

Brain Metaphor and Brain Theory

John G. Daugman

Then do you now model the form of a multitudinous, many-headed monster, having a ring of heads of all manner of beasts, tame and wild, which he is able to generate and metamorphose at will?

Plato, *The Republic*, Book IX

In its ideas about itself . . . as in all of its other endeavors, the mind goes from mastery to enslavement. By an irresistible movement, which imitates the attraction death exercises over life, thought again and again uses the instruments of its own freedom to bind itself in chains.

Roberto Mangabeira Unger, *Knowledge and Politics*

1 The Rise and Fall of Metaphors

Contemporary philosophers of science have emphasized the centrality of metaphor both in normal scientific discourse and in the transformative processes of scientific revolution.¹ The life cycle of the dominant metaphor in a scientific theory demarcates the life cycle of the scientific paradigm itself, and the adoption of new metaphors is said to be the signature of transitions between paradigms. More generally, it is believed that reasoning through metaphor is a pervasive and inescapable habit of thought itself (see, for example, MacCormac, 1985), both in the ordinary language of everyday life as well as in scientific and philosophical reasoning. In *Philosophy and the Mirror of Nature* (1979), Rorty claims that in general it is pictures rather than propositions, metaphors rather than statements, that determine most of our philosophical convictions. And perhaps nowhere else in the history of ideas has there been a more striking pattern of reliance on metaphors than in the history of reflection about the brain and the causes of behavior, and about the enigmatic relationship among brain, mental life, and personhood.

Invariably the explanatory metaphors of a given era incorporate the devices and the spectacles of the day, and, in perhaps subtler ways, they may reflect the preva-

lent social forms and daily texture of life. Theorizing about brain and mind has been especially susceptible to sporadic reformulation in terms of the technological experience of the day. For example, the water technology of antiquity (fountains, pumps, water clocks) underlies (see Vartanian 1973) the Greek pneumatic concept of the soul (*pneuma*) and the Roman physician Galen's theory of the four humours; the clockwork mechanisms proliferating during the Enlightenment are ticking with seminal influence inside La Mettrie's *L'Homme machine* (1748); Victorian pressurized steam engines and hydraulic machines are churning underneath Freud's hydraulic construction of the unconscious and its libidinal economy; the arrival of the telegraph network provided Helmholtz with his basic neural metaphor, as did reverberating relay circuits and solenoids for Hebb's theory of memory; and so on. In this historical framework, surely it would be folly for us to regard the recent computer bewitchment of theoretical work in psychology and neuroscience as an *entirely different* kind of breakthrough in the history of ideas. More than folly, it would constitute a "Whiggism": the tendency to view the past history of our subject as the stumbling progression toward its inevitable culmination in today's understanding that the brain turns out to be a computer.

Yet there are many among the forward ranks of theorists today who ask precisely that we not think of computation as just the contemporary metaphor, but instead that we adopt it as the literal description of brain function. Thus, for example, Zenon Pylyshyn complains that "there has been a reluctance to take computation as a *literal* description of mental activity, as opposed to being a mere heuristic metaphor . . . there is no reason why computation ought to be treated merely as a metaphor for cognition, as opposed to a hypothesis about the literal nature of cognition" (1980, p. 111). But, however meritorious the currently fashionable metaphor may be, it behooves us to regard it as the latest in a long sequence of compelling metaphors that, like the technologies from which *they* sprang, also had their day. Not that through such awareness we could somehow transcend metaphor in favor of some higher or more direct language. Rather, the salutary effects of considering the history of brain metaphors may be twofold. (1) It makes us aware of the origins, genealogy, possible hidden agendas, and latent content of our present metaphors. (Or, as King Lear observes, "'Tis a wise man who can say who his father is." This might be called the Problem of Progenitive Epistemology.) And (2), it alerts us to the markings of paradigm shifts when the old metaphors prove inadequate or transparent and novel metaphors take their place. Historians of science such as Thomas Kuhn identify these changes in the metaphorical language as being characteristic of radical and revolutionary science, the brief punctuated episodes of real intellectual vitality, separating the long intervals of stable and sluggish "normal" science. The enlivening effect of a new metaphor and the deadening effect of embracing one too literally or too ideologically or too long are eloquently (and again, metaphorically) described by the Brazilian philosopher Roberto Mangabeira Unger (1975): "in its ideas about itself and about society, as in all its other endeavors, the mind goes from mastery to enslavement. By an irresistible movement, which imitates the attraction death exercises over life, thought again and again uses the instruments of its own

freedom to bind itself in chains. But whenever the mind breaks its chains, the liberty it wins is greater than the one it had lost." In this spirit, we might review the history of metaphors for brain and mind not with the goal of exorcizing metaphor from scientific discourse, which may be neither desirable nor possible, but rather to recognize our enclosure within metaphors and to identify their risings and settings as transitions that are the harbingers of insight and intellectual vitality.

