Inquiry, 29, 217–40
(1986)

Symposium: Patricia Smith Churchland's *Neurophilosophy**

A Program for the Neurobiology of Mind

Martin Sereno
California Institute of Technology

Patricia Smith Churchland's *Neurophilosophy* argues that a mind is the same thing as the complex patterns of neural activity in a human brain and, furthermore, that we will be able to find out interesting things about the mind by studying the brain. I basically agree with this stance and my comments are divided into four sections. First, comparisons between human and non-human primate brains are discussed in the context, roughly, of where one should locate higher functions. Second, I examine Churchland's views on reduction and levels of organization, which I find mostly congenial. Third, a key point of disagreement about the relationship and importance of language to specifically human cognition is taken up. I like Churchland's critique of certain sentential paradigms, but I try to show using an analogy with cellular coding systems why we need to get a better theory of 'sentences'. Finally, I discuss how the models introduced in the last chapter might be extended to make better contact with neurobiology and language.

Patricia Smith Churchland's *Neurophilosophy* is an important and audacious book, not the least for its truly interdisciplinary stance. The demands it puts on the reader's philosophical and neuroscientific intuition are substantial and may be disappreciated by those who would keep philosophy and science under separate roofs when it comes to the mind-brain; but such readers should persist since the book makes the most competent case I know for bringing the two back together. Philosophy of science today often delves into science or history of science, but usually in a different way than we see here; this book is better described as philosophy using science, or perhaps as philosophical science. I think this is an entirely legitimate occupation; and I find the basic outlook and detailed arguments congenial enough for it to be easy to localize a few points of disagreement.

Churchland's main contention is that philosophers, psychologists, and neuroscientists should immediately begin to work on building interlevel theories about the mind-brain using both 'top-down' and 'bottom-up' strategies. A common complaint, even from those who think such theories will eventually be possible and interesting, is that in lusting after something

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

so unattainable at present we may be setting ourselves up to accept, in proximal frustration, a sorely inadequate, or worse, an insidiously misleading substitute. Her response is that one cannot know how much theories and models at different levels presently constrain each other without studying them in detail, and anyway, that interlevel negotiations are most exciting before there is an accepted framework; in any case it seems unwise to rule out certain kinds of evidence a priori, at least if the history of philosophy in relation to science is any guide. There is also a social agenda here; now some may think it tasteless and dissipating to mingle too closely with other natives, but clearly this is an efficient way of getting some of their ideas into your head.

A good part of Churchland's case, which I consider first, rests on comparisons between, on the one hand, the structure of the human brain and its normal and damaged function and, on the other, detailed studies of neuroanatomy, neurophysiology, and behavior in mammals (especially non-human primates). In the second part, I discuss her views on reduction, folk science, and levels of organization. Third, a key point of disagreement about the relationship between language and human cognition is discussed, mostly from a comparative stance. Finally, I comment on some of the specific models introduced in chapter 10, especially with regard to articulation with available data.

I

It has been apparent for at least a century that the human mind-brain evolved only recently (by paleontological standards) from non-human precursors. However, humans seem so much more adept at certain things (e.g. language and music) than other mammals (e.g. rats and cats, but also other primates) that one might have expected an obvious reflection of this difference in the brain – human brains should look quite distinct from all other animal brains. In fact, except for their somewhat larger size, human brains closely resemble the brains of apes, which together with other primate brains differ markedly from cat brains, which differ markedly from rodent brains. The similarity between chimpanzees and man is quite striking at the molecular level where the DNA sequence that generates a human only differs by about one or two per cent (single-copy base-pair mismatches) from the sequence that generates a chimp. The implication is that relatively minor 'tweaking' of primate developmental programs may be responsible for the human brain. Given that animals exhibit considerable non-linguistic intelligence, it seems we should be able to learn a lot about the human mind and brain by studying those of non-human primates. The key question in all this, to be taken up in the third section below, is the *degree* to which

the neurophysiological and mental 'patterns' that appear in human neural networks differ from those in other primate brains as a result of the slight readjustments. I nevertheless strongly agree that it would be rash (p. 388) to think that the linguistic ability, for example, so lately arrived, could have somehow sidestepped pre-existing non-linguistic cognitive architectures, or inserted itself in between them either metaphorically or literally in the cortex (as neurologists – especially those studying aphasia – sometimes seem to imply).

After providing in chapter 2 an up-to-date, readable introduction to the biophysics of nerve cells, Churchland presents in chapter 3, an adroit summary of the massive literature concerning the functional architecture and development of vertebrate (mostly cat and monkey) brains that contains remarkably few archaisms. The discussion of topographic maps in the nervous system is better than any advanced textbook treatment available. In this case, I think some of the surprising implications of the monkey–human comparison with respect to the cortex can be a little more explicitly stated, since they support some of her later contentions.

In brief, there had long been a distinction between 'primary' areas in the cortex, through which much of the information from subcortical centers must pass on its way to higher processing centers in 'secondary' areas. Each of the three main sensory modalities represented in the cortex – i.e. vision, audition, and somatosensation – was thought to have its own primary and secondary areas. In animals and humans, it had also long been suggested that lesions of primary visual cortex produce 'sensory' blindness, while lesions of the surrounding secondary areas result in a different sort of 'psychic' blindness where the subject can still see things, but seems unable to recognize or act appropriately toward them. William James's *Principles of Psychology*, chapter 2, provides an admirable account of this. In recent years, as Churchland notes, detailed mapping studies uncovered an unexpectedly large number of distinct, more or less topographic maps (of receptor sheets) in each of the secondary 'psychic' cortices in addition to the better known map in each of the primary cortical areas, bringing the total number of visual, auditory, and somatosensory areas in monkey cortex to somewhere between 30 and 40.

Perhaps as significant is what was not found. Very many philosophical, psychological, and neurobiological theories of the mind assume that there must be some 'place' where sensations from various modalities can come together to generate or interact with more abstract, logical, relational, and anyhow, supramodal representations. The 'central isotropic system' postulated in Fodor's *Modularity of Mind*, the cognitive scientist's 'conceptual system' or 'semantic network', and the neuropsychologist's 'association cortex' are but late versions of an idea with a long philosophical pedigree. The problem is, the new-found visual, auditory, and soma-

tosensory maps discovered in monkeys and cats consumed almost all of the remaining post-central cortex, leaving only diminutive strips of truly polymodal cortex in between.

