

Journal of Philosophy, Inc.

Neurophilosophy: Toward a Unified Theory of the Mind/Brain by Patricia Churchland

Review by: Kathleen A. Akins

The Journal of Philosophy, Vol. 87, No. 2 (Feb., 1990), pp. 93-102

Published by: [Journal of Philosophy, Inc.](#)

Stable URL: <http://www.jstor.org/stable/2026883>

Accessed: 18/09/2012 19:14

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Journal of Philosophy, Inc. is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Philosophy*.

<http://www.jstor.org>

BOOK REVIEWS

Neurophilosophy: Toward a Unified Theory of the Mind/Brain. PATRICIA CHURCHLAND. Cambridge: MIT Press, 1988. 546 p. Cloth \$27.50.

In the introduction to *Neurophilosophy*, Patricia Churchland states that its central task is to address “the overarching question of the nature of a unified, integrating theory of how, at all levels of description, the brain works”—that is, to provide good reasons why a unified theory of the mind/brain is possible and probable, and to make a start in defining that theory. Churchland’s secondary goals concern her audience. Because the central theme of the book is unification, Churchland seeks to address both philosophers and neuroscientists, to educate both groups. On the home front, she hopes to give philosophers a sufficient background in neuroscience to sustain the examples used and to enable philosophers to forage the literature themselves; for the neuroscientists, the goal is to provide a “get-around” knowledge of the philosophical issues.

Significantly, given our materialist goals and tendencies, *Neurophilosophy* is the first book of its kind. It is a vital and timely project—moreover, one to which Churchland’s talents are uniquely suited.

Neurophilosophy does not, however, accomplish all that it attempts. The central reason for this failure is not philosophical per se. Rather, given the magnitude of the goals coupled with the division in audience, the scope of this project is too wide—what *Neurophilosophy* does is to compress the work of several volumes into one. Desiring breadth, it has sacrificed detail, the result being a work which is both philosophically sketchy and, at points, unclear in its intent.

That said, this review will begin with a brief discussion of Churchland’s theory of the mind/brain. My aim is to say how the agenda for the book is set by the philosophical view—to say, on the theory’s own terms, what it would take to complete the project—and then to contrast this with what Churchland has actually written. The second half of the review is a discussion of a major piece of the project, Churchland’s theory of content.

I. THE PROJECT

At the foundation of Churchland’s view of the mind/brain are two central theses. The first thesis, from the philosophy of science (Pierre Duhem, W. V. Quine, Imre Lakatos), states that the terms of a scientific theory have meaning only within the context of that theory.

There are no “pure” observational predicates that take their meaning “from their referents”; there are only terms which are embedded in the theory as a whole, which are “theory-laden.” The second thesis, ascribable to Wilfrid Sellars, is about the nature of our understanding of mental events. Here, we are asked to regard the compendium of our knowledge of mental states as a theory—and, moreover, a theory whose hallmark features have been modeled, rightly or wrongly, upon the “visible” or intersubjective features of language, on utterances, not upon the results of introspection. Putting these two theses together yields the following conclusion. Just as there are no pure observational predicates in theories that purport to describe the external world, the inner eye offers no privileged “theory-free” insight about the nature of mental events.¹ Hence, *qua* theory, folk psychology might well be wrong, revisable in the light of a better theory—in this case, in the light of our future neuroscience.

Starting with these theses, the question of the future of psychology, of the “scientific” and “folk” variety, arises. Churchland’s central claim is reductivist. When all is said and done, the phenomena described by scientific psychology will be explained by the functional entities postulated by a future neuroscience. Even though a traditional “reduction” (of the entities postulated by the one theory to the entities of the other) will probably not be forthcoming, an understanding of the micro-level phenomena (neurophysiological events) will explain how the macro-level events (those described by scientific psychology) come about. On the other hand, *vis-à-vis* the properties and entities of our present folk psychology, the new theory may well prove less kind. Here, elimination (as opposed to reduction) is the most plausible outcome, although, as Churchland emphasizes, this issue is purely empirical.