2 A Taxonomy of Mind/Brain Metaphors

One of the edifying if uncomfortable lessons to be learned from the history of brain metaphors presents itself whenever we feel astonished by how inappropriate the categories and metaphors of past eras seem to be: when we marvel at the possibility that anyone could ever have thought *that*. At such times we must try to imagine the amused reception that our own are likely to receive from the intellectual historians of the future.

Something like this edifying lesson was presumably intended by Michel Foucault when he began *Les Mots et les choses* (trans. *The Order of Things: An Archaeology of the Human Sciences*) by citing Jorge Luis Borges's discovery of a tenth-century Chinese Encyclopedia,² in which the following classification of the world's animals was given: (1) Those Belonging to the Emperor; (2) Embalmed; (3) Tame; (4) Suckling Pigs; (5) Sirens; (6) Fabulous; (7) Stray Dogs; (8) Included in the Present Classification; (9) Frenzied; (10) Innumerable; (11) Drawn with a Very Fine Camelhair Brush; (12) *Et Cetera*; (13) Having Just Broken the Water Pitcher; and (14) That From a Long Way Off Look Like Flies (originally from Borges, *El Libro de Los Seres Imaginarios*).

Foucault proceeds: "In the wonderment of this taxonomy, the thing we apprehend in one great leap, the thing that, by means of the fable, is demonstrated as the exotic charm of another system of thought, is the limitation of our own. . . ."

3 Embodied Spirits and Helmsmen

It might be said that a cornerstone of Western thought and social institutions is the notion that persons are embodied spirits. Many of the antinomies and dichotomies of Judeo-Christian culture spring from this idea that we have on the one hand, bodies, and on the other hand, minds. These antinomies include determinism and accountability, cause and volition, the temporal and the eternal, the descriptive and the normative, immanence and transcendence, and the profane and the sacred. As Descartes asked, what then is the relationship, both moral and mechanical, between the body and the spirit that is embodied by it? Our art and literature are filled with images of spirit or consciousness being infused into an earthen body: recall Michelangelo's Sistine fresco of Adam, made from clay, being imbued

with spirit, or the Greek legend of the sculptor Pygmalion whose ivory statue is infused with life, or the variations in transmutation myths in which consciousness is conferred through the eyes or by touch (as in Pinocchio), by breath (as in Genesis), by heat (as for the alchemists), or by electricity (as for Frankenstein). In this same tradition, today, near my house in Cambridge there is a company named "Thinking Machines Corporation," which has apparently managed to confer thought through parallelism.

Michael Arbib points out in his book *The Metaphorical Brain* (1972) that much of modern thought about brain and behavior can be distilled to two fundamental metaphors: (1) the cybernetic metaphor ("Humans are machines"), and (2) the evolutionary metaphor ("Humans are animals"). The very term *cybernetics*, coined by Norbert Wiener to invoke the concepts of control and communication, reflects the tradition of embodied spirits, since the *cyber* (Gk. *kybernētēs*) is the helmsman controlling the ship. This concept of *control* is absolutely central to our reflection about the relationship between the mental and the physical. It is expressed as not only a mechanical problem but also a moral problem: *self-control* is the great Judeo-Christian problem, with ever-recurrent incarnations throughout our culture ranging from the moral allegories of the brothers Grimm, to Freud's psychodynamic conflict, to the anti-utopian solutions of Skinner's *Walden Two* or Huxley's *Brave New World*. Our social sciences of psychology and economics similarly conceive of personhood largely through expressions of control and choice: person as puppet, role-player, poker-player, script-follower, bundle of desires, optimizer. Very broadly, whether within the Judeo-Christian framework or the cybernetic one or the social scientific one, the recurrent and fundamental metaphor for personhood and consciousness in relation to brain, body, and action has been that of a ship's helmsman.

