that the bulbar mechanisms underlying these states can be best modeled as a set of distributed, coupled, nonlinear oscillators.[34] Such a system can behave in any number of ways, but given input constraints and interaction strengths it will tend to a definable state and maintain it until perturbed. If, as in the bulb, the system returns to the same state when perturbed, its behavior is said to be governed by what topologists refer to as an 'attractor'. Attractors mathematically represent the qualitative form of behavior exhibited by the system. A change from one qualitative form of behavior, or attractor, to another is referred to as a 'bifurcation' or state change. Bifurcations require parametric changes, that is, a change in the quality of the system and not merely a change in its input.

Freeman uses attractors to mathematically model bulbar activity reflected in the EEG. His claim is that the resting EEG can be modeled as a 'chaotic' attractor, manifest as nonperiodic activity that may appear random to the observer, while burst activity manifests a 'limit cycle' attractor characterized by periodic oscillation. A bifurcation due to parametric changes occurring with learning leads the system to change from low-amplitude chaos to the high-amplitude, spatially coherent limit-cycle state, and back again.[35]

Freeman's evidence suggests that when an animal is trained to respond to a particular odor the bulb creates an activity state that can be modeled as a limit-cycle attractor. With each inhalation this more ordered state emerges from the chaotic background state and collapses back again. A separate limit-cycle attractor forms for each learned odor given under reinforcement. Each attractor-modeled activity state is defined by a distinctive form of self-organized activity, by a set of parameter values including the strengths of synaptic interactions, and by a specific domain of input referred to as its 'basin'. Random access to each of these forms of activity is achieved by the chaotic background state that maintains bulbar readiness by placing the system far from equilibrium. Each inhalation can place the system in a different basin depending upon the initial conditions set by the odor input. Most of the time the odor is the background, but the presence of a conditioned stimulus places the system in a position in which it generates its spatially distinctive limit-cycle activity instead of the stereotypic oscillation characteristic of the control state.

The upshot of this is that odor discrimination depends upon the dynamics of self-organization taking place in the olfactory bulb. The process of learned sensitization to an odor that we interpret as an 'expectation' of that odor is realized as the formation of strengthened connections among a network of neurons which, when placed in a learned input domain, self-organize to generate a distinctive pattern of neural activity here modeled as an attractor. Odor expectation at the neural level, thus, consists of a tendency to adopt a certain qualitative form of behavior under definite circumstances. This 'tendency' exists after learning and prior to receipt of the learned odor, but the animal does not require a prior active neural state which can be correlated with the odor as a 'template' for the process of odor-specific expectation. The pattern of neural activity recorded in the EEG is the *result* of a more fundamental property of the bulb, namely, its readiness to respond with a given form of behavior to a learned stimulus. The process of learning is a process of differential self-organization, of the system generating different qualitative forms of behavior to deal with significant input. Note that the input is a spatial pattern of receptor activity that differs with each inhalation, that the output is a stereotypic spatial pattern depending on past experience, and that they do not co-exist and are not 'matched' or 'cross-correlated' in any way. The dynamic output pattern is *created* within, not caused by, the conditions set by the input and the previously altered synapses.

V. Behavior without Representations

These findings have important implications for how we view the neural processes that underlie behavior. The pattern of neural activity that Freeman records in the EEG is the result of a self-organized, acquired tendency to behave in a certain way given the goals of the system considered as a whole (e.g. to avoid aversive stimuli or seek rewarding stimuli). When a learned stimulus is presented to the animal this tendency is let loose, but the burst pattern that is the end result of this tendency does not play a role in its own selection by the stimulus; it exists only after the stimulus is delivered and not before. Thus, the burst pattern does not constitute a 'search image' as once thought; the notion is contradictory. The deciding factor in brain function is not neural activity patterns *per se*, on which Churchland's model depends, but the more fundamental self-organized tendencies to adopt a qualitative form of behavior on which these activity patterns depend; not the input pattern and its transformation within the animal, but the internally generated neural dynamics created by the system itself.

Churchland's proposed system reacts and adjusts, but it does not act or create. This fact is easily lost sight of in her presentation because relevant stimuli are the only ones Roger is offered; in the real world things are not that easy. Freeman's self-organizing model departs from the reflex-based models as well as from cognitivist models that attempt to explain responses in terms of represented features of the stimuli. The ability of the olfactory bulb to respond depends on self-organizing processes in the brain, not on features of the stimulus: the bulb does not respond to unreinforced stimuli after habituation, and it responds to novel stimuli with a characteristic generic pattern that forms a distinctive class of outputs easily distinguished from learned responses.