A commitment to the results of science notwithstanding, Churchland has her money firmly placed: she takes the recent data to suggest strongly, if not quite an outline of the new theory, at least an eliminative scenario. Here, we are likely to find that: (a) many folk psychological categories, such as memory and learning, will not be legitimated by the new functional categories of neuroscience (i.e., there will turn out to be many diverse and different mechanisms for phenomena we took to exemplify one psychological category); (b) the form of the mechanisms underlying thought will be in no way “sen-

¹ I should note here that, although Sellars does arrive at this same conclusion, it is not via the same route: Sellars’ argument depends upon a theory of thought coupled with a separate theory of sensation. This correction was pointed out to me by Wright Neely and I thank him for it.

tential” (in Jerry Fodor’s sense of the word); (c) the folk notion of “consciousness” will not delimit a natural kind; and (d) on the new theory, although it will still be legitimate to speak of the determinate, objective (albeit relational) meaning of a neural state, we will probably not be able to express that content as a propositional attitude—mental content will be best stated in neurophysiological/computational terms. (This is only a partial exposition of Churchland’s scenario.)

Given the above project, then, what would be necessary to give it substance? What “books” are required in order to complete Churchland’s philosophical and theoretical project?

(1) *Reduction and the Unified Theory of Mind*. This book gives an exhaustive accounting of the reductive strategy, including a defense of the central theses given above, the details of the envisioned theoretic reduction, and responses to its standard philosophical objections. Here, we want to know the details. What, for example, is the relation of folk theory to scientific theory? Assuming that we do not represent our folk knowledge by means of mental sentences, in what ways is folk theory a theory? And so on. This is philosophy of mind in the traditional mold.

(2) *Why Neuroscience Matters*. Part of the problem with claiming that we can treat all knowledge of ourselves as empirical is that the conclusion, however well-argued, remains intuitively unconvincing. Examples are needed. It is only by saying how specific folk psychological theories/presuppositions/dogmas about the mind are now in question, or which ones would be, given specific neuroscientific results, that the eliminativist position acquires bite. It is in this book, then, that the doctrines of folk psychology are made explicit and we are shown how the scientific data tell against them.

(3) *The Unified Theory*. This book represents a “best bet” about the new theory of the mind/brain. Building upon the conclusions of book 2, the elimination and revision of folk concepts, it gives the positive view—the kind of neurophysiological/computational theory that could incorporate the overall framework.

Next, as Churchland suggests, if the project is to reach both neuroscientists and philosophers, there is the problem of audience, of bringing the two groups up to speed. This would require two texts.

(4) *Neuroscience for Philosophers: A Guide to the Principles, Techniques, and Current Findings of Neuroscience*. For those philosophers who are skeptical of the project at the outset, such a lengthy primer will seem an onerous and unnecessary burden. Nonetheless,

learning the neurobiology is not merely a matter of gaining the wherewithal to understand specific examples. Rather, if Churchland's major goal is to find a theory of mind that "makes biological sense," it will be difficult to explain, at any critical juncture, the drift of that phrase. 'What makes biological sense' refers to the kinds of solutions evolution has found to informational/environmental problems. And just as grasping the nature of, say, proof theory is best accomplished by doing (or going over) a number of proofs, understanding the principles of neuroscience is best achieved by walking through a number of evolutionary "solutions." Hence, in order to understand the project as a whole, a neuroscience primer is needed.

(5) *Philosophy for Neuroscientists: An Introduction to the Philosophy of Mind and Science*. Unlike neuroscience, where there is an accepted body of introductory "facts," philosophy has no agreed-upon set of best theories. Not that it would be impossible to select a few "very bad ideas," those ideas we love to hate. Still, because most current theories are highly controversial, an introduction to the philosophy of mind and science would include a summary of alternative views, with the author's biases clearly flagged as such.

Let me turn now to *Neurophilosophy*. Part I, approximately the first half of the book, is a neuroscience primer (book 4) supplemented with some hints as to which experimental areas the author expects will prove theoretically interesting. This is a job well-done—a lucid and knowledgeable work.

Part II is an overview of the philosophical issues, "Recent Developments in the Philosophy of Science." In these chapters, Churchland gives a short history of the philosophy of mind, sketches the outlines for a theory of reduction, and takes up some standard objections. There is also a chapter on the antireductionist sentiments of "functionalist" theories or, more properly, sentential/computational views. The central problem here is that Churchland has combined books 1 and 5, the introduction to philosophy with the argumentation for the philosophical project. As one might suspect, this is a less than harmonious union: often the line between explication and argument is not clearly demarcated (of particular trouble to the neuroscientific reader); for the most part, the arguments needed to sustain Churchland's view are given only in skeletal form, referring the reader to previous articles for further detail.