4 Hydraulic and Mechanical Metaphors

Hydraulic and mechanical concepts of the psyche began in pre-Socratic thought³ (figure 2.1). In the fifth century BC, the School of Hippocrates advocated an early hydraulic model of mind based on the four humors, whose preponderance or imbalance could be related to mental dispositions. An excess of phlegm resulted in apathy; an excess of black bile led to melancholy; an excess of yellow bile produced biliousness; while an excess of blood yielded sanguinity. This basic humoral model evolved in the second century BC into the Roman physician Galen's theory of "animal spirits," a highly rarefied fluid that flowed inside the nerves. Through these animal spirits, the brain received sensory messages and sent back commands. Thus began the pneumatic/hydraulic concept of the mind, a concept whose many descendants include Victorian "hydraulic" theories of the psychodynamic libidinal forces of desire and repression (the libidinal economy).

Probably the first modern exposition of the mechanical metaphor for the brain was Descartes's contention that *animals*, as distinct from men, were pure automata



Figure 2.1 An early triumph of the mechanical metaphor for cognition. Baron Wolfgang von Kempelen, widely celebrated for his mechanical genius and many notable inventions, while holding the post of Aulic Counselor on Mechanics to the Royal Chamber of Austrian-Hungarian Empress Maria Theresa, in 1770 unveiled before the Habsburg Royal Palace in Vienna a chessplaying automaton. Operating only with brass gears and springs, the “Terrible Turk” had the upper body of an Ottoman sultan whose eyes scanned the chessboard and glowered, whose head nodded from side to side, and who puffed on a Turkish pipe held in his right arm while he moved the chess pieces with his left arm. His chessplay was legendary for its brilliance, even though the cabinet could be opened from all four sides and peered through to reveal nothing but whirring clockwork and springs. This early touring machine made the rounds of all the royal courts of Europe, drawing throngs in Paris, Dresden, Leipzig, Brussels, and London, defeating nearly all opponents (including Joseph II and Napoleon Bonaparte) and making a fortune for von Kempelen. It finally reached the New World via Cuba, and eventually perished in a museum fire in Philadelphia.

– the “bête-machine” doctrine. This difference between animals and men reflected both Descartes’s metaphysical dualism (that the mental and the corporeal are two different sorts of substance) and his Christian theology (which precluded awarding animals souls). Hence animals were unconscious bodies, and, for example, he argued that we should feel no qualms about inflicting pain on them. We cannot do so.

Just as Aquinas had earlier used the metaphor of clockwork mechanisms to explain animals’ instincts, Descartes (1664, from Vartanian, 1953, p. 136) attributed our own mental experiences to the function of the body’s organs as clockwork mechanisms within an automaton:

... the reception of light, sounds, odors, tastes, warmth, and other like qualities into the exterior organs of sensation; the impression of the corresponding ideas upon a common sensorium and on the imagination; the retention or imprint of these ideas in the Memory; the internal movements of the Appetites and Passions; and finally, the external motions of all the members of the body . . . I wish that you would consider all of these as following altogether naturally in this Machine from the disposition of its organs alone, neither more nor less than do the movements of a clock or other automaton from that of its counterweight and wheels. . . .

Among other major seventeenth-century and Enlightenment invocations of mechanical and clockwork metaphors, Hobbes (1658), *De Homine* (from Vartanian,

1953, p. 138), sought to trace the physical basis for ideas and associations to minute mechanical motions in the head, and thus to turn epistemology into a branch of the new physics of Newtonian mechanics. In the same spirit, Hartley (1749), *Observations on Man* (from Vartanian, 1953, p. 141), proposed that all mental phenomena result from *vibratory motions* in the brain, by virtue of which arise the laws of association, causing ideas to cohere. And perhaps the most seminal of all Enlightenment appearances of this metaphor was La Mettrie's *L'Homme machine* (1748, from Vartanian, 1953, p. 139) which, based on profuse evidence of correlation between mental and physical states, described the human brain and body as "a machine that winds its own springs – the living image of perpetual motion . . . man is an assemblage of springs that are activated reciprocally by one another."

