Brainwork: a review of Paul Churchland's *A Neurocomputational Perspective**

ROBERT N. MCCAULEY
Department of Philosophy, Emory University, Atlanta, Georgia 30322 USA

ABSTRACT *Taking inspiration from developments in neurocomputational modeling, Paul Churchland develops his positions in the philosophy of mind and the philosophy of science. Concerning the former, Churchland relaxes his eliminativism at various points and seems to endorse a traditional identity account of sensory qualia. Although he remains unsympathetic to folk psychology, he no longer seeks the elimination of normative epistemology, but rather its transformation to a philosophical enterprise informed by current developments in the relevant sciences. Churchland supplies suggestive discussions of the character of knowledge, simplicity, explanation, theory, and conceptual change. Many of his treatments turn on his prototype activation model of neural representation, which looks to the notion of a 'prototype' as it is employed in the psychological literature on concept representation, however, this and other features of Churchland's neurocomputational program do not square well with some of his views about cross-scientific relations.*

1. Neurocomputationalism in the philosophy of mind and the philosophy of science

Occasionally, someone does something genuinely creative in philosophy. Such accomplishments are rare enough that when they do occur they are not too hard to spot. If doubts had remained, the publication of Paul Churchland's *A Neurocomputational Perspective: the nature of mind and the structure of science* surely secures his position as one of today's most creative philosophers. This is not to imply that Churchland's work does not reflect others' influences—Sellars and Feyerabend come especially to mind. Still, with Patricia Churchland (1986), he is pursuing approaches to epistemology, the philosophy of science and the philosophy of mind that are empirically informed, philosophically promising and refreshingly imaginative.

Among philosophers, traditional critics have remained most skeptical about the claims for philosophical promise; more friendly, naturalistically minded critics, among whom I include myself (McCauley, 1986a) have worried about what frequently appears as unnecessary constraints on the range of empirical information to which the Churchlands are wont to attend. In addition, among this second group

*Although I am not sure that he agrees with even a word of this paper, I am grateful to Dick Neisser for many helpful comments on an earlier version.

various philosophical and scientific critics are cautious about the relative exclusivity of Churchland's zeal of late for connectionist models of mind. Still, the arresting imaginativeness of Churchland's arbitration of neuroscientific, connectionist and traditional philosophical notions is unassailable.

In sections 2 through 4 I shall lay out the major tenets of Churchland's neurocomputationalism, raising occasional questions, arbitrating some tensions in the texts and, where it seems helpful, filling the position out in small ways along lines Churchland indicates. Since the position is both sweeping and complex, space will permit examination of but a few of Churchland's arguments. Section 5 focuses on problems neurocomputationalism raises for Churchland's own account of cross-scientific relations.

2. The road to neurocomputationalism

Neither our familiarity nor our fluency with our own version of common sense diminishes in the least its conjectural character. Typically, we are blind to this—regarding our common sense pronouncements as obvious truths. We can usually gain glimpses of the deeply theoretical character of our own common sense perspectives only when we confront alternative versions of common sense from different cultures or different times. In *those* contexts even most of Churchland's critics readily endorse his criteria for the assessment of conceptual frameworks in terms of their relative explanatory, predictive and manipulative success and their overall coherence with everything else we know. When conceptual schemes we deem deficient on these counts conflict with our own positions (common sense or not), we have no trouble envisioning their elimination. The demon theory of disease is wrong and the demons it countenances do not exist; similarly, Stahl's account of combustion is false and phlogiston is no more real than the demons.

So far, so good—but now the argument proceeds down a path some find frightening. The next set of turns is the most difficult to negotiate (in both senses). Since these deflationary assertions hold no less for our common sense presumptions about human psychology than they do for common sense presumptions in other areas, our common sense (or folk) psychology is, ultimately, also conjectural to the core. On *most* fronts Churchland holds that our folk psychology constitutes a profoundly inadequate theory of human behavior and mental life. (A summary of his principal arguments follows in the next paragraph.) Our common sense psychology operates with an ontology of proposition-like mental contents and attitudes toward them that is strikingly discontinuous with the assumptions of our most accomplished explanatory theories in these domains—especially those in neuroscience. The critecriteria Churchland propounds (summarized in the previous paragraph) are for the adjudication of clashes between *any* conceptual frameworks—stretching from the most informal aspects of common sense to our most ceremonious theories in science, and such clashes sometimes *do* result in the total eradication of conceptual frameworks—when they prove utterly irreconcilable with those we judge their clear superiors. Churchland foresees such an elimination of most of folk psychology in deference to the superior explanatory, predictive and problem solving power of theories in neuroscience. Churchland's judgment that this will prove the likely fate of folk psychology and of the views of language, cognition and knowledge it informs is a conclusion most philosophers find unwelcome.