This feature is characteristic of self-organizing systems. Consider a solution brought to supersaturation. With supersaturation the chemical solution structures its underlying dynamics such that it will respond to a class of input with a qualitatively new form of behavior. This class of input may be one that previously had no qualitative effect on the system's behavior, or it may be a class of input that the system has never before received. No matter. When the container holding the supersaturated solution is tapped, jarred, jiggled, bumped, or otherwise disturbed, it responds with a qualitatively new form of behavior, i.e. the supersaturated chemical precipitates out. Other classes of stimuli, e.g. an increase in temperature, do not have this effect. The process of self-organization results in the system being 'poised to respond' in a certain way. The class of stimuli, moreover, to which the system can respond is quite general, and the specific stimulus need never have been delivered prior to the process of self-organization. The representations of cognitivist theories, on the other hand, are always specific and lack this generality and flexibility.

The recognition of self-organizing neural dynamics in the brain is important for a further reason: it contradicts the cognitivist view that the relation between the stimulus and brain events can be equated with the relation that holds between a representation and that which it represents. The supersaturated solution does not contain a representation of the class of stimuli to which it can respond, yet it is poised to respond to one class and not others. Likewise, the fertilized ovum does not contain a representation of the adult that it will become. Self-organizing dynamics in the olfactory bulb are similar in this respect; they do not contain an active neural representation of expected stimuli prior to their receipt. For the purposes of observation, the experimenter can feel free to refer to the active state that follows the stimulus as its neural representation, but Freeman and his associates have shown that the animal itself does not need such a representation in order to expect or recognize a learned odor.

Yet, cognitive theorists could respond as follows: they might accept Freeman's data and amend their view to identify neural representations with Freeman's attractors instead of with active neural states among a network of neurons. But this approach, too, will not succeed. Representations were introduced to 'guide' or 'control' a system that, apart from these representations, is viewed as inherently undirected in ways relevant to making its behavior adaptive. Cognitivists assume that if the system is to behave adaptively, and hence successfully, in its environment, it requires an inner representation of that environment. But the olfactory bulb disconfirms this hypothesis. It does not require anything like a representation in order to behave adaptively.

If brains are truly self-organizing, then each creates its own context and coordinate systems. These will not map one to another. When we see a stereotypic EEG pattern with a reinforced odor, we can say that it 'represents' the odor, but only to us as observers and not to the animal. What it 'means' to the animal is manifested in what the animal does, in how it interacts with its environment. What we need to understand are the neural dynamics of what happens in the brain, not our interpretation of that process.

What does all this imply for a cognitivist theory of the sort that Churchland envisions as capable of unifying with neuroscience? At the very least, it indicates that cognitivism is *not* the 'only available model'[36] about processes intervening between the input and the output. Cognitivism may have been the best available theory when the only alternative was behaviorism, but the situation is changed today. Today we have access to the brain in a way that was not possible 30 years ago, and we have new explanatory models available upon which to base experimental design. Churchland urges neuroscientists and philosophers alike to remain open to new forms of explanation, new ideas, and new findings. I join her in this, and I hope that the ideas presented here will encourage just such a growth process.

If the theory I have presented is supported by further experimentation, however, it has deeper consequences for cognitivism. What it says is that the brain does not need representations in order to produce goal-directed behavior. Researchers might need to appeal to representations in order to carry out their experiments and to draw the necessary correlations between input and output, but the brain does not work on the same principles as researchers do when trying to understand what it does, or when we try to build machine models of its functioning. If the theory presented here turns out to be supported by further experimental investigation, then the kind of 'unified theory' that Churchland envisions will not be the one to describe the biological reality of brain functioning.

As far as current neuroscience is concerned the implications are that talk of neural representations is misleading. Indeed, a survey of the current literature reveals that such talk serves more to obscure our present ignorance about what the brain is doing than to inform us in any positive sense. A fresh look at the entire subject would be welcome.