Part III, the shortest section, is designed to cross the divide. Churchland presents three very interesting and suggestive neurophysiologically-based theories—tensor-network theory, the vestibulo-

lar-ocular reflex, and PDP models—each chosen in the hope that it might exemplify “the foundations of a new paradigm for characterizing representations and computations.” The extent to which the reader will find these theories convincing, however, depends largely upon her previous sympathies for specific eliminativist/reductivist claims. More generally, without the folk views held firmly at bay, it is hard to see these theories as genuine alternatives.² Good arguments for the eliminativist claims—book 2 above—must first be given. Book 2, however, is not a part of *Neurophilosophy*.

II. THE QUESTION OF CONTENT

In this section, I want to talk about Churchland’s theory of content. As many people have said, *Neurophilosophy* contains—at least, explicitly—very little about the subject. Why, then, discuss it? Most importantly, part III assumes the success of Churchland’s philosophical view: the neurophysiological theories are largely suggestions about how to implement or realize that theory. In virtue of Churchland’s view that the folk notion of propositional content will be replaced by a two-part theory of nonintentional representations and computational relations, the new scientific theory acquires its goal—namely, to supply “the foundations of a new paradigm for characterizing representations and computations.” The theory of content sets the agenda.

On the Churchland theory of content (a view held and explicated by both Patricia and Paul Churchland³), folk psychology is viewed as having conflated two separate notions of content. On the one hand, there is “translational” content. Roughly speaking, this is “what the agent has in mind”—*how* the agent represents an external event to herself. So, for example, if I exclaim, “Zeus’s thunderbolts!” (sincerely and with feeling), whenever the gas oven lights with a bang, the translational content of the thought is about Zeus. On the other hand, because that thought—in whatever guise—reliably correlates with the ignition of gas, its “calibrational” content is about gas explosions. That is, I think *of* the explosion (calibrational content) *as* an act of Zeus (translational content).

On the Churchlands’ reductive view, the translational content of a neural state is identified with the “inferential/computational rela-

² For example, without first explaining in naturalistic terms (or explaining away) the intentionality of mental representations, it is hard to see how such states could possibly be “positions in phase space” (452). This issue is discussed in more detail below.

³ See, for example, Paul M. Churchland, *Scientific Realism and the Plasticity of Mind* (New York: Cambridge, 1979); and P. M. and P. S. Churchland, “Functionalism, Qualia, and Intentionality,” *Philosophical Topics*, XII (Spring 1981): 121–145.

tions it bears to all the rest of (the agent's) representations." Just as a linguistic term has meaning in virtue of its place within the "semantic net," so too is the meaning of a mental state a function of its inferential/computational relations to other representations. And representations—for one must now discharge the remaining intentional term—are those neural states with calibrational content. They are those states which reliably correlate with some property of the environment. Both theories of "meaning," of computation and of "representation," then, are placed squarely within the intellectual domain of neuroscience.⁴

Let me take a moment to put this theory into philosophical context. Despite the theory's origins, it is clear that the Churchland view differs significantly from those of Quine and Sellars. Quine, as we know, took the failure of the analytic/synthetic distinction plus the indeterminacy of translation to show that we ought to dispense with propositional attitudes and their semantic kin—or rather, that we would be wise to do so in our attempts to "limn the true and ultimate structure of reality." If there is no empirical solution to the problem of translation, then there is no sense in speaking of "a fact of the matter" about meaning—no phenomenon, "meaning," which stands in need of an explanation over and above that given by a theory of translation. Sellars' view is no less revisionary. If we are to regard our lore about mental states as fundamentally dependent on an analogy to linguistic practice, then it is open to question whether any "linguistic" properties are rightly attributable to the mental—and this includes both the notions of meaning and reference.

Contrast this with the Churchland view. It follows from the above that at least some neural states are meaningful, they yield a determinate semantic content, can be assessed for "synonymy," and are both representational and "referential."⁵ Clearly, the Churchlands

⁴ One important caveat should be noted. The Churchlands do not think that, thus, the folk psychological notion of content is reduced to neural/functional statements. Along with Stich, the Churchlands hold a "you have what I have" theory of content ascription. When I ascribe the belief that *p* to another agent, I assert (tacitly) that she possesses a "semantic analogue" to whatever state I have when I believe that *p*. I assert that she has a neural state that plays a cognitive/inferential role similar to some state in my own representational "economy." Nonetheless, a folk ascription does not serve to reliably "point at" some type of neural state. First, content ascriptions are both contextually sensitive and vague; second, folk ascriptions conflate calibrational with translational content (i.e., ascribe content on "nonindividualistic" grounds). The end result is a set of contradictory ascriptions—an inconsistency within folk theory that the Churchlands believe will preclude intertheoretic reduction.