The significance of this for human neuropsychology may have been greatly underappreciated. Of course, it is not entirely implausible that the 'extra' cortex in humans (relative to monkeys) between the primary sensory areas (which are more nearly monkey-sized) constitutes a large polymodal area for which there is no analogue in all other primates. But I think a more likely hypothesis is that human brains simply have more and possibly larger secondary visual, auditory, and somatosensory areas separated by the same small polymodal strips seen in other primates. This, incidentally, was William James's assessment – he wrote '[t]here is no "center of Speech" in the brain any more than there is a faculty of Speech in the mind'[1] – but it is also supported by non-invasive blood-flow experiments on humans, which show that stimuli in each of the three main modalities activate large, but non-overlapping parts of the post-central cortex;[2] and it is more in line with the differences in areas seen between prosimian primates and monkeys. This is not to deny any interactions between modalities; there are various pathways (e.g. through the limbic system and frontal cortex) whereby stimulus-initiated activity in, for example, somatosensory areas could be transmitted to say, secondary visual areas, and there used to generate a visual representation of some aspect of the somatosensory stimulus.[3] But there is more than a terminological quibble here, since the main character of the neural processing underlying peculiarly human abilities like language comprehension may be determined in large part by cortical areas whose main truck was (and probably still is) in visual, auditory, and somatosensory activity-patterns as opposed to the abstract, modality-free 'representations' more commonly invoked (Wernicke's area, for example, probably consists of several auditory areas). Interestingly enough, the only cortical areas reliably activated by more than one modality so far observed lie in the frontal cortex, which, if other primates are a guide, contains multiple motor maps.

I think these points actually reinforce Churchland's arguments about why it is worth studying animal brains to find out about their human counterpart and they are in line with the general thrust of the pivotal discussion (pp. 450–8) on the possible relation between mental states and sensorimotor control. Higher functions certainly are 'not a sphere unto themselves' (p. 451) and very probably are mediated by neural activation patterns playing across map-networks evolved to (non-linguistically) see, hear, feel, and thus avoid trouble. Higher functions may follow the 'rules' or 'laws' of such networks in the way a cell's proteins conform to the 'laws' of physics and chemistry that we presume also governed pre-biotic soups. The question to which I will return is whether the functioning human brain is as qualitatively

different from its animal precursors as the enzyme controlled metabolic networks in cells are from pre-biotic soups.

To conclude, although I think the possible connections between human and non-human primate cortical areas could have been discussed more, this is a small point and the exposition of neuropsychology and neurology in the succeeding chapters 4 and 5 is uniformly critical, insightful, and intriguing. As before, Churchland compactly surveys a massive and (even more) unruly literature with good judgment. The discussion of right hemisphere language is particularly good. The section on event-related potential research might have mentioned recent evidence from human depth recording suggesting that the P300 wave originates primarily from the hippocampus and amygdala[4] – which means that a large number of psychophysiological experiments may have unwittingly been done on these structures.

II

Part II of *Neurophilosophy* concentrates, after an historical preface, mainly on reduction as a general phenomenon in science, and then, specifically with respect to folk psychology, dualism, functionalism, and neuroscience. I like many of the points made here. In this section, I approach some methodological issues via two questions: (1) why do mental states seem so hard to reduce? and (2) what does reduction look like in practice?

Churchland makes a strong case for the eventual reduction, revision, or in some cases elimination of folk psychological mental states by a matured neuroscience. Most people find this view deeply implausible, even while granting that neuroscience may someday become mature and ecumenical. I think the case for revision and replacement is actually quite convincing and right-minded; but I think there is a possibility that some might misconstrue what I take to be the intended sense of this, as a result of a failure to distinguish reduction (or replacement) involving theories about stuff at approximately the same level of organization (i.e. folk physics/Newtonian physics) from reduction involving theories about stuff at greatly different levels of organization (i.e. folk theory about internal mental states/neurobiology of same) (see esp. Wimsatt).[5] Now surely many folk theories that seemed ineliminable at one time have been eliminated, and we have got over them. Given the complexity of the brain, and the poor record with respect to physics, chemistry, and biology, it would be, in Paul Churchland's words 'a *miracle* if we got *that* one right the very first time'. The implication is that mature neuroscience might to an extent actually transform our private introspection.

In the case of folk physics replaced by Newtonian physics, it is clear that

an experienced mechanical engineer watching a twirled weight-on-a-string fly away from him after the string breaks has mind-brain states that are quite different in some respects from what happens in the head of a 'folk physicist' ostensibly doing, watching, and feeling the same thing. The engineer knows possibly without conscious consideration that the direction of the (centripetal) force he or she was countering while twirling is always approximately orthogonal to the instantaneous direction of the weight, while the 'folk physicist' partly or wholly conflates the two directions, gives the proprioceptively detected pull during twirling precedence, and then often misrepresents to himself the actual tangential trajectory he just saw (presumably, the initial conflation of the two directions arises from experience with stationary suspended weights where the two directions are the same). Thus, an analogous transformation of introspections or even a removal from introspection for that matter could certainly be contemplated with respect to a neurobiologist versus a 'folk psychologist'.

Two key questions are first, what the new neurobiological stuff looks like, and second, whether there are certain aspects of experience only inchoately described by folk psychology that might nevertheless be ineliminable. On the first count, it seems clear that a lower level view of the new stuff could look quite unfamiliar. Personally experiencing the mental states $S(1) \ldots S(n)$ will never be much like experiencing the lower level brain states of somebody else who was experiencing the mental states $S(1) \ldots S(n)$ (presumably via the apparatus and theory of a more mature neuroscience – think of a late model cerebroscope with spatial resolution of a tenth of a millimeter and temporal resolution of a millisecond hooked up to a giant screen color graphics read-out). This is not to say that one would be unable to study one's own states. True, it would be impossible for a person to *simultaneously* experience both the mental states $S(1) \ldots S(n)$ and his or her own lower level brain states constituting $S(1) \ldots S(n)$ via the apparatus (because of the resulting positive feedback); in this restricted sense any given set of one's own brain states are 'occult'. But this is not much of a practical problem; one way to get around it would be simply to set up a recording apparatus for later playback and *then* think the states $S(1) \ldots S(n)$. Afterwards, even the mature neuroscientist would probably have to sit for a very long time in front of the video playing back and analyzing the data to get it into a suitable shape to publish. In the process, the investigator would experience millions more mental states, aggregations or recombinations of which would eventually modify the theoretical framework he or she started with. But, as noted above, something similar happens presently when we become an engineer, read a colleague's insightful paper, or go to a play. Confusion comes from the unfounded fear that the contents of our 1980's mental experience, folksy as it is, could somehow be directly replaced or eliminated by lower level neural activity. As Churchland notes (p. 396), the replacement will be the 'conceptual framework of a matured neuroscience' – i.e. new contents (and obviously, new constituent brain states) but not the elimination of private mental experience. By analogy, the eventual reduction of embryonic development to molecular biology will not make the phenomena of developmental biology disappear. This is not to maintain the indubitability of introspections, but simply to acknowledge that the public–private distinction is likely to remain with us – at least until the day that two brains can interact directly at the single neuron level (*two* brains in a vat perhaps?). Philosophers have long noted that one might have a head start in studying the mind-brain relative to studying, say geology, since we know what it is like to be a mind-brain but not an overturned anticline. Philosophers especially from Kant on began to realize that the advantage might be less than was first supposed. And since brains are rather harder to study than rock, the head start has become in some ways a hindrance.

On the second count, I think there is some preliminary neuropsychological evidence that certain aspects of mental experience described inchoately by folk psychology – but also prized by philosophers – may survive the onslaught of any future neuroscience. One such phenomenon in all seriousness is that when you look at a well-lit scene with your eyes open, everything else held constant, you get a 'bright' visual experience, while if you close your eyes, no matter what you think about or try to visualize, or how hard you think about it, the visual mental experience is much 'darker' (cf. Hume). In neurobiological terms, a 'bright' visual experience roughly requires that your secondary, tertiary, and quaternary visual cortical areas be activated from the input side via primary visual cortex.