Among Churchland's principal targets are those epistemological analyses for which the logical relations *between* and our manipulations *of* sentences (or propositions) are of central significance. Propositional knowledge is but a small and, in some respects, rather unusual part of the knowledge we command. Epistemological positions that construe knowledge in terms of some variation on the classic formula of 'justified true belief' (where beliefs require propositional formulation) supply few resources for explicating vast areas of human knowledge, including most of our intuitions and skills—both motor and intellectual (see Johnson 1987). Moreover, such positions are not even Darwinian in spirit. Incapable of illuminating the continuities between our favored forms of knowledge and those of either preverbal infants or members of other species, these traditional epistemological positions also rely fundamentally on the categories and assumptions of folk psychology, which, Churchland contends, offer:

(1) little explanatory insight into many relevant phenomena (such as mental illness, imagination, sleep, memory, intelligence differences, etc.);
(2) a history of 'retreat, infertility, and decadence' (1989a, p. 7*) as the inspiration or basis for a research program in psychological science (no matter how narrowly or broadly the concepts 'research program' and 'science' are construed); and
(3) few connections with the other theories in science that address behavior and cognition. (Occasional comments, such as some on pp. 102–108, 133–134 and 196, however, indicate that Churchland does acknowledge the possibility of *some* connections.)

Earlier, Churchland (1979) had seemed to forecast an end to normative epistemology, once an advancing neuroscience secured the elimination of folk psychology. That position earned extensive criticism in the subsequent decade (e.g. Putnam, 1988). In 1981, though, Churchland had already noted (pp. 16–17) that what would likely disappear was neither normative concerns nor normative judgments, but only the conceptual scheme of folk psychology within which we currently formulate them. Not only are the conjectures of common sense psychology open to revision in light of scientific change, so too are our normative conjectures in epistemology, which rely on the concepts of folk psychology for articulation. Churchland maintains that the substitution (without remainder) of descriptive for normative discourse was never the aim in naturalizing epistemology. (Quine 1990 has argued similarly.) What naturalists (p. 196) most wish to highlight and defeat are the stagnating consequences for epistemological research of insisting (at least, *de facto*) upon its autonomy (see, for example, Siegel, 1984).

Churchland's recent work offers ample evidence of his willingness to employ normative epistemic concepts such as 'truth'. Many of his critics have assumed, in

*All subsequent references to this text will include page numbers only.

the light of Churchland's fervent naturalism, that this amounts to glaring inconsistency. In a similar vein others (Baker, 1987; Putnam, 1988) have pressed one version or another of a *reductio* argument which declares that Churchland's position cannot be true, because its truth would obliterate the concept of 'truth'. The argument is usually some variation or other on the following. The elimination of folk psychology would leave enquirers with neither beliefs nor intentions to express them. So, Churchland's own assertions are just meaningless strings of sounds, and such noises are not candidates for truth. Since this holds for all assertions, there are no candidates for truth, rendering 'truth' an empty concept.

Churchland offers two different, but related responses to such arguments. The first embraces his critics' conclusion while trying to strip away its pejorative connotations. Not the acquisition of truth but rather the 'ever more finely tuned administration of the organism's *behavior*' (p. 150) may be the primary aim of cognition. The overlap may prove substantial, but the connection is not a necessary one. Moreover, if the underlying dynamics of cognition are not sentential, then neither sentences nor sentences' salient epistemic virtue (viz., truth) may be the most penetrating concepts with which to conceive our cognitive accomplishments. Churchland insists, however, that this *raises* our epistemic standards by searching for goals *more valuable* than truth (see Churchland, 1993a).

Churchland's second reply claims that such *reductio* arguments presume a view of meaning that he does not accept. Indeed, when defenders of folk psychology employ this counter-argument, Churchland argues that they assume the adequacy of theories of meaning rooted, in part, in the folk psychology that is at issue (thereby, begging a critical question). The incoherent outcome of the *reductio* argument can impugn that theory of meaning at least as easily as it can threaten the premises peculiar to Churchland's view. On the face of it, Churchland's assault on folk psychology, outlined earlier, increases the probability of the first option.

The relative success of these replies is not the issue, but rather (1) that neither reply expunges the normative, and (2) that neither requires the obliteration of the concept 'truth'—though both foresee its supplementation. Churchland raises the question of whether normative reflection can occur within conceptual schemes that are not wedded to the system of categories and commitments that common sense psychology provides. He suspects so.

If our perception of the external world can undergo transformations on the basis of our new theoretical commitments, there is no compelling reason to think that the same cannot occur with the categories of inner sense, i.e. with the categories by which we conceive our mental life. Churchland invites us to 'consider … the possibility of learning to describe, conceive, and introspectively apprehend the teeming intricacies of our inner lives within the conceptual framework of a matured neuroscience …' (p. 54) as opposed to that of the orthodox propositional attitudes of our common sense psychology. Section 3 summarizes some of Churchland's projections.

3. Making the first month's payments

Churchland makes some suggestive initial instalments on this promissory note he has incurred. He no longer merely envisions accounts from cognitive neurobiology for issues that have interested philosophers, he has begun to *supply* them. The general strategy of his neurocomputational approach is to reconceive both human psychology and the philosophy of science within the framework of promising theories of neural functioning. One major theme is that the network architecture of the brain permits the generation of astonishingly vast representational spaces (which, for some purposes, can themselves be mapped onto one another). These representational spaces arise in virtue of neural networks' instantiations of operations in matrix algebra.

Churchland offers revealing accounts of motor coordination in terms of multiple state space sandwiches and neural matrices that transform coordinates from one representational system (e.g. the visual system) into those of another (directing motor skill). The joint profile of architecture and function fits the superior colliculus and cerebellum reasonably well.

Employing the concept of neural state spaces Churchland also offers a plausible account of both a vast range of intuitions we have about *our* sensory qualia and the relative sensory sensitivities that *members of other species* display. Although he does not explicate what it is like to be a bat, Churchland does a fair job of outlining what some aspects of taste are like for a rat or a cat. For our own sensory experience the story Churchland tells looks much like the classical identity theory.