⁵ Although the conclusions of the Churchland theory appear thoroughly un-Quinean, the roots of the divergence are to be found in a disagreement about the

believe that we need a theory of meaning—that something needs to be explained—and they want to do so using largely traditional semantic concepts. They are not attempting, however, to explain the folk notion of content, to say how, in the naturalistic order of things, certain neural states come to have a particular propositional content. Theirs is a theory of “meaning”—but not of propositional content ascription.

What, then, are the phenomena that the Churchlands want explained and, assuming the robustness of the phenomena, how good are those explanations?

Let me begin with the notion of calibrational content. Here the phenomenon is not intentional states. Rather, what calibrational states—“representations”—explain is simply how we get about. Churchland writes,

I take it as obvious that if there were no systematic relations between the external world and states of the brain, the animals could not survive or if they did it would be a miracle in the strict sense of the word. An owl would not know where the mouse is and he would not intercept it as it runs.⁶

There is not much to quarrel with in this quotation. I want to point out, however, that there is a significant difference between a systematic relation and the kind of correlational (representational) state to which the Churchlands appeal in their explanation of content. Imagine, for example, a simple creature with a digital instrument that registers set increases in light intensity. The function of the instrument is to control the amount of light into the eye, which can be done either with “shades” (i.e., filters) or by closing the lens aperture of the “eye.” It works as follows. Each registered increment in intensity causes a filter to be lowered over the “eye.” When all the ten filters have been used successively (i.e., ten intensity increases have been registered), the “eye” aperture clicks down one notch. The filters go back to starting position, the intensity meter falls back to zero, and the process begins again, with the ten filters being lowered over a new aperture. For this creature, the light intensities in the world cause systematic state changes in the light sensor, but the increments on the “meter” do not correlate with any particular light

proper method of translation. This way of framing the disagreement was suggested by Patricia Churchland in Paul Churchland's *Scientific Realism and the Plasticity of Mind*, sect. 9, pp. 63–74.

⁶ “Replies to Comments” from “Symposium on Patricia Smith Churchland's *Neurophilosophy*,” *Inquiry*, xxix (June 1986): 139–273, p. 260.

intensity. Once the meter is reset to zero, its settings (alone) no longer indicate the actual value of light intensity. Functionally speaking, then, this value does not matter to the “motor” mechanism. What this organism has, in other words, is a simple “servo-mechanism.”

Systems of the above kind have systematic relations with the environment. But they do not function in virtue of correlations *qua* correlations. Contrast this with, say, our best guess about one function of the retinotopic map. Given that light travels in straight lines, any given place on the retina correlates with a spot in visual space. A place on the retina “represents” a place in the world. If we have guessed right, these co-ordinates are used, *qua* co-ordinates, for various tasks, say, to guide motor movements. For example, on one version of the tensor-network theory, movement is made possible by functions that map coordinates encoded in sensory space onto co-ordinates in motor space. In such systems, it is states *qua* representations (unchanging correlations) and transformations between co-ordinates that do the explanatory work.

Why does this matter? Most anyone would agree with the conjecture that the brain has systematic relations with the environment. But that these relations consist in functional correlations is a much stronger empirical hypothesis. The question is: Which view does Churchland endorse? If the former weaker view, then our understanding of “content” is not furthered, for “content” requires an indicator relation. Nor does postulating systematic relations give neuroscientists any mandate for future research—it does not lay down a theoretical basis for explaining how we get about.

If Churchland endorses the stronger thesis,⁷ however, there is a clear link with the question of content—functional correlations with external properties provide the neural analogue of reference. Still—and this is what I want to emphasize—the representational view is a strong empirical hypothesis with ready alternatives. If, as a matter of empirical fact, the brain manifests “servo” type mechanisms *and* “representational” type mechanisms *and* mechanisms for keeping track of and comparing internal states *and* mechanisms of functional kinds we have yet to discover, then very few of the brain’s states and processes can be characterized by Churchland’s theory. My question here is not whether to call these various functional types “representations.” I do not think it much matters. Rather, I am arguing that, if the basis of brain function is not of the kind the Churchlands sug-

⁷ This would seem to be the case. All the vector-tensor systems suggested by Paul Churchland (for such diverse sensory tasks as color vision, olfaction, face recognition, and vision) are of this form.

gest, then we would not be able to rest our explanation of content solely upon this one kind of function, the “representational” state.