Now let us examine a pile of non-standard cases from the literature.[6] For instance, you might have a lesion in part of primary visual cortex. Since there is a fine-grained map of the visual field there that projects powerfully to secondary maps, a stable 'scotoma' will be generated. This would not be visible as a hole in your visual field (cf. the blind spot that everybody has, which subtends an angle about ten times that of the moon) even if it was rather large, and would only gradually be discovered in the course of bumping into things that happened into it. (Traumatic visual cortex wounds from penetrating missiles or surgery do not result in Anton's syndrome of blindness denial [p. 228]; though patients never become phenomenally aware of the defect as a visual 'hole', they know that it is there and learn where it is.) Perhaps the lack of awareness occurs because the higher areas that receive from primary cortex are quite 'used to' things being partially occluded (since this happens ubiquitously in real scenes) and thus 'assume' that something must have got in the way or that something is in the eye when an object falls into the scotoma. Your eyes-open experience is still 'bright'. If all of primary cortex is suddenly lost, for

example in a gunshot wound to occipital cortex without loss of consciousness, the removal generates bright, colored visual phenomena, but seconds later, complete phenomenal 'blindness' is experienced. Months afterwards, residual conscious visual experiences often return, though the ability to orient to objects, or recognize patterns, is often much better than would be predicted by the patient's performance on tests of conscious awareness of the stimuli (i.e. blindsight). In these circumstances, some of the secondary and tertiary areas are being driven weakly from the input side through an alternate pathway (retino-colliculo-pulvino-cortical) but this apparently cannot produce conscious 'brightness'. Patients like these, however, often report preserved visual imagery (i.e. 'dark' conscious visual experiences) as well as visual dreams. If an intact primary area is electrically stimulated with a small current, the patient sees a localized, 'bright' image called a 'phosphene', even if he has lost his eyes. When eyes are undamaged (and open), the phosphene blots out the underlying part of the visual field. If you are having a vivid dream (the kind normal people have at night, not the kind sceptical philosophers have during the day) and have somebody handy to pry open your eyes with the light on, you will admit that the light (which unlike the dream, activates your primary visual cortex) looks much brighter than your dream. If you take some LSD or mescaline and go in a dark room to hallucinate, turning on the light (and probably your primary cortex) will mostly suppress visual hallucinations and the light will look 'bright' relative to them; sometimes, faint moving geometrical patterns scaled with eccentricity will appear that are probably due to periodic waves of endogenously generated activity in your primary area being 'seen' by higher areas. Hallucinations from sleep deprivation behave somewhat similarly. If you accidentally hit the part of your skull overlying visual cortex smartly, you will see 'stars' (phosphenes) and then persisting, moving, zigzag corrugations of the visual field similar to those induced by drugs. Finally, if you look at anything while in a PET scanner (or other device) set up to detect cerebral blood flow, primary (and secondary and tertiary) visual cortex will show increased flow indicating considerable activity; if your eyes are carefully taped shut, just thinking about (visual) things causes no detectable activation in primary and secondary cortex.

Thus, although the 'theory' presented is exceedingly schematic and incomplete (it says nothing about how 'bright' visual experiences are actually generated in cortical networks), and depends on introspective reports, I think it suggests that certain molar distinctions already approximately made by folk psychology (and philosophers) – like 'bright visual experience' versus 'dark internally-generated visual experience or image' – may have reasonably coherent neurophysiological requirements in the awake brain – in this case, lots of neural activity coming into secondary and tertiary visual areas from primary visual cortex for phenomenal brightness. The implication is also that this correlation may be basically unrevisable by learning the conceptual framework of any future neuroscience – i.e. the future neuroscientist will not be able to generate bright visual experiences (and the concomitant massive excitation of striate-extrastriate pathways) endogenously without direct electrical or chemical intervention in his or her cortex. This is perhaps only a minor solace to someone interested in arguing for the constancy of other somewhat more abstract, linguistically-mediated or -initiated aspects of the 'lived-in-world' and of the social structure of science, but I am inclined not to be dogmatic in the face of *any* future science of brains (and scientists). However, I would not be surprised, for instance, if neurobiology and psychology remained within the realm of small laboratory research groups, or if scientists in those fields continued to be little affected by the neurobiology of discovery, rationality, and theory choice in the next century.

The question of what reductive explanation looks like *in practice* is of relatively recent vintage in philosophy of science. Historically, more effort has been directed toward extremely idealized situations and claims often centered exclusively on 'in principle' reduction, deduction, or translation of one theory (often just a sentence) to another. For example, one might have examined the implication of the in principle reduction of the brain to the atomic level in a single stroke via an extra-giant Schrodinger wave equation (while a real molecular neurobiologist might perform a quantum electronic structure calculation for a single neurotransmitter molecule in empty space). In turning toward 'in practice' examples of reduction, we find a much more unruly process with theories and special purpose models at many intermediate levels interacting in complex ways through time. Churchland is especially concerned to describe reduction in real neurobiology, and psychology (and philosophy). Somewhat surprisingly, in light of that, and given her general pessimism about logic and sentences as an understructure for human cognition, she provisionally adopts (p. 294) a modified deductive-nomological model of intertheoretic reduction; her version emphasizes how an old theory is extensively reconfigured in the process of explaining it by (reducing it to) the new theory. To be sure, this is only given as an eminently revisable (or discardable) framework, and examples are not formalized.

Two of Churchland's key points about reductive explanation in neurobiology and psychology, in fact, concern levels of organization – a topic rarely approached explicitly in the context of a D-N model. First, and I think it is a crucial point, there are good reasons for not slavishly carrying over the three-level computation/algorithm/hardware ontology most closely associated with von Neumann computers into psychology and neurobiology, as von Neumann himself was at pains to point out[7] (it is not even clear that this ontology does justice to the underlying architecture of any

but the most primitive computers). The attractiveness of it has always been the apparently loose coupling between the computational/algorithmic levels and the hardware level; for example, different hardwares can run the same program (I say 'apparently' for those who may have actually tried to transport a program). The hope so eloquently expressed in David Marr's book, *Vision*, is that one might be able first to determine certain fundamental hardware-indifferent constraints intrinsic to a particular subtask the brain must solve and then worry about the neural implementation separately. One can see how this could go wrong, however. Say, as chemists and biologists, we were trying to make 'artificial life' with an aim toward understanding the real thing. We might have isolated a subtask necessary for life – e.g. the breakdown of a sugar – and furthermore, have devised a series of special-purpose solid-phase catalysts quite unlike enzymes to do this, arguing that the constraints inherent in sugar bond-making and -breaking were more important to study than the probably irrelevant and arcane multi-leveled structure of actual sugar-metabolizing enzymes (which we shall assume were not well understood). The problem is, we might very well have overlooked important constraints involved in integrating our artificial catalysts into a working cell. For example, a catalyst's bond-breaking specificity might not be high enough to work in close quarters along with all the rest of the artificial cell contents we haven't got around to modeling yet; or it might be impractical for the artificial cell to construct some of the catalysts. This is not necessarily to imply that the only way to run a metabolic network is the way actual cells do it, using enzymes in all their baroque excesses; but there is nothing so good as an existence proof when it comes to something so functionally interlaced as a cell, or as the brain structures and states mediating human cognition. Thus, somewhat paradoxically, the same reason often given for ignoring the architecture of the brain's multi-level 'hardware' may be the best reason to study it.