The implications are also far-reaching for philosophy of mind. We find that there is a directed process that lies between the representations of cognitive theory and the mechanistic reactions of behaviorism. This realization is significant. Philosophers have traditionally attempted to explain behavior in one or the other of these terms; the wars over the mind–brain problem, materialism vs. dualism, have drawn their battle lines in terms of this dichotomy. The discovery of a middle ground should, therefore, be welcome. Indeed, philosophers such as Heidegger and Merleau-Ponty built their philosophies around the recognition of a directed, non-representational process that is responsible for behavior. The idea is not new to philosophy, and while Churchland does not discuss it in her book, neuroscientists deserve to know that that notion has a degree of philosophical respectability.

In conclusion, the point is this: there is evidence that the brain does not need representations to produce goal-directed behavior, that representations do not figure in the production of behavior. Moreover, models that are based on transformations of input, no matter how complex, will never be adequate. The brain creates, it does not just react. Churchland's model is intriguing, but there are experimental data that indicate that her model is inadequate to express the biological reality of brain functioning. And it is the biological reality of the brain, after all, that any unified theory must recognize if it is to succeed.[37]

NOTES

1 H. Putnam, 'Philosophy and Our Mental Life', in *Philosophical Papers 2: Mind, Language, and Reality* (Cambridge: Cambridge University Press, 1975), pp. 291–303.

2 See J. Fodor, *The Language of Thought* (New York: T. Y. Crowell, 1975); D. C. Dennett, *Brainstorms* (Cambridge, Mass.: MIT Press/Bradford Books, 1978), pp. 174–89; N. Block (ed.), *Imagery* (Cambridge, Mass.: MIT Press/Bradford Books, 1981), *passim*; S. Kosslyn, *Image and Mind* (Cambridge, Mass.: Harvard University Press, 1980).

3 E. Kandel and J. H. Schwartz (eds.), *Principles of Neural Science* (New York: Elsevier/North Holland, 1981).

4 A. J. Pellionisz, 'Coordination: A Vector-Matrix Description of Transformation of Overcomplete Central Nervous System Coordinates and a Tensorial Solution Using the Moore-Penrose Generalized Inverse', *Journal of Theoretical Biology* 110 (1984), No. 3, pp. 353–76.

5 A. J. Pellionisz, 'Tensorial Approach to the Geometry of Brain Function: Cerebellar Coordination Via Metric Tensor', *Neuroscience* 5 (1980), pp. 1125–36.

6 See E. S. Luscei and L. J. Goldberg, 'Neural Mechanisms of Mandibular Control: Mastication and Voluntary Biting', in J. M. Brookhart and V. B. Mountcastle (eds.), *Handbook of Physiology*, sect. 1, vol. 2, pt. 2 (1981), pp. 1237–74.

7 V. B. Brooks and W. T. Thach, 'Cerebellar Control of Posture and Movement', in J. M. Brookhart and V. B. Mountcastle (eds.), *Handbook of Physiology*, sect. 1, vol. 2, pt. 2 (1981), pp. 877–946.

8 W. T. Thach, 'The Cerebellum', in V. B. Mountcastle (ed.), *Medical Physiology*. 14th ed. (St. Louis, Missouri: Mosby 1978).

9 Brooks and Thach, op. cit., p. 882.

10 C. Sherrington, *The Integrative Action of the Nervous System* (New Haven: Yale University Press, 1906).

11 I. P. Pavlov, *Conditioned Reflexes* (Oxford: Oxford University Press, 1927).

12 W. Koehler, *Gestalt Psychology* (London: G. Bell, 1930).

13 J. Dewey, 'The Reflex Arc Concept in Psychology', *Psychological Review* 3 (1896), pp. 357–70.

14 K. Goldstein, *Der Aufbau des Organismus* (The Hague: Martinus Nijhoff, 1934).

15 Sherrington (1906), op. cit., p. 7.

16 C. Sherrington, 'Quantitative Management of Contraction in Lowest Level Co-ordination' (The Hughlings Jackson Lecture 1931), in D. Denny-Brown (ed.), *The Selected Writings of Sir Charles Sherrington* (New York: P. B. Hoeber, 1940), pp. 497–514.

17 M. Merleau-Ponty, *La Structure du Comportement* (Paris: Presses Universitaires de France, 1942).

18 D. H. Hubel and T. N. Wiesel, 'Receptive Fields of Single Neurones in the Cat's Striate Cortex', *Journal of Physiology* 148 (1959), pp. 574–91.