Frank Jackson's (1986) qualia-based objections to physicalism maintain that it cannot account for the 'feel' of much conscious experience and of sensory experience in particular. However, since he has never accepted the view that even most knowledge is propositional, Churchland is not troubled by the notion that we may know many things directly, including, perhaps, our sensory qualia, by *acquaintance* only—where our knowledge of such phenomena by *description* can only be characterized as indirect (by comparison). Churchland (p. 70) accepts the possibility that Jackson's sensorially deprived neuroscientist, Mary, lacks some knowledge in virtue of the fact that she has, throughout her entire life, been systematically deprived of seeing red. This would still be true even if Mary might know (by description) all of the propositions expressing all of the physical truths of the universe.

A point about which, most of the time, neither Churchland nor his critics are sufficiently clear is that their criticisms employ language that can point to either of *two* sorts of phenomena. What-it-is-like to see the redness of a ripe tomato (i.e. the feel of that sensory experience) in addition to what the redness of a ripe tomato looks like are each things that Mary may not know. Churchland holds that what the redness of the ripe tomato looks like is identical with a triplet of electromagnetic reflectance properties of said ripe tomato (see pp. 56–57), whereas how that experience feels, i.e. what-it-is-like to see the redness of the ripe tomato, is identical with the pattern of neural activation (in response to stimuli like that same triplet of electromagnetic reflectances) that defines the relevant location in the brain's color state space (see pp. 103–105).

The latter identity claim may seem particularly counter-intuitive. The assertion that the feel of seeing red just *is* some brain state will, no doubt, leave many feeling short-changed. However, if anti-physicalists insist that their foes provide a (propositionally formulated) physical theory that will, as it were, recreate the *feel* of the experiences in question, then they are almost certainly asking for the impossible. If anti-physicalists respond by stressing that *that* is just their point, then Churchland offers good reasons for suspecting that they demand too much of physicalism.

Churchland maintains that knowledge by acquaintance of the feel of sensory qualia need not menace physicalism, if a *physical* account of *that form of knowing* the feel of the qualia in question is available—and that is just what his discussion of the neurally represented color state space does for the *feel* of seeing red. Anti-physicalists' appeals to their intuitions about the immediacy (or the privacy or the vividness or the uniqueness) of such feels will prove less and less convincing, if the hypothetcal identification of the feel with an appropriate pattern of neural activation yields further insights about or systematic explanations of aspects of our mental life, of our brain function or of their connections with one another.

Churchland's discussion of the neural representation of color qualia seems to do just that. For example, when two colors are perceived as closer to one another than either is to a third, those relations are preserved, *as a direct function of the details of their neural representation*, in the activation space defined. If this hypothetical identity generates a progressive program of research leading to unified explanations and richer understanding, it will not take long for it to overshadow dogged insistence about the alleged metaphysical implications of intuitions (McCauley, 1981) [1].

For more centrally cognitive phenomena, Churchland appeals to insights that arise from connectionism, which he regards as the most promising of recent approaches to neural modeling. The important bonus of connectionist designs is that they offer Churchland a mechanistic means, which preserves some prominent features of neural structures, for explicating virtually all of the theses in the philosophy of psychology he has advanced—as well as a number of central notions in the philosophy of science.

Thus, neurocomputationalism takes the knowledge an organism possesses as the array of synaptic connection strengths between its neurons. The general assumption is that distributed rather than localized systems more closely approximate neural functioning. Churchland reviews an army of advantages that accompany such conceptions (also see Bechtel, 1987). The following is but a sampling. Unlike propositional models, a neurocomputational approach accounts for the graceful degradation in the cognitive performance of organic systems in the face of impoverrished input, minor damage, disease and aging. It readily explains the rapidity of so much recall, since it involves no lists of propositional entries in memory to be surveyed (never a very plausible model for creatures who have no command of propositions) but only the activation of networks as the natural and immediate outcome of cues. Connectionism also makes sense of how individuals with different training histories and idiosyncratic arrays of connection strengths can still partition their activation spaces into what are basically the same set of conceptual divisions relative to the inputs. For example, in supervised learning neither the specific sample of training items nor their order of presentation usually matter much.

Churchland indicates that the activation spaces that arise from neural computations are frequently partitioned in ways that broadly coincide with many of our standard concepts. (Examples Churchland cites include the distinctions between the various colors and between vowels and consonants.) This consonance between neurocomputational and traditional accounts of the products of our cognitive activity will tempt epistemologists into confining their enquiries to the conceptual level. Churchland demurs. He thinks exclusive focus on partitions would obscure the fact that systems with virtually identical partitions would still react and learn differently in response to novel inputs. Systems which differ substantially at the level of synaptic weights may still support quite similar partitions of their activation spaces, thus, principles of cognitive *change* formulated at the level of the networks' details *may* well prove more penetrating than generalizations about the partitions:

> Knowing ... vector-space partitions may suffice for ... accurate short-term prediction ... but that knowledge is inadequate to predict or explain the evolution of those partitions over the course of time and cruel experience. Knowledge of the weights, by contrast, *is* sufficient for this task. This gives substance to the conviction ... that to explain the phenomenon of *conceptual change*, we need to unearth a level of subconceptual combinatorial elements within which different concepts can be articulated, evaluated, and ... modified ... The connection weights provide a level that meets all of these conditions. (p. 178)

His somewhat misleading, atomistic language, notwithstanding, Churchland holds that this level of analysis provides vital resources for understanding cognitive processes that seem critical in accounting for knowledge.