Second, Churchland rightly emphasizes the *co*evolution of theories at different levels; as often as a lower level theory corrects and informs an upper level theory, an upper level informs and corrects a lower. It may well be as Wimsatt has suggested,[8] that upper to lower corrections are particularly common in the early stages of lower level theory construction when failed lower level models are eliminated wholesale. Certainly the neurobiology of higher functions is in the early stages. But even with more mature lower level theories like those in molecular biology, one finds robust two-way interactions. An example is the recent flurry of excitement and progress in homeotic genes in the fruit fly. In contrast to some other bottom-up attempts on developmental mechanisms in higher organisms, this work built on a substantial base of ongoing classical genetic research into homeotic mutants. Churchland gives some other nice examples of influences going both ways. Though it must be admitted that the connection between

patterns in neural networks and language is presently much more remote than that between genes and cell metabolism, the answer to how neural networks represent may turn out to be at least as interesting and exciting as the slowly unfolding explanations of how complex interaction between biochemicals, but chemicals nonetheless, conspire to make life.

III

A theme running through this book is that logic-based sentence-manipulation may not be the main reason why people are smart. Churchland presents a battery of supporting arguments, most of which I basically agree with. Some of the main points are: (1) folk psychology uses sentences and will probably be revised or replaced, (2) sentential paradigms have trouble with tacit beliefs (difficult to deduce only appropriate ones), (3) and with knowledge access (difficult to determine what is relevant in a given context), (4) pattern recognition and other intelligent behaviors seem un-sentence-like, (5) animals are pretty smart without language and we only recently evolved from them, and (6) language is learned without a sentence-interpreter. But I would rather not accept the conclusion – if it is taken also to refer to human language in general and not just to the somewhat peculiar, stripped down model here called folk psychology that contains only isolated sentences expressing beliefs, desires, and a few other attitudes *that p*. My main arguments for language differ from those Churchland attacks, and are, I think, compatible with the book's stance on many other issues. A comparison with the earlier discovery of the very much simpler 'symbolic-representational' system in living cells helps to illustrate how we might eventually come to understand language naturalistically.[9] At the risk of extreme overcondensation, I shall consider, in turn, the prominence of language, language versus communication, and the architecture and origin of language systems.

Language-related activity, even from a strictly external, behavioristic viewpoint, is an exceedingly prominent and unique characteristic of human primates. While immersed in language, it is easy to forget how peculiar an animated, hour-long conversation consisting of perhaps 30,000 closely connected speech symbol segments in largely non-repetitive sequences must appear to a contemplative non-linguistic primate. The only examples of serial vocal behavior in present-day animals that even remotely resembles this in scale are from songbirds; mockingbirds, for instance, typically sing hundreds of distinct songs, each consisting of a small group of 'syllables' with a few sound segments per syllable. It is easy to tell the human and avian behaviors apart, though, even while retaining an external perspective, since in humans, the ordering of the sequence at intermediate scales

(spanning hundreds to thousands of sound segments) makes a big difference, while the songbird's intermediate range ordering (e.g. as reflected in song order), though not random, is quite obviously of little importance to him or his intended avian audience. That linguistic symbol chains are invariably immense (e.g. the million or so segments arranged into over a thousand paragraphs in the book under review) is perhaps obvious, but is sometimes underplayed in the course of discussions concentrating on the properties of single sentences; it is certainly a rare person (or philosopher) who would be content to say 'I believe a cat is on the mat' and leave it at that. The point is not just that language involves many sentences, but that there is surely a great deal of contextual information in the order in which (real) sentences are heard, read, spoken or written (I suppose one could claim that this is included in the case of 'John believes that *p*' where *p* is a whole paragraph or a whole book, but this does not interestingly engage such information). It is another question how such information might be used, but it seems unfair to dismiss sentence-like, or better, *serially-generated* internal representations without exploring more realistically configured models.

The unique prominence of these immense strings of symbols (and their intended sequences of word meanings) is brought out in considering the various ape-language studies that have attempted to examine the linguistic competence of our nearest primate relatives. It seems quite clear especially from the carefully controlled experiments of Savage-Rumbaugh *et al.*[10] that chimpanzees can acquire quite a number (at least a hundred) of unitary concepts referring to classes of real world objects and actions, as well as the ability to 'activate' one of these concepts internally (bring one to mind even when a real world example of the concept is not present) upon viewing a non-iconic symbol for it that had been previously learned. The concepts include not only concrete categories like 'banana', 'sweet potato', and 'wrench' but also a few somewhat more abstract concepts like 'same', 'on (top of)', 'toy', 'food', and 'tool'. I think the labeling phenomena exhibited in these experiments point to the acquisition of unit concepts rather like the human concepts in the concrete examples (e.g. banana) though perhaps not as rich in the more abstract cases as human concepts (cf. very useful polysemous concepts such as 'line' and 'put'). However, pigeons (and probably many other animals not yet tested) can also rapidly learn to categorize certain classes of natural stimuli (e.g. scenes with trees vs. scenes without trees) while parrots have learned symbol sets that are smaller but otherwise similar to those in the ape experiments.

The most striking finding, however, has so far been negative. After a number of spectacular claims (especially in popular accounts) about 'syntax', and about the various sage and apposite things said by the apes, were gradually winnowed out in more careful studies (e.g. those cited above), there remained little convincing evidence that words were being productively recombined into sequences – even at the two-word level. One problem was selective reportage; for every 'cookie-rock' signed in the presence of a hard, dry sweet roll, there were so many hundreds of less obviously appropriate 'cookie-banana's, 'cookie-dirty's, and 'cookie-cookie's that random word recombination and imitation of the human interlocutor are now thought to explain away almost all the data suggesting production of any symbol *sequence* beyond the single word. Thus, apes (and perhaps parrots!) seem to be able to acquire concepts that are human-like in some respects, but seem entirely unable to 'bond' them together, as it were, to form linguistic sequences. It is intriguing that the millions-of-segments-long symbol sequences (DNA molecules) and the thousands-of-several-hundred-segments-long molecules they code for (proteins) found in cells are as distinct in a pre-biotic milieu (thought to contain no long covalent polymers) as human symbol sequences are in the context of the pre-linguistic substrate presumably examined in the studies cited above.

Questions as to the primary function of language and how much it accounts for or resembles thought have occupied philosophers, linguists, psychologists, and even neurobiologists for a long time. I think a new and enlightening perspective on these issues comes from examining the only other – and a much simpler and much more ancient – example of a self-contained symbolic-representational system, namely the genetic apparatus in every living cell. The first surprise is that cells have no mechanism for turning the three-dimensional information in proteins directly back into coded DNA messages for the purpose of directly communicating with other cells – i.e. no 'speech production'. Rather, each cell 'listens to' and 'comprehends' only its own coded DNA 'speech stream' (cells communicate, of course, but by using a variety of other non-code-based media). The fact that sequence information only goes from DNA to protein and never the other way was quaintly christened the 'Central Dogma of Molecular Biology' by Crick in 1957. Instead of mediating communication, the long code sequences in cells contain information on how to build thousands of special-purpose chemical reaction controlling devices (enzymes) that interact to maintain a complex self-reproducing *metabolic network*. Since all these devices, their chemical substrates, and the code-stuff must all function in close quarters, there is a high premium on specificity of action.

