

1 Introduction

Major advances in science often consist in discovering how macroscale phenomena reduce to their microscale constituents. These latter are often counterintuitive conceptually, invisible observationally, and troublesome experimentally. Thus, for example, temperature in a gas turned out to be mean kinetic energy of the constituent molecules; the varied properties displayed by matter turned out to be a function of the component atoms and their arcane properties such as electron shells; bacteria—not Divine vengeance—were found to be the proximal cause of smallpox and bubonic plague; and the reproduction of organisms, we now know, depends on the arrangement of four bases in the molecule DNA.

Our psychological life, too, is a natural phenomenon to be understood. Here as well, the explanations will draw on properties of the infrastructure that are certainly veiled and probably arcane, an infrastructure whose *modus operandi* may seem alien to our customary self-conception. Perhaps this is inevitable, since the very brain we wish to understand is also the brain whose unaided observation is focused at the macrolevel and whose design seems to favor large-scale concepts for the explanation of its own behavior; for example, superstructure concepts such as "is hungry," "wants food," "believes honey is in the hole up the oak tree," and "sees the grizzly bear approaching."

Neurons are the basic structural components of the brain. A neuron is an individual cell, specialized by architectural features that enable fast changes of voltage across its membrane as well as voltage changes in neighboring neurons. Brains are assemblies of just such cells, and while an individual neuron does not see or reason or remember, brains regularly do. How do you get from ion movement across cell membranes to memory or perception in brains? What is the nature of neuron-neuron connectivity and interactivity? What makes a clump of neurons a nervous system?

At this stage in the evolution of science, it appears highly probable that psychological processes are in fact processes of the physical brain, not, as Descartes concluded, processes of a nonphysical soul or mind. Since this issue has been discussed at length elsewhere (for example, P. M. Churchland 1984, P. S. Churchland 1986), and since Cartesian dualism is not taken very seriously either in mainstream philosophy or mainstream neuroscience, it is not necessary to repeat the details of the arguments here. Suffice it to say that the

Cartesian hypothesis fails to cohere with current physics, chemistry, evolutionary biology, molecular biology, embryology, immunology, and neuroscience. To be sure, materialism is not an established fact, in the way that the four-base helical structure of DNA, for example, is an established fact. It is possible, therefore, that current evidence notwithstanding, dualism might actually be true. Despite the rather remote possibility that new discoveries will vindicate Descartes, materialism, like Darwinian evolution, is the more probable working hypothesis. That being so, it does not seem worthwhile to modify the basic neuroscience research program and its scaffolding of physicalistic presuppositions to accommodate the Cartesian hypothesis, though scientific tolerance counsels that the door not be closed until the facts themselves well and truly close it. Whether modifications to micro/nano/pico level sciences such as quantum physics will be called for as a result of advances in neuropsychology is likewise conceivable (Penrose 1989), but so far there is no moderately convincing reason to expect that they will.

Arguments from ignorance are to be especially guarded against in this context. Their canonical form is this: neuroscience is ignorant of how to explain X (consciousness, for instance) in terms of the nervous system; therefore it cannot be so explained. Rather, it can eventually be explained in terms of Y (pick your favorite thing, for example, quantum wave packets, psychons, ectoplasmic retrovibrations, etc.). The canonical form lends itself to endless seductive variations, particularly ones in which failures of imagination massage intuition: "We cannot *imagine* how to explain consciousness in terms of neuronal activity . . . ; how could physical processes like ions crossing membranes explain the awfulness of pain?" In its denuded rendition, the argument from ignorance is not mildly tempting, but in full regalia, it may seem beguiling and exactly what reharmonizes such "intuition dissonance" as is provoked by reflecting on the physical basis of the mental. A version of the argument convinced the German mathematician and philosopher, Leibniz (1714), and in the past two decades, variations on Leibniz' basic theme have surfaced as the single most popular and appealing justification for concluding that neurobiological explanations of psychological phenomena are impossible. (For instances of the argument in many different and alluring guises, see Thomas Nagel 1974, J. C. Eccles 1977, John Searle 1980, 1990, and Roger Penrose 1989.) From the revolutions wrought by Copernicus, Galileo, Darwin, and Einstein, it is all too apparent that "intuition dissonance" is a poor indicator of truth; it is a good indicator only of how one idea sits with well-favored others. Establishing truth or probability requires rather more.