Churchland seems to concede at one point, though, that knowledge of the weights may not be quite sufficient. The partitions are much more useful for 'reckoning the cognitive and behavioral similarities across individuals in the short term' (p. 234). It would seem, then, that explanations of behavior are usually best cast at the level of the partitions: 'people *react* to the world in similar ways not because their underlying weight configurations are closely similar on a synapse-by-synapse comparison, but because their *activation* spaces are similarly partitioned' (p. 234, some emphasis added). Since Churchland asserts that 'the partitions and the functions they serve ... [are] the closest available neural analog of ... "our conceptual framework" ' (p. 234), it appears that a satisfactory account of conceptual change will not minimize those partitions in quite as stark a fashion as the earlier citation seemed to suggest.

The important point, though, is that the dynamics of connectionist systems are not propositional. Construing connectionist networks as the mere 'implementation' of propositional systems (Fodor & Pylyshyn, 1988) is not likely to reveal the rich resources connectionist analyses offer—for reasons Churchland has helped to furnish (see Bechtel & Abrahamsen, 1991).

4. Toward a neurocomputational philosophy of science

For the philosophy of science, where, in contrast to most Anglo–American epistemology, concern with conceptual change has been paramount, the implications of Churchland's proposals are particularly intriguing. In the wake of Kuhn's (1970) work, though, the categories philosophers have used to depict conceptual change in science have remained rather coarse-grained and the analogies inspiring their analyses continue to reflect propositional preoccupations. Philosophers have restricted themselves to talk of 'incommensurability' while construing the phenomena in question in terms of the *translation* of languages or the *interpretation* of (alien) texts. Churchland's neurocomputational perspective promises a very different and more fine-grained picture of theories, explanations and conceptual change.

Kuhn maintained that the theory, as a set of sentences, was not the most useful unit of analysis for understanding science, introducing his subsequently much abused notion of a 'paradigm' in its place. Although Churchland continues to speak of 'theories', he substantially expands the concept in directions Kuhn's discussion anticipates. What he explicates is not the traditional notion of a 'scientific theory', *qua* propositional structure, but rather the range of information processing that would count as 'theoretical' on an account of knowledge that is not absolutely wedded to propositional representation. From such a standpoint the *theoretical* stretches far beyond the theories of science and formal inquiry to all cognitive (though not necessarily conscious) accomplishments that enjoy the same sorts of neural representation and function as these central cases.

How broadly the theoretical ranges on this view turns primarily on our standards for assessing the sameness of neural representation and function. Defining precise standards would be difficult, and Churchland does not try. He clearly regards some standards as too narrow, though. Specifically, his account of perceptual judgments requires that he conceive of the theoretical as sufficiently broad as to remove all hope of theory-neutral observation (see, especially, pp. 255–279). The first step beyond the sensory transducers involves the transformation of activation vectors in accordance with synaptic weights, which, of course, on Churchland's view is in accordance with *prior knowledge* (however it was obtained) (see Churchland, 1993b).

The alternative to propositional representation to which Churchland hearkens most frequently is the psychologists' notion of a prototype (Rosch, 1981). Variously construed in the psychological literature, Churchland conceives of these representational structures as canonical patterns of activations across a neuronal population delineating a special region in the network's activation space. Not only does this view of prototypes preserve Churchland's non-propositional emphasis, it suggests solutions to a host of problems that have arisen both in psychologists' work on prototypes and in the philosophy of science.

This connectionist story explains prototypes' representation and implementation while making sense of, among other things, both the efficiency and flexibility with which they are employed. Churchland offers a general characterization of how the organism settles on the dimensions that define the graded structure in the abstract space the prototype dominates. The attributes the neurally instantiated connectionist system detects are 'those that allow the system to respond to diverse examples … in a distinctive and uniform way … that reduces the error messages … to a minimum'. (p. 124) (See Barsalou, 1993 for a discussion of the problems that connectionist models face in satisfactorily capturing the link between attribute-value pairs in, among other things, the representation of prototypes). More importantly, though, Churchland's proposal offers a basis for explicating naturalistically that ever-elusive notion (for both psychologists and philosophers) of 'similarity'. 'The trained network has succeeded in finding a set of dimensions, an *abstract space*, such that all more-or-less typical *F*s produce a characteristic profile of neuronal activity across those particular dimensions, while deviant or degraded *F*s produce profiles that are variously *close* to that central prototype' (p. 124). The relative similarities of prototypes and of instances prototypes can represent are a straightforward function of the relative distances from one another of the points for the various patterns of activation in question in the abstract representational space the connectionist system defines.

Churchland proposes that most human knowledge (including most practical, scientific, moral and even philosophical knowledge) should be understood in terms of the activation of prototypes within connectionist frameworks. This requires rethinking virtually all of the central concepts in the philosophy of science (in addition to the notion of a 'theory').