I think that the main purpose of language might be, by analogy, to maintain a stable network of mental 'reactions' (i.e. modifications of neural firing patterns) by the construction of special purpose 'reaction-controlling' devices (i.e. other neural firing patterns) – in other words, a mental metabolism, as it were.[11] From this perspective, the ability to communicate some of these internal reaction-controlling patterns into other people's brains

by turning them *back into code* is an added bonus (with far-reaching consequences to be sure), but something that might be at least conceptually distinct from the 'perceptual' process of generating and maintaining the internal network. Cells, at any rate, communicate by less direct means, each maintaining a common but *private* language. The idea that communication is the *sine qua non* of language has, of course, been challenged before, though perhaps not on these grounds. The linguist Edward Sapir, for example, wrote:

> The primary function of language is generally said to be communication . . . [but] the purely communicative aspect has been exaggerated. It is best to admit that language is primarily a vocal actualization of the tendency to see reality symbolically, that it is precisely this quality which renders it a fit instrument for communication. . .[12]

Possibly, the development of such a 'mental metabolism' has allowed hominids to take control of the highly patterned, but nevertheless, pre-linguistic 'soup' (cf. the pre-biotic 'soup') of mental patterns in their brains in a way that is qualitatively quite different from the way apes or other animals do. The problem now is how to explain this development in terms of relatively minor modifications of the pre-linguistic brain. Churchland also argues (p. 396) that we should treat communication and generalized smartness, especially of humans, separately; our agreement breaks down over how much language has to do with the latter.

The analogy with the much simpler symbolic-representational system in cells (only four 'sounds', twenty 'word meanings', and no language production) may be of some help at this point. In examining the transformation of ideas in the 'forties, 'fifties, and 'sixties surrounding the discovery of the code-stuff and the mechanism by which proteins (e.g. enzymes) are made from it, two areas – the architecture of the system and its relation to pre-existing materials – will be of concern here; since the neural mechanisms underlying human cognition, linguistic or otherwise, are so poorly known at present, the simpler, much more completely understood molecular level system may be able to help us in general questions about basic structural units and their relations. It is somewhat sobering to consider what a mechanistic understanding of the human brain might entail given how complicated things got (i.e. in the history of life) with a twenty-word system.

First, although biologists had consciously manipulated 'genes' since the end of the nineteenth century, discovering in the process that a given gene often had quite specific effects, the way in which this specificity turned out to be expressed at the molecular level was entirely unexpected, and in several ways much simpler than anyone could have guessed.[13] Three major findings were: (1) information in each gene consisted solely of a one-dimensional sequence (genes were long known to be linearly arrayed, but their information-carrying internal structure was thought to be vaguely three-dimensional), (2) proteins had unique, self-assembling, three-dimensional structures, and these folding patterns were determined entirely by their one-dimensional sequence (the composition of proteins was previously thought to be in 'dynamic equilibrium' with their structure and specificity affected by a variety of factors), and (3) the gene sequence determined the protein one in a simple manner with small groups of gene segments (DNA bases) standing for each protein segment (amino acids) (protein synthesis was earlier thought to involve a series of different reactions as in metabolism, instead of a straightforward template process).

In turning to language, the sequence specificity of the 'one-dimensional' code-stuff – i.e. the order of phonemes – has always been apparent. However, when it comes to what happens *after* the sequence of phonemes arrives at the primary auditory cortex as a neural firing pattern sequence, our level of knowledge is nearer to that of the incipient molecular biologists in 1930. A simple proposal based on the architecture of the molecular level system is that the phoneme sequence (cf. DNA sequence) is recognized in small groups (cf. the genetic code) to stand for a simple sequence of word-sized concepts, each thought of as a coherent, neural firing pattern taking perhaps one-quarter of a second to develop (cf. amino acids), and that the resulting chain of concept-patterns, perhaps in higher visual cortical areas, 'folds up' in a determinate fashion (cf. non-adjacent amino acids coming together as the polypeptide of a protein folds) – i.e. concept-patterns that were not adjacent in the original sequence come to interact strongly with each other. The resulting composite 'folded' firing pattern develops an extreme specificity in its interactions with the millions of other latent and active firing-patterns that must coexist in the cortex. Now surely this is too simple a model of language perception and its relation to a 'mental metabolism' of neural firing patterns underlying human cognition. But I present it as a suggestive image of how linguistic symbol segment streams may in part relate in an unexpectedly straightforward and direct way to the internal neural patternings that characterize human cognition. It contrasts in several ways with the non-language-based image Churchland argues for at several places in *Neurophilosophy*. The scenario presented here, however, is quite different from the ones she criticizes, which mostly consist of more or less isolated sentences operated on by a CPU-like logic device; also, it is much less at odds with the sort of theorizing Churchland applauds in chapter 10.

A second area where biology and chemistry turned out to be simpler and more direct than expected has to do with the pre-biotic origin of some of the components of living cells; I think there are important morals here for

thinking about human mental representations. The basic units of proteins – the twenty amino acids – are not particularly complex molecules (containing an average of about nineteen atoms), but before 1953 (the same year the structure of DNA was discovered) their origin was obscure. In that year, it was reported that a number of the biotic amino acids (and some other things) are easily generated by sparking presumed pre-biotic gas mixtures. This was quite surprising since it was expected that the spark would generate an indescribably complex mixture. Instead, it appears as if some of the biotic amino acids can be thought of as naturally occurring 'representations' (or better embodiments) of different categories of reactive chemical collisions that occur in simple pre-biotic contexts (amino acids have even been found on meteorites). The major advance of life, therefore, was apparently not to invent the basic amino acid 'concepts' but rather to find a reliable, standardized way to attach these pre-existing units together to form chains that would then self-assemble, without any further intervention, to form highly specific reaction controlling machines.

The implication for language is that there may be a subset of 'word meanings' (instantiated as a set of neural firing patterns) that are essentially pre-linguistic – i.e. they arise in the course of the primate visual system, for instance, learning to interact with and categorize things, actions, events, directions, places, manners, and so forth in the real world. The issue of how language and non-linguistic or pre-linguistic perception are related has been a durable one – one not likely to be dispatched in a single stroke. Nevertheless, the present analogy suggests that the major advance in the origin of language may not have been to invent the basic word meaning firing patterns; the ape language experiments suggest that chimpanzee brains might already support patterns like these. Rather, the trick was to find a standardized way to 'bond' together these pre-existing units into long chains and then let them 'fold up' without further intervention according to the rules of primate neural networks, to form 'devices' for operating on other patterns in the network.