Among contemporary incarnations of the hydraulic metaphor, probably the most pervasive one is not so much a literal theory of brain function as a theory of the drives of the psyche. Psychodynamic theory holds that both individual behavior and social history are the outgrowth of invisible forces of desire and repression. For structuralists such as Freud and Marx, the visible is to be interpreted in terms of the hidden: economic and social forms are the superstructure thrown up by an invisible "deep structure" of class struggle in the Marxist perspective, just as in the Freudian framework, the individual's conscious experience and behavior are the manifestation of a surging unconscious libidinal struggle between desire and repression. From the psychodynamic viewpoint, the present is pregnant with the past ("the child is father to the man"), and from the social-historic viewpoint, the present is pregnant with the future (class struggle being the midwife), but in both cases, what is visible in the behavior of the individual or of society must be explained in terms of underlying hydraulic forces that eventually will have their way. Similar to water or steam pressure, which cannot build up indefinitely without release, the internal psychic or social pressures must inevitably express themselves in one form or another. According to these modern incarnations of the hydraulic metaphor, war, artistic movements, work, passion, religion, revolution, and new economic forms are the result.

In contemporary brain theory, sometimes the hydraulic metaphor appears in reverse form, in which something analogous to pressure (or voltage) *drops*, rather than rises. One phenomenon that this *complementary* hydraulic metaphor is meant to illuminate is the hemispheric lateralization of the brain. Kosslyn (1987) has recently proposed that the "activation" of either hemisphere consumes a depletable resource, like energy; the more "activation" is used up by one hemisphere, the less there is available for the other, rather like water pressure or electric power. Kosslyn uses the metaphor of an electric toaster (not because it resembles the brain's bilateral morphology), with an intuition that some sort of depletable resource is consumed when part of the brain is activated: "If so, then the hemisphere that receives input directly may draw more activation (rather like the way in which using a toaster will draw more current from the power grid), leaving less for the other hemisphere" (1987: 160). Although Kosslyn refers to an electric toaster on a power grid, this is in essence a reincarnation of the hydraulic metaphor.

Finally, perhaps one of the more beautiful twentieth-century incarnations of the mechanical metaphor, albeit in an expressly poetic form, was that of Sherrington (1906), who described the brain as an “enchanted loom,” in which “millions of flashing shuttles weave a dissolving pattern.”

5 Electronic and Optical Metaphors

The notion of the neuron as a switch, gate, or relay probably reflects earlier electro-mechanical devices for communication and computation. Long before Howard Aiken, John von Neumann, and ENIAC, mechanical devices had been proposed for adding, subtracting, and multiplying by Wilhelm Schickard (1623), Blaise Pascal (1642), and Charles Babbage (1822); even Leibniz had imagined⁴ a universal logical machine in which by a general method, “all the truths of reason would be reduced to a kind of calculation.” Echoes of electromechanical relays appear in Hebb’s (1949) theory of reverberating circuits underlying memory. Electronic communications systems such as the telegraph provided a basic nerve metaphor for Helmholtz, and indeed in the 1950s, the differential equation that describes coaxial cable transmission (the spatiotemporal “Telegrapher’s Equation,” which had been developed to model signal propagation for the design of the transatlantic undersea cable), was adopted directly by Hodgkin and Huxley (1952) in their Nobel Prize-winning studies of nerve action potential generation and propagation.

Optical metaphors for cognition have a long history, reaching back to antiquity, and appear even in such vernacular expressions for understanding as, “I see.” Greek epistemology held that our knowledge of the external world was mediated through optically transmitted copies (*eidola*) that objects emitted of themselves, entering our brains through the *pneuma*. In contemporary brain theory, one optical metaphor that recently became widely popularized was the holographic metaphor. Inspired by the obviously distributed character of neural information storage, and its relative insensitivity to brain injury, Pribram (1969) applied van Heerden’s (1963) research on holographic information storage properties to propose a general holographic metaphor for perception and memory.

6 Networks and Societies of Simple Automata

Earlier we examined seventeenth- and eighteenth-century visions of simple mechanical automata underlying brain and mental function. Important recent theoretical movements have been based on whole armies of simple automata, each of which obeys very simple rules, unintelligently or “automatically.” A key motivation for such models is the notion of *synergy*, that the whole is significantly greater than the sum of its parts, so that remarkably intelligent behavior might emerge from surprisingly elementary subunits if they are collectively configured properly, and with sufficiently rich interconnections. This idea lies at the core of many

contemporary brain theories based on “cellular automata,” “neural network,” “connectionism,” and “parallel distributed processing.”