The working hypothesis underlying this book is that emergent properties are high-level effects that depend on lower-level phenomena in some systematic way. Turning the hypothesis around to its negative version, it is highly improbable that emergent properties are properties that cannot be explained by low-level properties (Popper 1959), or that they are in some sense irreducible, causally *sui generis*, or as philosophers are wont to say, "nomologically autonomous," meaning, roughly, "not part of the rest of science" (Fodor 1974, Pylyshyn 1984). The trouble with characterizing certain properties as irreduc-

ibly emergent is that it assumes we can tell in advance whether something can be explained—*ever* explained. Obviously such a claim embodies a prediction, and as the history of science shows all too clearly, predictions grounded in ignorance rather than knowledge often go awry. In advance of a much more highly developed neurobiology than currently exists, it is much too soon to be sure that psychological phenomena cannot be explained in terms of neurobiological phenomena. Although a given phenomenon such as protein folding or awareness of visual motion cannot *now* be explained, it might yield to explanation as time and science go on. Whether it does or not is a matter of empirical fact, not a matter of *a priori* divination. Searching for reductive explanations of emergent properties does not entail that we should expect the explanations to be simpleminded or breezily cobbled up or straightforwardly readable off the data points; it means only that the betting man keeps going.

Two groundbreaking discoveries in the nineteenth century established the foundations for a science of nervous systems: (1) macro effects displayed by nervous systems depend on individual cells, whose paradigm anatomical structures include both long tails (axons) for sending signals and treelike proliferations (dendrites) for receiving signals (figure 1.1); (2) these cells are essentially electrical devices; their basic business is to receive and transmit signals by causing and responding to electric current. Within this elegantly simple framework, truly spectacular progress has been made in unravelling the intricate story of exactly how neurons work. In this century, and especially within the

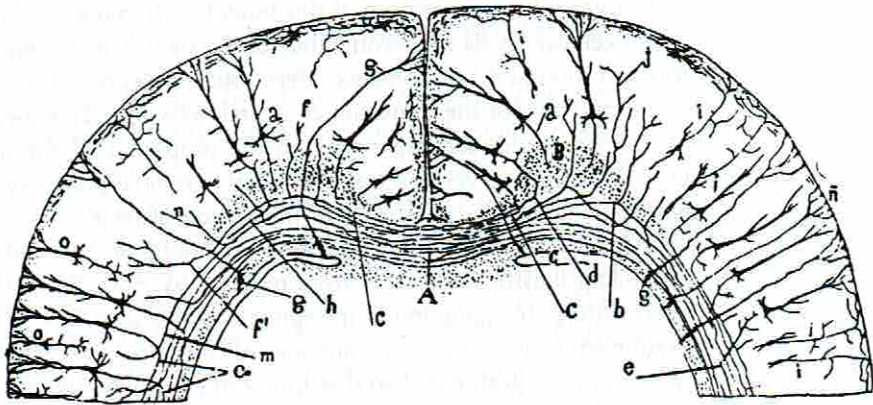


Figure 1.1 Drawing by Cajal based on his Golgi-stained sections of the superior part of the cerebral hemispheres and corpus callosum of a mouse of 20 days. A, corpus callosum; B, antero-posterior fibers; C, lateral ventricle; a, large pyramidal cell; b, callosal fiber bifurcating into a branch that is arborized in the gray matter and another that continues in the corpus callosum; c, callosal fiber that comes from an axon of the white matter; d, callosal fiber the originates in a pyramidal cell; e, axons of lateral pyramidal cells which follow a descending course in the corpus callosum without forming part of the commissure; f, f', the two final branches coming from a fiber of the corpus callosum and arborizing in the gray matter; g, epithelial cells; h, fiber from a large pyramid, giving off a fine collateral to the corpus callosum; i, fusiform cells whose axons ascend to the molecular layer; j, terminal arborization of a callosal fiber originating on the opposite side. (With permission. Santiago Ramón y Cajal, 1890. Reprinted in DeFelipe and Jones, eds., 1988, *Cajal on the Cerebral Cortex*. Oxford: Oxford University Press.)