It is important to emphasize how different the 'self-assembly' model of language comprehension suggested by protein synthesis is from many currently popular models emulating a computer software/hardware analogy (like those criticized by Churchland on pp. 380–99). One way of stating the difference is that the basic level word-sized category representations in the present case interact directly on the basis of their pre-linguistic properties (which arise from constraints in the architecture of the brain and the world) rather than via a set of rules insulated from the pre-linguistic details. For concreteness, one might imagine the pre-existing units as a set of complex, irregularly shaped objects with many protrusions. In the present analogy, the strategy is simply to stick these objects together into chains so that they interact directly with each other, and then pick out the resulting

contraptions that work the best for the task at hand. By contrast, the computer software/hardware analogy suggests that we first attach onto each object a convenient standardized 'handle' that summarizes important properties of the object in a more 'logical' way – say by a pattern of notches – and then set up a system of rules isolated from all the lumpiness that manipulates the meaningless notch patterns in a functionally interesting way. As it turns out, the cell actually uses such standardized 'handles' (i.e. tRNA's) to build up the chains in the first place; but folding subsequently takes place entirely unassisted. The 'handles' would be unsuited for direct use as the elements constituting the self-folding chain for the very reason they make good handles – namely, their standardized, uninteresting structure.

The hope that one might be able to stick with the 'notch patterns' has sprung up repeatedly in twentieth-century philosophy and psychology, starting with the early attempts of the logical atomists to replace messy real world semantics with neater logical syntax. I think some of the more recent debates about 'methodological solipsism', the 'formality constraint', 'cognitive penetrability', and the 'syntactic theory of mind' have skirted around the issue here – roughly, whether it is better to use standardized 'handles' or the real 'lumpy' *but also internal* things they stand for. So far as I understand the discussions, there seems to be an instinctive craving for the 'standardized handles' approach. In cells, at least, it is clear that *both* types of internal unit – standardized and lumpy – are essential. What is most remarkable about cells from a psychological viewpoint is that the 'lumpy', pre-biotic amino acids should somehow be well-suited to do something – i.e. fold up determinately and derive a high specificity of action when put into chains – they clearly weren't 'designed' to do. Perhaps when the internal economy of human linguistic patterns is eventually found out, it will prompt us to ask in a similarly incredulous tone – why should what are basically animal concept activation patterns be any good for building up highly specific human linguistic 'devices' in the brain? And perhaps the only answer beyond a catalogue of their most fortunate properties will be – as one would answer for the molecular level – because they were there before the system existed.

To conclude, I disagree with Churchland's claim that sentences are not the reason people are smart (remembering that the 'sentences' I refer to are longer, and perhaps better thought of as short discourses containing from several to fifteen to twenty sentences; also, I imagine no logical central processing unit). The analogy with the (only other existing) symbolic-representational system in cells was used to argue that it is at least plausible that there might be a fairly direct mapping between speech streams and the cortical activation patterns that might explain why people are smart (and how people might build and maintain a 'mental metabolism', quite in

contrast to pre-linguistic animals). However, I think the position taken here is very much in the spirit of Churchland's general program for a neurobiology of mind; I would say, hominids became smart when they got control of the complex pre-linguistic 'soup' of neural firing patterns in their brains by exploiting the unique properties of 'devices' built up by the simple concatenation of a type of pre-existing pattern that might best be called an 'animal concept'. The complete story of how the human brain and language works couldn't possibly be restricted just to this, but I think the stripped down, better understood cellular system provides a refreshingly concrete starting point for understanding our distinctive minds.

IV

Churchland presents in chapter 10, a vision of what an interlevel theory of brain function might eventually look like by way of three current attempts in this direction. What I would like to do here is critically look at the articulation of theory and data, and then very briefly consider prospects for modeling language functions. Theory in the neurosciences has a curious history. Though mathematically sophisticated theorizing in neurobiology has been around for many years, it has always been rather isolated from neurobiology as a whole. But more than that, as Churchland points out, it is often considered to be premature and somewhat disreputable. In physics and chemistry, communities of experimentalists and theorists seem to interact much more productively than in neuroscience; though experimentalists and theorists in those fields may not particularly like each other, any more than their counterparts in neurobiology do, there is much more intimate feedback. Given that the brain is considerably more complex than, say, a laser or a reactive chemical collision, and that sophisticated mathematical apparatus is required for any interesting explanation of even those things, the attitude toward theory in neurobiology is perhaps to be expected – but not condoned. Interactions between mathematized theory and experiment are much more pronounced in ecology and evolutionary biology, for example, in spite of the great complexity – seemingly more on the order of brains than stimulated emission or colliding molecules – of the phenomena involved. It appears in any case that the generally anti-theoretical stance of contemporary neuroscience is slowly shifting, at least in part, because interest in more brain-like parallel computational architectures has picked up outside neurobiology as absolute constraints on von Neumann devices (e.g. signal propagation delays) began to be more keenly felt. The theory of brains, however, will not be easy, especially if it is to meaningfully engage what is already known about the layout and functioning of brains. My criticisms below are emphatically not anti-theoretical,

and reflect only an impatience for even better, and more testable theories.

The overall aim of the tensorial network theory of Pellionisz and Llinas treated first by Churchland is to consider sensorimotor phenomena – from sensory input patterns to motor output patterns – as vectors expressed in various reference frames; since with respect to a particular goal, these different expressions represent the same physical entity, Pellionisz and Llinas claim that brain networks must therefore be embodiments of tensors – i.e. mathematical objects that remain invariant as they are represented in different (e.g. sensory and motor) coordinate systems. Tensor mathematics has applications in many areas of physics ranging from general relativity and electromagnetic theory to the physics of anisotropic solids – sedimentary rocks, for example. The central nervous system hyperspace in which the vector transformations take place has, in contrast to the spaces used in general relativity or for rock, very many dimensions, at least if neurons or groups of neurons specify separate dimensions. In addition, the space is almost certainly non-Euclidean (unlike the space for rock, but like the Riemannian manifold in general relativity), and the frames in it are probably non-rectilinear, non-orthogonal, and worst of all, non-linear.[14]

In the interest of clarity, however, Pellionisz and Llinas concentrate on examples in Euclidean two-dimensional space, using sensory and motor reference frames with two or three non-orthogonal coordinates and cerebellar networks with a few Purkinje cells (see also the vestibulo-ocular models). The question is how these abstract, simplified models might be extended to make contact with real neural networks (or real experiments on them) where the dimensionality is typically in the millions. The problem is that Pellionisz and Llinas's examples and simulations (e.g. a hand drawing 'OK' on a 2-D surface) embody a direct connection between the three- to six-dimensional (i.e. three to six neurons) internal spaces and *summable* external components in output space (e.g. of the three joint arm). In the real cerebellum, by contrast, mossy fibers and Purkinje cells contain, as groups, very high-dimensional signals that are not obviously interpretable as summable physical components in motor output space, not the least for their extremely high redundancy relative to this space. This does not mean that explicit connections between vectors in extremely high-dimensional neuron-firing rate spaces, in lower-dimensional muscle contraction force spaces, and in Euclidean point-on-a-limb spaces couldn't be made (the nervous system obviously does it), but the vagueness on these issues makes it difficult to test these abstract models on a real cerebellum or relate them to presently existing data.