Dennett (1978) has pointed out the philosophical merit of this approach to implementing the reductionist program. He argues that whereas visions of a *single* internal homunculus (who “watches” the internal movie screen of the world, and so on) are of course philosophical anathema, a vision of an *army* of homunculi, provided they are reductively stupid, may be philosophically felicitous and fruitful. The reductionist project might be achieved by positing hierarchies of successively more stupid homunculi, each of which behaves in a way that can be understood by its reduction to still less intelligent homunculi, until at the base of this hierarchy, the homunculi are true automata.

The metaphors based on networks or societies of simple automata may be coarsely divided into deterministic and stochastic versions.

Deterministic

Most of the seminal notions of networks of automata have been deterministic, with each formal neuron or daemon obeying specific and simple rules. Thus the ancestral prototype of such models, by its combination of microscopic simplicity with unpredictable emerging macroscopic pattern formation, is the *cellular automaton* first investigated systematically by John von Neumann. An idea with far-reaching influence, its descendants include the classic McCullough and Pitts (1943) treatment of formal neurons whose behavior is governed by a simple logical calculus, and the negatively influential study of simple perceptual automata by Minsky and Papert (1969), *Perceptrons*. Much current connectionist and neural network modeling makes the argument that well-orchestrated combinations of simple and stupid elements (thresholded summators, adaptive linear combiners, Boolean gates) can collectively achieve marvelous things.

Other variants of this idea include the general approach to human cognition as a society of simpler daemons, envisioned in Minsky's (1988) *Society of Mind*, which portrays the mind as emerging from its multitudinous cognitive components in much the same way as society emerges from its collectively interacting individuals. (This socio-political brain metaphor is clearly indebted to Hobbes's *Leviathan*, and offers the prospect of diverse forms of neuro-political organization such as social democratic, autocratic or winner-take-all, anarchic, and fascist.) Finally, in the psychodynamic dimension, an earlier vision of the mind as a deterministic collection of unconscious, automatic mechanisms is Freud's (1904) *The Psychopathology of Everyday Life*, which, in a similar manner, decomposes the life of the psyche into ensembles of numerous unconscious automata.

Stochastic

Various of the “daemon” network models recently proposed have possessed stochastic spirits, either specified by probabilistic transfer functions, Markov transition

rules, or more explicitly thermodynamic characteristics governed by stochastic differential equations. Those models which are serious about their thermodynamics, such as the Boltzmann machine (Ackley et al., 1985), Markov random field models, and relaxation labeling schemes, exploit certain analogies with statistical physics in which the evolution toward equilibrium or asymptotic configurations bears some resemblance to such cognitive phenomena as categorization, learning, recognition, memory, and decision-under-uncertainty.

Solid-state physics has provided salient metaphors for such theories. Perhaps the best-known of these is the spin-glass lattice popularized by Hopfield (1982) as a neural network formalism, which inspires Ising models of local domain formation by nearest-neighbor interactions. The dynamics of such networks are represented by trajectories (or flows) through a state space of Ising spins (Gutfreund et al., 1988), with these flows corresponding to the minimization of a global energy function. An essential element of the spin-glass metaphor is that in these networks every neuron is connected to every other neuron, so that distance has no meaning. Indeed, all "neurons" in such a lattice are nearest neighbors and, usually, all their interactions are symmetric; these two properties lead to a formal representation in an infinite-dimensional Hilbert space, in which all pairs of points are separated only by a unit distance. Although such globality is clearly a significant departure from brain connectedness, there have been several compelling demonstrations in which the chaotic dynamics that emerge can solve certain optimization and classification problems (Amari, 1988; Shaw, 1981). Recently, based on experimental observations in rabbit olfactory bulb, Skarda and Freeman (1987) have explicitly proposed that chaotic dynamics and their attractors may have great functional significance in neural mechanisms of olfactory recognition.

7 The Computational Metaphor

At its core the computational metaphor of brain function invokes the notion of formal rules for the manipulation of symbols, as well as certain ideas about data structures for representing information. Surprisingly, given the pervasive popularity of this metaphor, there remains today no well-established evidence of symbolic manipulation or formal logical rules at the neurobiological level in animal physiology. Perhaps the closest supporting neural evidence for symbol manipulation comes from human neuropsychological studies of language use (Geschwind, 1974), which demonstrate that specific brain lesions can result in specific *syntactic* deficits. But in general, while the computational metaphor often seems to have the status of an established fact, it should be regarded as a hypothetical, and historical, conjecture about the brain.