19 See D. O. Hebb, *The Organization of Behavior* (New York: Wiley, 1949); S. Grossberg, 'How Does the Brain Build a Cognitive Code?', *Psychological Review* 87 (1980), pp. 1–51; J. Hopfield, 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities', *Proceedings of the National Academy of Science, USA*, 79 (1982), pp. 2554–8; and J. P. Changeux, T. Heidemann, and P. Patte, 'Learning by Selection', in

P. Marler and H. S. Terrace (eds.), *The Biology of Learning* (Berlin: Springer-Verlag, 1984), pp. 115–31.

20 C. A. Skarda and W. J. Freeman, 'Behavior Without Representations: Self-Organizing Neural Dynamics', submitted.

21 W. J. Freeman and C. A. Skarda, 'Spatial EEG Patterns, Nonlinear Dynamics, and Perception: The Neo-Sherrington View', *Brain Research Reviews* 10 (1985), pp. 147–75.

22 W. J. Freeman, 'Spatial Properties of an EEG Event in the Olfactory Bulb and Cortex', in *Electroencephalography and Clinical Neurophysiology* 44 (1978), pp. 586–605.

23 W. J. Freeman and W. S. Schneider, 'Changes in Spatial Patterns of Rabbit Olfactory EEG with Conditioning to Odors', *Psychophysiology* 19 (1982), pp. 44–56.

24 See J. Atema, K. Holland, and W. F. Ikehara, 'Olfactory Responses of Yellowfin Tuna (Thunnus Alfacares) to Prey Odors', *Journal Chem. Ecology* 6 (1980), pp. 457–65.

25 W. J. Freeman, 'A Physiological Hypothesis of Perception', in *Perspectives in Biology and Medicine* 24 (1981), pp. 561–92.

26 W. J. Freeman, 'Nonlinear Neural Dynamics in Olfaction as a Model for Cognition'. Chapter for *Proceedings* of the International Conference on the Dynamics of Sensory and Cognitive Processing in the Brain, Berlin, 19–22 August 1985, to be edited by E. Basar and published by Springer-Verlag.

27 Skarda and Freeman, op. cit.

28 W. J. Freeman, 'Nonlinear Neural Dynamics in Olfaction as a Model for Cognition', op. cit.

29 W. Rall and G. M. Shepherd, 'Theoretical Reconstruction of Field Potentials and Dendrodendritic Synaptic Interactions in the Olfactory Bulb', *Journal of Neurophysiology* 31 (1968), pp. 884–915.

30 See C. M. Gray, W. J. Freeman, and J. E. Skinner, 'Associative Changes in the Spatial Amplitude Patterns of Rabbit Olfactory EEG are Norepinephrine Dependent', *Neuroscience Abstracts* 10, 36.2 (1984), p. 121.

31 D. O. Hebb, *The Organization of Behavior*, op. cit.; G. Viana Di Prisco, 'Hebb Synaptic Plasticity', *Progress in Neurobiology* 22 (1984), pp. 89–102.

32 See W. J. Freeman, *Mass Action in the Nervous System* (New York: Academic Press, 1975); W. J. Freeman, 'Techniques Used in the Search for the Physiological Basis of the EEG', in A. Gevins and A. Ramond (eds.), *Handbook of Electroencephalography and Clinical Neurophysiology* (Amsterdam: Elsevier, in press).

33 A. Garfinkel, *Forms of Explanation* (New Haven: Yale University Press, 1981).

34 W. J. Freeman and C. A. Skarda, 'Spatial EEG Patterns, Nonlinear Dynamics, and Perception: The Neo-Sherrington View', op. cit.; A. Garfinkel, 'A Mathematics for Physiology', *American Journal of Physiology* 245 (1983), 14 ('Regulatory Integrative Comp. Physiol.'), pp. R455–66.

35 W. J. Freeman and C. A. Skarda, 'Spatial EEG Patterns, Nonlinear Dynamics, and Perception: The Neo-Sherrington View', op. cit.

36 Fodor, op. cit., p. 51.

37 I am deeply indebted to my colleague, Walter Freeman, without whose help I could not have written this paper, as well as to Mike Shadlin, Lou Goldberg, and especially Alan Garfinkel.

Received 26 October 1985

Christine A. Skarda, Department of Physiology-Anatomy, University of California, Berkeley, California 94720, U.S.A.