Obviously, a theory should not attempt to explain all apparently relevant data since some of the data is usually wrong, misleading, or not actually relevant; but there are several recent findings about the physiology of the cerebellum that seem to suggest substantial revisions of standard con-

ceptions of how it works, and which very likely will be relevant to mechanistic models of cerebellar activity.[15] Part of this work was stimulated by the unexpected discovery of intricate, exceedingly fine-grained maps of the body surface (cutaneous receptors, not muscle receptors) in the granule cell layer of several parts of the cerebellum of the rat. Like cortical somatosensory maps, these maps are locally somatotopic, but unlike the cortical maps, the cerebellar maps are 'fractured' into many small, internally highly ordered 'patches' (many well under a square millimeter in area), and there are multiple representations of many parts of the body surface (other parts of the cerebellum receive muscle and tendon receptor, vestibular, auditory, and visual inputs, but the topographic organizations of these other inputs has not yet been studied physiologically at such a fine grain). Granule cell axons travel up into the molecular layer and bifurcate to form parallel fibers, which contact a 'beam' of many Purkinje cells. This divergent, one-to-many connectivity is built into many cerebellar models, including that of Pellionisz and Llinas (as the matrix that transforms 'covariant intention' into 'contravariant execution' vectors) and activated 'beams' were produced in early experiments in which the parallel fibers were directly activated by a surface electrode. However, when a part of one of the granule cell patches is activated by natural peripheral stimulation of the skin, only the Purkinje cells directly overlying the patch are excited (apparently by multiple synapses onto Purkinje cells from the ascending part of granule cell axons); those Purkinje cells 'down-beam' are actually inhibited. Thus, like other cortical structures, the cerebellum seems to show a strong vertical organization, despite the existence of parallel fibers, which must have other, more subtle effects. The presence of 'fractured' body surface sensory maps in the Purkinje cell layer implies that the granule cell–Purkinje cell excitatory connectivity matrix is approximately the identity matrix (off-diagonal numbers are small) in contrast to the matrices used in most cerebellar models.

A second recent finding is that if the deep cerebellar nuclei – through which most cerebellar output must pass – are microstimulated, a wide variety of different, well defined, discrete and synergistic movements of many different body parts are generated, depending on where the stimulating electrode tip is located. Since somatosensory input reaches the cerebellum before it gets to the cortex, and since the cerebellum can generate movements even after a motor cortex lesion, the cerebellum is in some respects like a self-contained sensorimotor transformer – cf. the superior colliculus – with a sensory map 'overlying' a motor map, as much as it is the sort of 'add-on' unit postulated by Pellionisz and Llinas to modify what are already motor-like intentions into more accurate executions.

Clearly, the process of mapping, remapping, and combining fundamentally different sorts of information is vital to the coordinated function

of the several hundred different map-like and non-map-like structures (nuclei, areas) that make up a vertebrate brain. The question is how to represent firing patterns in *maps* in a comprehensible way (which as we saw are prominent even in the cerebellum). The problem with simply writing out a two-dimensional, map-like neural firing pattern as a many-dimensional vector is that the two-dimensional – or in the case of a laminated cortex, three-dimensional – location of each neuron is only opaquely represented across the thousands or millions of dimensions of the firing rate space. Churchland's exposition of the 'phase-space sandwich' approach (pp. 441–5) illustrates one way of constructing a more intuitive representation of activity in map-like structures. Basically, the technique is to move one level up and condense the immense sensory firing rate vector (which one would get from interrogating each neuron in a sensory map like the one in the upper layers of the colliculus) to an activated point – i.e. a vector – in the two-dimensional space of the sensory map itself (Churchland doesn't do this explicitly because of the novel proprioceptive method used in the model to detect target location, but such a condensation would be required to represent retinal input to the real colliculus, or somatosensory input to the cerebellum). Then, this two-dimensional vector can be used as an input to a motor map (like the one in the deep layers of the colliculus). As with the sensory map, a many-dimensional motor map firing rate vector is condensed to an activated point (vector) in a two-dimensional space. To recover an explicit motor output in terms of a pair of *firing frequency* coordinates (e.g. to specify a graded muscle contraction) from an activated point whose information is carried as a pair of *location* coordinates, one also needs a 'spatial-to-temporal' transformation (this is implicit in Churchland's motor output mechanism). The way I have reconstructed the phase-space sandwich approach here explicitly indicates how higher level representations (like a vector in 2-D sensory phase space) relate to lower level single neuron dynamics (many-dimensional sensory neuron firing rate space) – a necessity when trying to relate a model to lower or upper level data from a real neural network. It may sometimes be necessary to retain a lower level perspective, especially when trying to describe the effects of presumably non-linear local circuit interactions; but I suspect we will eventually need multi-level models like the one just discussed, where lower level complexity is 'summarized' at a higher level, particularly for heuristic purposes.

Before leaving tensor network theory, I find it difficult to avoid commenting on Pellionisz's 'tensorial blueprint' for the amphibian brain reproduced as Figure 10.13 (p. 441). I agree with Churchland (p. 408) that circuit diagrams are not 'theories of the brain', but also that they are 'essential' in making such theories. Amphibians (and reptiles) do not have a corticospinal (or corticopontocerebellar) pathway as implied by this

diagram. Cortical output is instead relayed through the basal forebrain and thence to the tectum; there it and other information (e.g. retinal) gets to motoneurons mainly via the tectoreticular pathways, which have inexplicably been omitted from this diagram. My several publications on tectoreticular neurons, however, may have predisposed me to overestimate the significance of this error.

Churchland also gives a nice discussion of several connectionist and parallel distributed processing models. Like tensor network models, these models can be thought of as devices that take an input vector and then transform it via a connectivity matrix and some dynamical assumptions (that produce gradient descent, for example) into an output vector – the reference frame is the same for the input and output vectors, but the transformations are more complex. Many of these models can be thought of as performing a classification of the input vector – i.e. a number of different input vectors (e.g. noisy input, inputs with missing parts) will all generate the same output vector. Often, 'hidden units' are used that do not directly see the input, but become active only after some of the other 'sensory units' to which they are connected are 'clamped' by the input during an epoch of relaxation into a minimum energy state; the clamping has the effect of changing the energy landscape (otherwise defined by the connectivity of the units) across which the network tries to minimize its energy. One of Churchland's criticisms is that these models seem too slow; but the classifications they perform are inherently more complex than the hard-wired one-to-one vector transformations described earlier, and such tasks do have a longer turnaround time in real people than the lightning fast sensorimotor integration required to play a musical instrument (or locomote gracefully through trees). Another criticism is that these models ignore the sensorimotor interface. On this, I quite agree; it is often forgotten that *every* visual cortical area, for example, has direct projections into one or more motor system structures (in this case, to the frontal eye fields, basal ganglia, superior colliculus, pontine nuclei). These models are difficult to relate to real neuronal network data, because unrealistic, task-specific information is usually consciously introduced into the connectivity matrix or used to interpret the output (see, e.g., the Necker cube simulation, or networks in which single 'neural-like units' are assigned abstract interpretations such as letters, word letter sequences, syntactic categories, or concepts). This strategy, however, allows application to a wider range of phenomena, and generates network behaviors that are interesting in their own right. There is a substantial gap to be bridged (from either side) before these models will begin to interact more with neurophysiology.