The computational or "Boolean brain" notion was originally expressed in the seminal proposal by McCullough and Pitts (1943) that nervous activity embeds a logical calculus, and was explored further in very general terms by John von Neumann (1958) in *The Computer and the Brain*. Alan Turing had proposed in 1950

the famous "Turing test" for justifying the assertion that under certain circumstances machines can be said to think, based on a positivistic criterion for distinguishing between artificial and human intelligence. A consequence of the Turing test is that if, in a teletype conversation with a computer no basis can arise for proving that there is not a real person producing the responses, then it is not valid to deny to the machine, in principle, any of the potential attributes or accoutrements of personhood, such as beliefs and intentionality.

Through formal work on the foundations of computation, augmented by results in mathematical logic by Alonso Church and Kurt Gödel on decidable propositions and computable functions, and second-order logic containing predication over predicates, the notion of the Universal Turing Machine was born together with the suggestion in some quarters that human brains might be instances of such machines. A central aspect of the Turing Machine concept is its focus on abstract and formal properties, independent of their hardware realization. The principle of the irrelevance of implementation has allowed cognitive scientists to adopt this metaphor in fairly direct form without requiring brains in any way to resemble computational devices physically. With the emergence of the terms "artificial intelligence" and "cognitive science" after the famous 1956 Dartmouth conference, which led to the creation of AI laboratories at major universities, the computational metaphor for brain function had become a slogan and a banner without par. Among the myriad recent expressions of adherence to the computational metaphor, perhaps the strongest statement advocating its adoption as the paramount theoretical framework for the cognitive and brain sciences is Pylyshyn's *Computation and Cognition* (1986).

Today's embrace of the computational metaphor in the cognitive and neural sciences is so widespread and automatic that it begins to appear less like an innovative leap than like a bandwagon phenomenon, of the sort often observed in the sociology and history of science. There is a tendency to rephrase every assertion about mind or brains in computational terms, even if it strains the vocabulary or requires the suspension of disbelief. In cognitive usage the meaning of the phrase "computational theory" has changed, so that now the term can be invoked to describe any theory that includes a *task analysis*. This appears to be an error of overextension (i.e. computational models include task analysis, so a theory with task analysis is a computational theory); and the descriptor "computational" then implies that the theory possesses properties and formal notions that in fact it does not. The irresistible temptation to announce "a computational theory of *X*," where *X* could be anything from the gill-withdrawal reflex of *Aplysia* (Gluck and Thompson, 1987) to the hemispheric lateralization of the entire brain and its variability among individuals (Kosslyn, 1987) typically reveals more about the sociological diffusion of slogans than about any proven or natural connection between *X* and computation. What yesterday would have been called a theory must today be called a computational theory, whether or not any deep relationship with computability (beyond some general notion of process or transformation) is established or even explicitly articulated.

As Earl MacCormac (1985) put it: "Theories require metaphors to be both hypothetical and intelligible." But perhaps like the goslings of Konrad Lorenz (to adopt

another metaphor), we are too easily imprinted with the spectacle of the day, the "duckiest" device in sight. That we should have chosen to model the brain as a computing machine, and even to believe that we have found at last the essence of our personhood in computation, is perhaps in historical perspective no more surprising than the fact that Freud's pre-eminent metaphor for personhood involved hydraulic forces or that La Mettrie's was based on a clockwork mechanism.