Churchland highlights both the diversity of philosophical proposals about explanation and their respective limitations. Faithful naturalist, he focuses on what is involved when we possess *explanatory understanding* rather than on the (various) abstract relations in virtue of which we might come to regard a set of propositions as the *explanation* of some other proposition. Like his treatment of theories, Churchland's approach to explanatory understanding is inclusive. From the standpoint of explanatory understanding explanations in science are not unique. Explanatory understanding (even in science) can just as well arise from perceptual recognition as it can from deductive inference involving causal laws. Explanatory under-standing, categorization and perceptual recognition are 'essentially the same kind of computational achievement' (p. 198). Each involves activating a vector across the system's hidden units, explaining, in the bargain, why, even in the most complex scientific cases, explanatory understanding so often happens all at once. Once biologists recognize those squiggles the microscope discloses as *mitochondrial* DNA, they gain explanatory understanding of why these structures play no central role in meiosis.

This prototype activation model unmasks some of the most off-putting features of reductive explanation for the metaphysically (and informationally) kinder and gentler implications that they are. Churchland estimates that a typical prototype vector contains information on at least 10^8 elements. Construing some phenomenon, then, within the descriptive framework of a lower level theory does not diminish available information but rather substantially enhances it. In short, reductive explanation amplifies our descriptive and explanatory resources. Nor must it be the logically complicated process the logical empiricists had outlined in such excruciating detail. Churchland (1979) had argued that reductive explanation rests on little more than preserving an 'equipotent image' of the higher level theory's explanatory resources (its mechanisms, laws and principles) within those of the lower level theory. Now he has proposed an account of the cognitive mechanism that underlies such recognition.

Learning, conceptual change and some forms of reductive explanation can all be (and often are) instances of what Churchland calls 'conceptual redeployment'. In the philosophy of science the most prominent illustration is when conceptual change appears to be radically discontinuous over time within a particular science (McCauley, 1986a). With its emphasis on small, gradual adjustments of connection strengths, how does Churchland's neurocomputationalist account of prototypes handle such discontinuities?

Churchland notes four available strategies. The first two look directly to the findings of neuroscientific and neurocomputational research concerning (1) the rapid change in both the number and surface area of synaptic connections and the long-term potentiation of neuronal response that can sometimes arise (Desmond & Levy, 1983, 1986), and (2) the substantial changes in the partitions of conceptual space that can occur from rather minor changes in connection strengths.

The third strategy, which Churchland considers but is disinclined to develop, is to downplay these problems by showing the discontinuities in question as far more tractable than first imagined. Thagard (1992) offers historical and computational evidence that indicates this strategy may yet prove more promising than anyone, including Churchland, has supposed.

The fourth strategy, which Churchland prefers, involves the appropriation of an already-developed conceptual framework in a completely new domain. Churchland thinks that such conceptual redeployment will cover the bulk of the conceptual revolutions in the history of science. Huygens' proposal of the wave theory of light is an illustration. Huygens 'had only to apprehend a familiar class of phenomena in a new cognitive context, one supplied largely by himself, in order to have the inputs activate vectors in an area of his conceptual space quite different from the areas they had previously activated. The difference was the context fixers brought to the problem' (p. 237, see also pp. 219–220). Huygens discovers and exploits this analogy in virtue of the fact that input about one domain (light) results (when supplemented by auxiliary inputs from other cognitive centers) in patterns of activation that are sufficiently close to existing prototypes for some already familiar domain (sound) that it produces many of the same effects that activation of the existing prototypes would produce. (For a detailed treatment of closely related materials from a traditional computational approach, see Thagard, 1988.)

Thus, analogical reasoning is at the heart of conceptual redeployment. Although Churchland describes the achievements of the requisite processes, he does not supply an account of how conceptual redeployment is, in fact, accomplished. How does the scientist come to provide the new *cognitive* context? How does the scientist recognize the analogy between it and an already familiar domain? Churchland states that Huygens had to supply this context 'largely by himself', but unless the 'context-indicating elements ... [which] accompany ... inputs' (p. 235) come *exclusively* through recurrent pathways, i.e. unless *none* of the context fixers are a direct function of the input, then Churchland must explain how the net comes to distinguish, within the input, targets from their contexts.

Churchland's prototype model of representation within neural networks is of a piece with his semantic holism in the philosophy of language. His view of prototypes as vectors in neural nets is not only consonant with that position analogically, its dynamics stand in stark contrast with the presumptions of extensional theories of meaning and of causal theories especially. This is so, even though Churchland concedes that the connection between language and the world is richly causal.

The problem does not reside in any skepticism about the pertinent causal relations but rather in the notion of 'reference' that such theories presume (see pp. 284 and 287). Churchland holds that to the extent that language refers at all, it does so in virtue of the relevant terms' positions in a framework whose global excellences undergird its staying power. Finally, though, 'reference is uniquely fixed neither by networks of belief, nor by causal relations, nor by anything else, because there *is no* single and uniform relation that connects each descriptive term to the world in anything like the fashion that common sense supposes' (pp. 276–277). The standard theories' plausibility has rested on untenable assumptions about referential connections between language and the world and in the case of causal theories on historically unjustifiable assumptions about a referential continuity for natural kind terms (such as 'gold')—allegedly sustained by smooth intertheoretic reductions over time. Churchland notes (pp. 283–288) that such smooth reductions are the exceptions and not the rules (even in the case of 'gold') in the long histories of transitions between the relevant theories that stretch from alleged, inaugural naming ceremonies to the present time.