A problem in extending these models beyond recognition and categorization tasks – appreciated by both practitioners (Hinton) and critics (Pylyshyn) – is that it is difficult to get them to produce the *sequences* of

category-like states that everyone assumes must be involved in manifestly serial tasks like language comprehension and other cases of 'symbol manipulation'. But we need more than just sequences of isolated states; we need to find out how to 'bond' category-like firing pattern states together into long 'chains' that are *themselves* capable of spontaneously relaxing into low energy states (analogous to the manner in which proteins intricately but spontaneously fold up after they are serially assembled and exposed to water). More concretely, we might look for a 'dumb', local way by which a pair of state space vectors could be modified with respect to an axis between them; the previously 'bonded' vectors in the chain would have to persist in some way so that a composite pattern could be built up. In understanding this paragraph, for example, the comprehender not only has to recognize each word sequentially and activate its associated meaning at some level, but he or she must also construct, over tens to hundreds of seconds, a transiently existing 'representation' of the discourse meaning in working memory capable of interacting specifically with the myriad other neural firing patterns that must be simultaneously present in the reader's brain in a latent or active state.

The picture argued for here is no more and no less than an inspiration for how to make an interesting model; whether a scheme along these lines could be realized in a model network is at least an empirical question. Looking for such a higher order device in real neural networks, much less explaining how it might interact with other active and latent devices and substrates of all sorts, is far beyond our abilities at present; we barely have a metaphor for what the real thing must look like. But models like those described in this book may eventually move from metaphor to more concrete links, until one day we shall perhaps be able to catch a full-field glimpse of the strange jungle of patterned neural activity that we casually experience from the inside every day.

V

I hope it is clear that despite some disagreement about the place of language in human cognition, I found this a very interesting book. There is a wealth of information here, even for those who may dislike some of the arguments. This book deserves to be closely read by philosophers, psychologists, and neurobiologists.

NOTES

1 William James, *Principles of Psychology*. 2 vols. (New York: Dover, 1950 [orig. ed. 1890]), I, p. 56.

2 N. A. Lassen and P. E. Roland, in A. Kertesz (ed.), *Localization in Neuropsychology* (New York: Academic, 1983), pp. 141–52; and P. Fox *et al.* (unpublished PET studies and personal communication).

3 The recent experiment of Haenny, Maunsell, and Schiller suggests this interpretation – 'State Dependent Activity in Monkey Visual Cortex; Visual and Non-visual Features in V4' (in press).
4 E. Halgren, J. Engel, C. L. Wilson, R. D. Walter, N. J. Squires, and P. H. Crandal, in W. Siefert (ed.), *Neurobiology of the Hippocampus* (London: Academic, 1983), pp. 529–72.
5 See esp. William Wimsatt, 'Reduction, Levels of Organization, and the Mind–Body Problem', in G. G. Globus, G. Maxwell, and I. Savodnick (eds.), *Consciousness and the Brain* (New York: Plenum, 1976), pp. 199–267 and 'Reductive Explanation', in R. S. Cohen, A. C. Michalos and J. van Evra (eds.), *PSA 1974* (Dordrecht: Reidel, 1976), pp. 671–710; and T. Nickles, 'Two Concepts of Intertheoretic Reduction', *Journal of Philosophy* 70 (1973), pp. 181–201.
6 See, e.g., H. L. Teuber, W. S. Battersby, and M. B. Bender, *Visual Field Defects after Penetrating Missile Wounds of the Brain* (Cambridge, Mass.: Harvard University Press, 1960), W. Penfield and L. Roberts, *Speech and Brain Mechanisms* (Princeton: Princeton University Press, 1959); H. Kluver, *Mescal and the Mechanisms of Hallucinations* (Chicago: University of Chicago Press, 1966); G. S. Brindley *et al.*, *J. Physiol. Lond.* 225 (1973), pp. P57–58; H. R. Rodman *et al.*, 'Removal of Striate Cortex Does not Abolish Responsiveness of Neurons in Visual Area MT of the Macaque', *Neurosci. Abstr.* 11 (1985), p. 1246.
7 John von Neumann, *The Computer and the Brain* (New Haven: Yale University Press, 1958).
8 Wimsatt, op. cit., p. 231 (see note 5).
9 An extended consideration of these issues is found in Martin Sereno, *'DNA' and Language: The Nature of the Symbolic-Representational System in Cellular Protein Synthesis and Human Language Comprehension*, Ph.D. Dissertation, Committee on the Conceptual Foundations of Science, University of Chicago, 1984.
10 E. S. Savage-Rumbaugh, J. L. Pate, L. Lawson, T. S. Smith, and S. Rosenbaum, 'Can a Chimpanzee Make a Statement?', *J. Exp. Psych.: General* 112 (1983), pp. 457–92; see also L. Weiskrantz (ed.), 'Animal Intelligence', *Phil. Trans. R. Soc. Lond. B*, 308 (1985), pp. 1–216.
11 Lest this be thought a reversion to the 'psychological chemistry' presented in J. S. Mill's *A System of Logic*, vol. II, bk. IV (London: Longmans, Green, & Co., 1843), let me hasten to point out that here we consider biological macromolecules and coding systems, while Mill who knew nothing of these stuck mostly to vastly simpler inorganic salts!
12 E. Sapir, *Language* (New York: Harcourt, Brace, & Jovanovich, 1921), p. 159.
13 See H. F. Judson, *The Eighth Day of Creation* (New York: Simon & Schuster, 1979) for a history of these developments and esp. his pp. 605–16.
14 M. A. Arbib and S. Amari, 'Sensorimotor Transformations in the Brain (With a Critique of the Tensor Theory of the Cerebellum)', *J. Theor. Biol.* 112 (1985), pp. 123–55.
15 See G. M. Shambes, J. M. Gibson, and W. Welker, 'Fractured Somatotopy in Granule Cell Tactile Areas of Rat Cerebellar Hemispheres Revealed by Micromapping', *Brain Behav. Evol.* 15 (1978), pp. 94–140: J. M. Bower and D. C. Woolston, 'Congruence of Spatial Organization of Tactile Projections to Granule Cell and Purkinje Cell Layers of Cerebellar Hemispheres of Albino Rat: Vertical Organization of Cerebellar Cortex', *J. Neurophysiol.* 49 (1983), pp. 745–66; L. Rispal-Padel, F. Cicirata, and C. Pons, 'Cerebellar Nuclear Topography of Simple and Synergistic Movements in the Alert Baboon (*Papio papio*)', *Exp. Brain Res.* 47 (1982), pp. 365–80; W. Schultz, E. B. Montgomery, and R. Marini, 'Proximal Limb Movements in Response to Microstimulation of Primate Dentate and Interpositus Nucleus Mediated by Brainstem Structures', *Brain* 102 (1979), pp. 127–46; M. Ito, *The Cerebellum and Neural Control* (New York: Raven, 1984).

Received 26 February 1986

Martin Sereno, Division of Biology 216–76, California Institute of Technology, Pasadena, California 91125, U.S.A.