What seems remarkable about today's pervasive computational metaphor in neuroscience and psychology is how great is the distance between everything that is actually known about neurobiology and cognition, on the one hand, and the basic logical structure, formal sequence, tokens, powers, and limitations of computation on the other hand. With the exception of human linguistic competence, there is very scarce evidence indeed in the nervous systems of animals for any biological mechanisms of formal symbol manipulation. In some ways, the properties of pattern recognition and representation found in the immune and genetic systems, as Edelman (1987) has recently suggested, may provide a more germane model for information processing (and intentionality) in biology. Even in very mundane terms, it almost belabors the obvious to point out that where syllogisms and symbols are concerned, people are terrible at computation, and that conversely, any current implementation of computation is terrible at doing what people do or simulating what they are like. Although conscious experience obviously may not reveal anything about underlying mechanisms, who among us finds any recognizable strand of their personhood, or of their experience of others and of the world and its passions, to be significantly illuminated by, or distilled in, the metaphor of computation? Perhaps the ascendancy of this metaphor within the discourse of neuroscience and psychology, like the images of its many technological predecessors in a great hall of mirrors, is more a reflection of the sociology of science, its bandwagons and its bewitchments, than any of us may care to admit. We should remember that the enthusiastically embraced metaphors of each "new era" can become, like their predecessors, as much the prisonhouse of thought as they at first appeared to represent its liberation.

Acknowledgments

I am grateful to Cathryn Downing, Gary Hatfield, and Scott Weinstein for their helpful criticisms and suggestions.

Notes

- 1 Expressions of this view have taken many forms, from Michel Foucault's perspective (in *The Order of Things: An Archaeology of the Human Sciences*) that scientific thought and belief should be regarded in the same way as a literary text, to Earl MacCormac's view (in *Metaphor and Myth in Science and Religion*) that scientific explanations would simply not work as explanations if not for their metaphors, to Thomas Kuhn's view (in *The*

Structure of Scientific Revolutions) that the preferred or permissible metaphors within a scientific paradigm determine what will be accepted as an explanation and what problems remain as unsolved puzzles.

- 2 The *T'ai P'ing Kuang Chi* Encyclopedia, known as the "Extensive Records Made in the Period of Peace and Prosperity." Completed in the year 978 and published in 981.
- 3 For the review of historical material in this section, I am indebted to Vartanian (1953, 1973).
- 4 For this material I am indebted to Resnikoff (1988).

References

- Ackley, D. H., Hinton, G. E., and Sejnowski, T. J. 1985: A learning algorithm for Boltzmann machines. *Cognitive Science*, 9, 147–69.
- Amari, S. 1988: Statistical neurodynamics of associative memory. *Neural Networks*, 1 (1), 63–74.
- Arbib, M. A. 1972: *The Metaphorical Brain*. New York: Wiley Inter-science.
- Barthes, R. 1964: *Elements of Semiology* (Translated from French.) New York: Hill and Wang.
- Borges, C. 1967: *El Libro de Los Seres Imaginarios* (trans. *The Book of Imaginary Beings*). Buenos Aires: Editorial Kier, S.A.
- Churchland, P. 1986: *Neurophilosophy: Towards a Unified Science of the Mind-Brain*. Cambridge, MA: MIT Press.
- Dennett, D. C. 1969: *Content and Consciousness*. London: Routledge and Kegan Paul.
- Dennett, D. C. 1978: Artificial intelligence as philosophy and as psychology. In M. Ringle (ed.), *Philosophical Perspectives on Artificial Intelligence*, New York: Humanities Press.
- Descartes, R. 1664: *Traité de l'homme*. Trans. by T. Hall (1972), *Treatise on Man*. Cambridge, MA: Harvard University Press.
- Dreyfus, H. L. 1972: *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper and Row. See also H. L. Dreyfus, *Alchemy and Artificial Intelligence*, RAND Publication P-3244. Santa Monica, December 1965.
- Dreyfus H., and Dreyfus, S. 1986: *Mind over Machine*. New York: Free Press.
- Edelman, G. 1987: *Neural Darwinism: The Theory of Neuronal Group Selection*. New York: Basic Books.
- Egecioglu, O., Smith T., and Moody, J. 1987: Computable functions and complexity in neural networks. In J. Casti and A. Karlqvist (eds), *Real Brains, Artificial Minds*, New York: Elsevier.
- Foucault, M. 1973: *Les Mots et les choses* (trans. *The Order of Things: An Archaeology of the Human Sciences*). New York: reprinted by Random House.
- Geschwind, N. 1974: *Selected Papers on Language and the Brain*. Dordrecht: Reidel.
- Gluck, M., and Thompson, R. 1987: Modelling the neural substrates of associative learning and memory: A computational approach. *Psychological Review*, 94 (2), 176–91.
- Gunderson, K. 1985: *Mentality and Machines*, 2nd edn. Minneapolis: University of Minnesota Press.
- Gutfreund, H., Reger, J. D., and Young, A. P. 1988: The nature of attractors in an asymmetric spin glass with deterministic dynamics. *Journal of Physics A: Mathematical and General*, 21, 2775–97.
- Haugeland, J. 1985: *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.