6

last three decades, an enormous amount has been learned about neurons: about their electrophysiology, microanatomy, connectivity, and development; about the large assortment of neurochemicals that mediate signaling from one neuron to the next; inside a neuron, about the cell's membrane, its roster of channel types, and their specific roles in receiving, integrating, and sending signals; about transmitter release, and about the range, structure, and mechanisms of receptors. Even the genetics of the proteins that constitute the various receptors is now steadily coming into view. (Nathans 1987, 1989, Gasic and Heinemann, 1991, Heinemann et al. 1990).

Recent progress in neuroscience is genuinely breathtaking and deservedly captivating. But, the naïf might wonder why, if we know so much about neurons, do we not yet understand how the brain works—or at least how, say, the visual system or the motor system works? Assuming that detailed knowledge of the parts automatically confers (or nearly so) knowledge of the whole, then we ought to understand—more or less, at least in silhouette—how animals see, learn, and take action. In fact, however, we do not. Perhaps the hitch is that microlevel progress notwithstanding, we still do not know nearly enough about the fine-grained neural facts. All that is needed, runs this argument, is more of the same—indeed, much, much more of the same. This strategy is sometimes referred to as the pure bottom-up approach. It counsels that if brains are, after all, just assemblies of cells, then once we truly understand every facet of cell function, the principles of brain function will be evident, by and large. Perhaps. But perhaps not.

The overarching contention of this book is that knowledge of the molecular and cellular levels is essential, but on its own it is not enough, rich and thorough though it be. Complex effects, such as representing visual motion, are the outcome of the dynamics of neural networks. This means that while network properties are dependent on the properties of the neurons in the network, they are nevertheless not identical to cellular properties, nor to *simple* combinations of cellular properties. Interaction of neurons in networks is required for complex effects, but it is dynamical, not a simple wind-up doll affair.

A telling illustration derives from research by Allen Selverston (1988) on the stomatogastric ganglion of the spiny lobster (figure 1.2).¹ The network in question contains about 28 neurons and serves to drive the muscles controlling the teeth of the gastric mill so that food can be ground up for digestion. The output of the network is rhythmic, and hence the muscular action and the grinders' movements are correspondingly rhythmic.

The basic electrophysiological and anatomical features of the neurons have been catalogued, so that the microlevel *vitae* for each cell in the network is impressively detailed. What is not understood is how the cells interact to constitute a circuit that produces the rhythmic pattern. No one cell is responsible for the network's rhythmic output; no one cell is itself the repository of properties displayed by the network as a whole. Where then does the rhythmicity come from? Very roughly speaking, from the pattern of interactions among cells *and* the intrinsic properties of component cells. What, more precisely speaking, *is* that? How does the network create rhythm? How is it

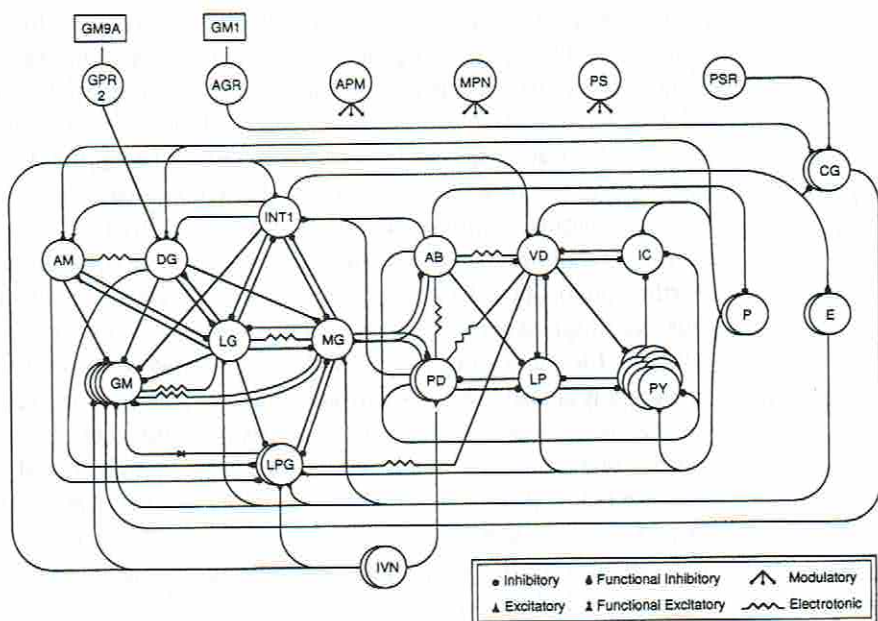


Figure 1.2 Diagram of the circuit in the stomatogastric ganglion of the spiny lobster. The circuit normally has 28 neurons, and for each, its connectivity (whom it affects and who affects it), sign of connectivity (excitatory or inhibitory), and mode of effect (chemical or electrical) have been discovered. Labels on cell bodies stand for their individual names. (Courtesy Allen Selverston.)