What, then, of “translational” content? Here the phenomenon to be explained is the “theory-laden” nature of mental states. As Paul Churchland has put it,⁸ this is the problem of how the semantic content of a state can diverge from its causal couplings to the world—how the brain interprets its signals as representing one property rather than another. What, for example, is going on such that I see the gas explosion as an act of Zeus? Why might someone else see that same explosion as an explosion? As I said above, the Churchlands’ answer is in terms of computational relations: for a representation to be theory-laden is for it to occupy a particular place in the computational “web.”

To judge this answer, take a simple neurological example, that of the vestibular-ocular reflex (see *Neurophilosophy*, pp. 433–444 for a discussion and competing theory thereof). The problem that the VOR solves is how to keep the eyes fixated on one part of the world despite head movements—it must rotate the eyes in the opposite direction of the head movement but at the same angular velocity. In understanding this process, David Robinson⁹ has postulated a *neural integrator* (“integration” in the mathematical sense). First, the signal, which emerges from the vestibular system (semicircular canals), encodes the velocity of the head movement in the plane of the semicircular canals. This signal is relayed directly to the motor neurons. But in order to move the eyes, both the desired velocity and the desired deviation—position—are needed. Hence, a second parallel line goes to the *prepositus* nucleus where, through integration, position is calculated from velocity. This information is then relayed on to the motor neuron for a completed output signal to the eye muscles.

Given that neurological computations of this sort will probably prove quite common in brain function, how does understanding this kind of neural system clarify the nature of “theory-laden” states? At first the signal encodes velocity information; after integration, it encodes position. But what is velocity represented *as* at any given

⁸ “Semantic Content: In Defense of a Network Approach,” a reply to Kenneth Sayre’s “Intentionality and Information Processing: An Alternative Model for Cognitive Science,” *Behavioral and Brain Sciences*, ix, 1 (March 1986): 121–155, p. 141.

⁹ For a review of this literature see his “Control of Eye Movements,” in *Handbook of Physiology, The Nervous System*, II, 2 (Baltimore: Williams & Wilkins, 1981): 1275–1320; for an informal lecture by the same author, see “The Windfalls of Technology in the Oculomotor System,” in *Investigative Ophthalmology and Visual Science*, xxviii (December 1987): 1905–1924.

time? Is it represented as velocity (“change in distance as a function of change in time”?) or as something else, and if so as what? Even for this simple system, it is unclear which are the telling facts.

There is, of course, something unfair about this way of posing the question. If the Churchland theory does not attempt to ascribe propositional content to neural states, we cannot in fairness demand to know what velocity, or anything else, is represented as. We cannot demand to know the propositional content of the individual neural states. The very goals of the theory preclude this line of criticism.

Nonetheless, there is, by the Churchlands’ lights, a phenomenon to be explained. What that phenomenon is, however, is none too clear. On the initial phrasing of the problem, we are trying to understand the “theory-ladenness” of neural states, how the content of a neural state “diverges” from its causal origins. But there is, of course, no physical thing, the meaning, which is carried about, divided, and reshaped by the internal workings of the brain. In more general terms, we might ask how a neural state “comes to have meaning for its owner” or how a stimulus is “interpreted” by the brain. Again, these phrases are largely metaphorical. We do not have any clear sense of what neural “interpretation” or “meaning” amounts to. The task for the reductivist/eliminativist, then, is to understand what processes, if any, are pointed at by these (quasi) folk phrases.

Referring to computational relations as a whole does not solve the problem. Assuming that there is a principled physiological difference between you and me, when I think, “Zeus’s Thunderbolts!”, and you think, “Gas Explosion!”, the explanation of our capacities will advert to (among other things) various principles of functional organization. Whereas the difference at hand may be explained by the particulars, the general ability to “represent as” will be understood with reference to functional kinds, common to different systems, at various levels of organization. The reason, then, that the VOR example appears so empty is that we have only the particulars. Given no principles, we are no closer to unpacking the folk psychological notions.

What the above problems with the theory of content illustrate is that explaining and making plausible an eliminativist/reductivist view is an enormous task. Although *Neurophilosophy* does not supply the details, it contributes a vantage point from which to define that project. Being the first and leading step, we should acknowledge it as such.

KATHLEEN A. AKINS

University of Illinois/Urbana-Champaign