- Haugeland, J. 1987: *Mind Design: Philosophy, Psychology, Artificial Intelligence*. Cambridge, MA: MIT Press.
- Hebb, D. 1949: *The Organization of Behavior*. New York: Wiley.
- Hinton, G. E., Sejnowski, T. J., and Ackley, D. H. 1984: Boltzmann machines: Constraint satisfaction networks that learn. *Technical Report CMU-CS-84-119*, Carnegie-Mellon University.
- Hodgkin, A. L., and Huxley, A. F. 1952: A quantitative description of membrane current and its application to conduction and excitation in nerve, *Journal of Physiology*, 117, 500–44.
- Hopfield, J. J. 1982: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences (USA)*, 79, 2554–8.
- Jay, M. 1973: *The Dialectical Imagination: A History of the Frankfurt School and the Institute of Social Research, 1923–1950*. Boston: Little, Brown.
- Kosslyn, S. 1987: Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, 94 (2), 148–75.
- Kuhn, T. 1962: *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- La Mettrie 1748: *L'Homme machine*. Holland: Jean-Jacques Pauvert (ed.), Nr 40, 1966.
- MacCormac, E. R. 1976: *Metaphor and Myth in Science and Religion*. Durham, NC: Duke University Press.
- MacCormac, E. R. 1985: *A Cognitive Theory of Metaphor*. Cambridge, MA: MIT Press.
- McCulloch, W. S., and Pitts, W. 1943: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115–33.
- Medawar, P. 1968: *Induction and Intuition in Scientific Thought*. Philadelphia: American Philosophical Society.
- Minsky, M. 1988: *Society of Mind*. New York: Simon and Schuster.
- Minsky, M., and Papert, S. 1969: *Perceptrons*. Cambridge, MA: MIT Press.
- Pribram, K. 1969: The neurophysiology of remembering. *Scientific American*, 200, 73–86.
- Putnam, H. 1960: Minds and machines. In A. Anderson (ed.), *Mind and Machines*, Englewood Cliffs, NJ: Prentice-Hall.
- Pylyshyn, Z. W. 1980: Cognition and computation: Issues in the foundations of cognitive science. *Behavioral and Brain Science*, 3 (1), 111–32.
- Pylyshyn, Z. W. 1986: *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Resnikoff, H. L. 1988: *The Illusion of Reality: Topics in Information Science*. Heidelberg: Springer-Verlag.
- Rorty, R. 1979: *Philosophy and the Mirror of Nature*. Princeton, NJ: Princeton University Press.
- Rosen, R. 1987: On the scope of syntactics in mathematics and science: The machine metaphor. In J. Casti and A. Karlqvist (eds), *Real Brains, Artificial Minds*, New York: Elsevier.
- Scheffler, I. 1967: *Science and Subjectivity*. Indianapolis: Bobbs-Merrill.
- Shaw, R. 1981: Strange attractors, chaotic behavior, and information flow. *Zeitschrift für Naturforschung*, 36 (a), 80–112.
- Sherrington, C. 1906: *The Integrative Action of the Nervous System*. New Haven, CT: Yale University Press.
- Simon, H. A. 1981: *The Sciences of the Artificial*, 2nd edn. Cambridge, MA: MIT Press.
- Skarda, A., and Freeman, W. J. 1987: How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10 (2), 161–95.
- Unger, R. M. 1975: *Knowledge and Politics*. New York: Free Press.

- van Heerden, P. 1963: Theory of optical information in solids. *Applied Optics*, 2, 393–400.
- Vartanian, A. 1953: *Diderot and Descartes: A Study of Scientific Naturalism in the Enlightenment*. Princeton, NJ: Princeton University Press.
- Vartanian, A. 1973: *Dictionary of the History of Ideas: Studies of Selected Pivotal Ideas*, ed. P. P. Wiener. New York: Scribners.
- von Neumann, J. 1958: *The Computer and the Brain*. New Haven, CT: Yale University Press.
- Wiener, N. 1948: *Cybernetics, of Control and Communication in the Animal and the Machine*. Cambridge, MA: MIT Press.