that the network can produce different rhythms under different biochemical conditions?

Research on the stomatogastric ganglion is legendary in neurobiology, partly because it is a fair test case for the bottom-up strategy: if the purely bottom-up approach works anywhere, it should work on the stomatogastric ganglion. If the macrolevel answers are supposed to fall out of the microlevel data, they ought to do so here. Yet we are disappointed. As Selverston ruefully points out, the purely bottom-up strategy has all the earmarks of a half-strategy. Moreover, the plea, "If only more microlevel details of the neurons were discovered, then the explanation would be evident," tends now to fall on skeptical ears. What the stomatogastric ganglion seems to be telling us is that we need to figure out the interactive principles governing the system, and that although interactive hypotheses should be constrained by microlevel data, their job is to characterize higher-level features. Boiled down, the lesson is that microlevel data are *necessary* to understand the system, but not *sufficient*. To echo a remark of Maxwell Cowan, even if we did know about all the synapses, all the transmitters, all the channels, all the response patterns for each cell, and so forth, still, we would not know how an animal sees and smells and walks.²

There is a broader rationale for modeling that goes beyond neuroscience in particular and applies to science generally. Why bother with models at all, one might ask? Why not just perform experiments and record the observations? Though the answers may be obvious, they are perhaps worth listing. First,

models help organize the data and motivate experiments; they suggest how data might fit together to yield an explanation of a phenomenon. It is, therefore, better to have some model than none at all. In fact, of course, scientists do always have some hypothesis or other that provides the motivational and interpretive framework for their research, though background hypotheses may be neither cleanly articulated nor well-honed. A quantitative model is a step forward because it brings background assumptions to the light of day and permits a more exacting analysis of why they might or might not be true. The further philosophical point is that models increase in believability as they survive tough experimental tests (Popper 1959, P. S. Churchland 1986). Especially in the pioneering days of a discipline, when data are relatively sparse, progress is closely tied to *ruling out* a class of models and hypotheses. Indefinitely many models can be equally consistent with a set of data; to make real strides one must seek to falsify an ostensibly plausible model. Consequently, models that suggest potentially falsifying experiments are critical.³ Should a model survive a demanding experimental test, to that degree it is more probable; saved from the scrap heap of dead hypotheses, it lives on to be tested against yet further experimental data. Should it be falsified, it then becomes a springboard for the next model.

Computational neuroscience is an evolving approach that aims to discover the properties characterizing and the principles governing neurons and networks of neurons. It draws on both neurobiological data and computational ideas to investigate how neural networks can produce complex effects such as stereo vision, learning, and auditory location of sound-emitting objects. To put it crudely, it has one foot in neuroscience and one foot in computer science. A third foot is firmly planted in experimental psychology, and at least a toe is in philosophy, so evidently the enterprise is multipedal. Of which more anon.

Probably the closest academic kin of computational neuroscience is systems neurobiology, a branch of neuroscience that traditionally has focused on much the same set of problems, but did not explicitly ally itself with computer modeling or with an avowedly information-processing framework for theories. A precocious ancestor went by the name of "cybernetics," which, inversely to systems neurobiology, generally leaned more heavily on the engineering and psychophysical sides, and more lightly on the neurobiological side. Coined more recently, "connectionism" usually refers to modeling with networks that bear only superficial similarities to real neural networks, while "neural net modeling" can cover a broad range of projects. Ironically perhaps, "neural net modeling" is usually identified with computer modeling of highly artificial nonneuronal networks, often with mainly technological significance such as medical diagnoses in emergency wards.⁴ "PDP" ("parallel distributed processing") is generally the preferred label of cognitive psychologists and some computer scientists who seek to model rather high-level activity such as face recognition and language learning rather than lower-level activity such as visual motion detection or defensive bending in the leech.