For too long such approaches in the philosophy of language have driven too many philosophers of science to fret over problems of communication between partisans of incommensurable paradigms, however, 'bitheoretics' are no more unusual than bilinguals. For example, as his chapter by chapter replies to the French translation of Kirwan's defence of Stahl's work made clear, Lavoisier was utterly conversant with the chemistry of phlogiston—so much so that his replies persuaded Kirwan to switch his own allegiance to the oxygen theory soon thereafter (Thagard, 1992). Unfortunately, the philosophers' fretting hasn't helped much in addressing what Churchland insists is the important epistemological problem, viz. how individuals rationally evaluate theories without some neutral touchstone.

Surely, from an implacable foe of normative epistemology such insistence would be unexpected! Churchland is not out, or at least is no longer out, to demolish normative epistemology. Instead, he is progressively helping to clarify the shape it must take once we are forbidden neutral touchstones and permitted reference only as a function of the virtues of our conceptual frameworks taken as coherent wholes, i.e. once we recognize the need to naturalize epistemology (McCauley, 1988). Numerous comments throughout his paper on explanation broadcast the fact that Churchland recognizes Putnam's (1983) claim that 'explanation' is a normative notion. The job of the neurocomputational epistemologist is not to eliminate normative concerns but rather to provide scientifically informed illumination and amplification of the epistemological project. So, then, in light of his considerable discussion of the italicized terms in preceding pages, it is no small step when Churchland contends that:

> A virtuous mode of explanatory understanding (that is, an activated prototype vector) should be a *rich* portrait of the general type at issue … strongly *warranted* (that is, have low ambiguity in the input that occasions it) …

> *correct* (relative to the library of currently available alternative prototypes); and ... part of the most *unified* cognitive configuration possible. (p. 223)

Indeed, Churchland shows (pp. 179–181, 221–222 and 1989b) how examining the performance of connectionist networks with varying numbers of hidden units offers *independent* evidence upholding the *epistemic* (not merely the aesthetic) value of the unity and simplicity of explanations. The lesson, in short, is that excess hidden units provide nets with so many computational resources that they never need to formulate responses that will effectively generalize to new cases. They manage the training set with a series of *ad hoc* solutions. This provides *detailed* insight into just what the relation between the unity and simplicity of explanations might amount to in such contexts (as well as, for example, the ties between explanatory unity and multiple constraint satisfaction). The critical points are (1) that such evidence was entirely unanticipated otherwise, and (2) that the scarcity of such detailed discussions in conventional treatments of these topics is notorious. Contra Klein (1992), the issue is no longer whether a normative naturalism is possible. Churchland has already directed one version of that enterprise through some of its initial steps.

His apparent allegiance (p. 151 and 1979) to scientific realism, notwithstanding, Churchland's positions concerning:

(1) the breadth of the theoretical,
(2) the conceptual discontinuities in science (see, for example, pp. 236, 239 and 285), and
(3) the need for a 'more global story of the nature of theoretical justification and rational belief' (p. 255, see also pp. 268 and 273),

leave no more room for classical scientific realism than do those of any other contemporary pragmatist (say Putnam, for example). In recent papers, Churchland is explicit about his pragmatism and less confident about exactly what unreconstructed realism amounts to (see pp. 293–294 and 1993a).

Many champions (such as Putnam) of the *reductio* argument discussed in section 2 subscribe to some or all of the positions that drive Churchland to his pragmatism and therefore, likewise, claim the mantle of pragmatism for themselves. From the standpoints of metaphysical and scientific realists, of causal (and other) theorists of reference, and of correspondence theorists of truth, Churchland's neurocomputationalism is no more self-defeating than is any other version of pragmatism. These philosophical positions imply that *all* forms of pragmatism obliterate the concept of 'truth'.

That charge sticks, however, only on the assumption that these philosophical positions have the corner on what that concept amounts to. Pragmatists, including Churchland and Putnam (and myself), are unwilling to concede that assumption. Churchland maintains that 'the quality of one's knowledge is measured not by any uniform correspondence between internal sentences and external facts, but by the quality of one's continuing performance' (p. 298). As both naturalist and pragmatist Churchland holds that progress in the sciences will continue to aid us in clarifying what to make of 'truth'. Although pragmatic accounts of 'truth' encompass a wider range of considerations than those science occasions, it does not follow that the directions and findings of science should have anything less than a cardinal role in shaping such accounts (McCauley, 1988).

Eschewing metaphysical and scientific realisms and traditional accounts of truth—can natural kinds be far behind? No. If all empirical knowledge possesses an unavoidably theoretical spin propagated through an overarching semantic network, then we have few, if any, assurances that our claims about natural kinds are any more secure than are the theoretical perspectives in which they are embedded (McCauley, 1986b). Natural kinds are those that figure centrally in scientific laws. They are what Churchland calls 'law-bound kinds' (p. 288). But anything that is law-bound is bound even more completely by theoretical presumptions. Without assurance of a final, uniquely privileged, theoretical description of the world, all kinds, even those of our currently most basic physics, are practical kinds.

5. The Diversity of Cross-Scientific Relations or Giving Psychology its Due

Naturalists' differences concerning both the range of scientific findings that should bear on the ongoing reconfiguration of epistemology and the levels of analysis that, initially at least, most merit our attention on this front turn on their preferred models of cross-scientific relations. The Churchlands' models have been richer than most and are still evolving.

Churchland's (1979) *continuum* of intertheoretic commensurability offered a basis for comparing models of theory reduction and scientific revolution (see too Hooker, 1981). Churchland construed the interface between theories in psychology (especially folk psychology) and those in neuroscience as one involving overwhelming discontinuity and, consequently, calling for the outright elimination of the psychology (in the light of its weaknesses). Churchland's gradual surrender of scientific realism, however, has coincided with his relinquishing some of his eliminativist talk.

One of connectionist modelers' favorite collateral activities is to display how various sets of connections in hidden layers amount to possessing approximations of many of our everyday concepts applicable to the domains in question. What are sometimes preserved in these reductions via connectionist networks are recognizable approximations of familiar distinctions.

Churchland now often countenances approximate reductions in terms of specialized connectionist networks' preservation (in their partitions of their activation spaces) of 'equipotent images' of many of our ordinary notions about topics as diverse as sensory qualia (pp. 102–108), the intricacies of culture (pp. 132–134) and many of the categories that populate our everyday theories and explanations. Neurocomputational analyses corroborate some of the patterns our customary concepts capture. Though conceding that 'perhaps the familiar propositional attitudes will be smoothly reduced by the computational structures we find' Churchland emphasizes that whatever their fate—'reduction by something superior or elimination by something superior—the categories of folk psychology remain displace*able* in favor of some more penetrating categorial framework' (1993b). Thus, none of this implies that Churchland is ready to rehabilitate either folk psychology or propositional forms of representation. Still, to the extent that theory reduction yields explanatory understanding, these approximate reductions involving the redeployment of prototypes *enrich* our appreciation of the higher level categories and the phenomena they address. Instead of threatening them, reductions shore up higher level notions—Churchland's talk of their displaceability to the contrary notwithstanding. Crucially, this predicts a *far richer integration* of the pertinent sciences than all of Churchland's eliminativist talk and virtually all of his reductionist talk would seem to anticipate. And well it should. The Churchlands have been slow to acknowledge the pivotal role of *systematic theorizing in the psychological sciences* in the overall project of naturalizing epistemology. Even the model of intertheoretic relations that is emerging from *their own* work suggests that this has been unfortunate (see Churchland & Sejnowski, 1990).

Experimental psychology proffers an array of useful theoretical constructs (e.g. representation via prototypes) that can lend both unity and insight to neurocomputational proposals—as Churchland's own undertaking so plainly illustrates! It is from cognitive psychology that empirically informed explication *of* and support *for* prototypes and network theories of meaning have materialized over the past two decades. Furthermore, even types of experimental psychology (such as research on social cognition) that are unabashedly intentional have suggested, at least in the hands of Stephen Stich (1983, 1990), the same sorts of questions about our common sense psychological framework that the Churchlands wish to raise.

More importantly for present purposes though, the roots of connectionism extend as deeply into the soil of cognitive psychology as into that of neuroscience. Connectionist modeling occurs at a level of analysis considerably higher than that at which most neuroscientific research is carried out. The resemblances of connectionist networks to neural structures notwithstanding, most neuroscientists work at much lower levels of analysis concerned with intercellular and intracellular neuronal mechanisms. In fact, most neuroscientific research is not computational and most computational research is not neuroscientific. (I should note straightaway that Churchland is well aware of this. More than once he explores problems surrounding the neural plausibility of connectionist models.) But the contribution of psychology does not stop there. Much, if not most, of the evidence cited in support of connectionist networks' human plausibility arises from results in experimental psychology. Sejnowski and Rosenberg's (1988) emphasis on NETtalk's manifestation of the spacing effect is an apt illustration (developed at length in McCauley, in progress) as is Bechtel and Abrahamsen's (1991) discussion of the verisimilitude from a developmental standpoint of the Rumelhart and McClelland past tense model (1986).

Higher level theories do more than provide the categories and questions in cross-scientific contexts. Cognitive psychology does not merely generate problems for neuroscience to solve. In the course of pursuing its own agendas, it systematically delineates phenomena which can serve as a body of *independent* evidence to which neurocomputational and neuroscientific models should prove responsive and against which they can be partially assessed. The point, in short, is that it is a philosopher's fiction (which all too many of the Churchlands' writings on these topics abet) that reductive explanation exhausts the epistemic interest of cross-scientific contexts.

These comments are intended to clarify the status of connectionist modeling and of Churchland's proposals in particular *as cross-scientific endeavors*. In that light they should make even more clear the extent of his accomplishments. Not only has he begun to clarify how neurocomputational models can augment standard epistemic notions, he has also offered penetrating proposals that demonstrate how connectionist modeling may forge links between neuroscientific and psychological theories—with an eye toward their *mutual* improvement. In the bargain, Churchland has found further bases for elucidating and promoting views he has propounded for over a decade concerning such topics as the nature of mental representation, the semantics of natural language and the character of explanations and theories.

The substantially expanded view of the theoretical which issues from Churchland's neurocomputational perspective leaves no claim above criticism. To the extent that this doctrine expands the domain of our conjectural knowledge, it is profoundly anti-dogmatic (see pp. 278–279 and McCauley, 1991). It is also seamlessly commensurate with the best values of science and with some of the most valuable features of human association generally—all of this and more from just a bit of brainwork.

Note

[1] I am grateful to Kirby Griffis and Stuart Rachels for numerous stimulating discussions about qualia-based objections to Churchland's version of physicalism.

References

BAKER, L.R. (1987) *Saving Belief: a critique of physicalism* (Princeton, Princeton University Press).

BARSALOU, L. (1992) Frames, concepts and conceptual fields, in: A LEHRER & E. KITTAY (Eds) *Frames, Fields, and Contrasts*, (Hillsdale, NJ, Erlbaum).

BECHTEL, W. (1987) Connectionism and the philosophy of mind: an overview, *Southern Journal of Philosophy*, 26 (Suppl), pp. 17–41.

BECHTEL, W. & ABRAHAMSEN, A. (1991) *Connectionism and the Mind: an introduction to parallel processing in networks* (Oxford, Basil Blackwell).

CHURCHLAND, P.M. (1979) *Scientific Realism and the Plasticity of Mind* (Cambridge, Cambridge University Press).

CHURCHLAND, P.M. (1989a) *A Neurocomputational Perspective: the nature of mind and the structure of science* (Cambridge, MIT Press).

CHURCHLAND, P.M. (1989b) Simplicity: the view from the neuronal level, in: N. RESCHER (Ed.) *Aesthetic Values in Science*

CHURCHLAND, P.M. (1993a) Activation vectors versus propositional attitudes: how the *brain* represents reality, *Philosophy and Phenomenological Research*, in press.

CHURCHLAND, P.M. (1993b) A deeper unity: some Feyerabendian themes in neurocomputational form, in: G. MUNEVAR (Ed.) *Beyond Reason: essays on the philosophy of Paul Feyerabend* (Boston, Kluwer).

CHURCHLAND, P.S. (1986) *Neurophilosophy* (Cambridge, MIT Press).

CHURCHLAND, P.S. & SEJNOWSKI, T.J. (1990) Neural representation and neural computation, in: W. LYCAN (Ed.) *Mind and Cognition: a reader* (Oxford, Basil Blackwell).

DESMOND, N. & LEVY, W. (1983) Synaptic correlates of associative potentiation/depression: an ultrastructural study in the hippocampus, *Brain Research*, 265, pp. 21–30.

Desmond, N. & Levy, W. (1986) Changes in the numerical density of synaptic contacts with long-term potentiation in the hippocampal dentate gyrus, *Journal of Comparative Neurology*, 253, pp. 466–475.

Fodor, J.A. & Pylyshyn, Z.W. (1988) Connectionism and cognitive architecture: a critical analysis, *Cognition*, 28, pp. 3–71.

Hooker, C. (1981) Towards a general theory of reduction, *Dialogue*, 20, pp. 38–59, 201–236 and 496–529.

Jackson, F. (1986) What Mary didn't know, *Journal of Philosophy*, 83, pp. 291–295.

Johnson, M. (1987) *The Body in the Mind* (Chicago, University of Chicago Press).

Klein, E. (1992) Is normative naturalism an oxymoron? A response to McCauley, *Philosophical Psychology*, 5, pp. 289–297.

Kuhn, T. (1970) *The Structure of Scientific Revolutions* (2nd ed) (Chicago, University of Chicago Press).

McCauley, R.N. (1981) Hypothetical identities and ontological economizing: comments on Causey's program for the unity of science, *Philosophy of Science*, 48, pp. 218–227.

McCauley, R.N. (1986a) Intertheoretic relations and the future of psychology, *Philosophy of Science*, 53, pp. 179–199.

McCauley, R.N. (1986b) Truth, epistemic ideals, and the psychology of categorization, in: A. Fine & P. Machamer (Eds) *PSA-1986*, Vol. 1 (East Lansing, Michigan, Philosophy of Science Association).

McCauley, R.N. (1988) Epistemology in an age of cognitive science, *Philosophical Psychology*, 1, pp. 143–152.

McCauley, R.N. (1992) Models of knowing and their relations to our understanding of liberal education, *Metaphilosophy*, 23, pp. 288–309.

McCauley, R.N. (in press) Three ways for psychology and neuroscience to co-evolve, in: R. N. McCauley (Ed.) The Churchlands and Their Critics. (Oxford, Basil Blackwell).

Putnam, H. (1983) *Realism and Reason* (New York, Cambridge University Press).

Putnam, H. (1988) *Representation and Reality* (Cambridge, MIT Press).

Quine, W.V.O. (1990) *The Pursuit of Truth* (Cambridge, Harvard University Press).

Rosch, E. (1981) Prototype classification and logical classification: the two systems, in: E. Scholnick (Ed.) *New Trends in Cognitive Representation: challenges to Piaget's theory*. (Hillsdale, NJ, Erlbaum).

Rumelhart, D.E. & McClelland, J.L. (1986) On learning the past tenses of English verbs, in: J. L. McClelland, D. E. Rumelhart and the PDP Research Group (Eds) *Parallel Distributed Processing*, Vol. 2 (Cambridge, MIT Press).

Sejnowski, T. & Rosenberg, C. (1988) Learning and representation in connectionist models, in: M. Gazzaniga (Ed.) *Perspectives in Memory Research* (Cambridge, MIT Press).

Siegel, H. (1984) Empirical psychology, naturalized epistemology, and first philosophy, *Philosophy of Science*, 51, pp. 517–537.

Stich, S. (1983) *From Folk Psychology to Cognitive Science: the case against belief* (Cambridge, MIT Press).

Stich, S. (1990) *The Fragmentation of Reason* (Cambridge, MIT Press).

Thagard, P. (1988) *Computational Philosophy of Science* (Cambridge, MIT Press).

Thagard, P. (1993) *Conceptual Revolutions* (Princeton, Princeton University Press).