As we use the term, "computational neuroscience" aims for biological realism in computational models of neural networks, though *en route*, rather sim-

plified and artificial models may be used to help test and explore computational principles. Academic garden-plotting is a comically imprecise trade because the carrots regularly wander in with turnips and the turnips with the potatoes. Each of us (P.S.C. and T.J.S.) is cheerfully guilty of wandering into neuroscience from his mother discipline, so we emphatically do not mean to tut-tut academic "cross-fielding." On the contrary, we view the blurring of the disciplinary boundaries between neuroscience, computer science, and psychology as a healthy development to be wisely encouraged. In any case, perhaps a crude survey will help orient the greenhorn—or even the old hand—to the clustering of goals, tactics, and prejudices manifest in the "network" game.

The expression "computational" in computational neuroscience reflects the role of the computer as a research tool in modeling complex systems such as networks, ganglia, and brains. Using the word in that sense, one could have also computational astronomy or computational geology. In the present context, however, the word's primary force is its descriptive connotation, which here betokens the deep-seated conviction that what is being modeled by a computer is itself a kind of computer, albeit one quite unlike the serial, digital machines on which computer science cut its teeth. That is, nervous systems and probably parts of nervous systems are themselves naturally evolved computers—organically constituted, analog in representation, and parallel in their processing architecture. They represent features and relations in the world and they enable an animal to adapt to its circumstances. They are a breed of computer whose *modus operandi* still elude us but are the mother lode, so to speak, of computational neuroscience.

A number of broad clues about computation in nervous systems are available. First, unlike a digital computer which is general purpose and can be programmed to run any algorithm, the brain appears to be an interconnected collection of special-purpose systems that are very efficient at performing their tasks but limited in their flexibility. Visual cortex, for example, does not appear able to assume the functions of the cerebellum or the hippocampus. Presumably this is not because visual cortex contains cells that are essentially and intrinsically visual in what they do (or contain "visions" instead of "auditions"), but rather it is mainly because of their morphological specialization and of their place in the system of cells in visual cortex, i.e., relative to their input cells, their intracortical and subcortical connections, their output cells, and so on. Put another way, a neuron's specialization is a function of the neuron's computational roles in the system, and evolution has refined the cells better to perform those roles.

Second, the clues about the brain's computational principles that can be gleaned from studying its microstructure and organization are indispensable to figuring out its computational organization because the nervous system is a product of evolution, not engineering design. Evolutionary modifications are always made within the context of an organization and architecture that are already in place. Quite simply, Nature is not an intelligent engineer. It cannot dismantle the existing configuration and start from scratch with a preferred design or preferred materials. It cannot mull the environmental conditions

and construct an optimal device. Consequently, the computational solutions evolved by Nature may be quite unlike those that an intelligent human would invent, and they may well be neither optimal nor predictable from orthodox engineering assumptions.

Third, human nervous systems are by no means exclusively cognitive devices, though the infatuation with cognition fosters a tacit tendency to assume so. Nervous systems must also manage such matters as thermoregulation—a very complex function for mammals—growth, aspects of reproduction, respiration, regulation of hunger, thirst, and motor control, and maintenance of behavioral state, such as sleeping, dreaming, being awake, and so forth. Thus an evolutionary modification that results in a computational improvement in vision, say, might seem to have the earmarks of an engineering prizewinner. But if it cannot mesh with the rest of the brain's organization, or if it marginalizes critical functions such as thermoregulation, the animal and its "prize-winning" vision genes will die. Given these reasons, *reverse* engineering, where the device is taken apart to see how it works, is a profitable strategy with respect to the brain. By contrast, a purely *a priori* approach, based entirely on reasonable principles of engineering design, may lead us down a blind alley.

Fourth, it is prudent to be aware that our favorite intuitions about these matters may be misleading, however "self-evident" and compelling they be. More specifically, neither the nature of the computational problems the nervous system is solving nor the difficulty of the problems confronting the nervous system can be judged merely by introspection. Consider, for example, a natural human activity such as walking—a skill that is typically mastered in the first year or so of life. One might doubt whether this is a computational problem at all, or if it is, whether it is a problem of sufficient complexity to be worth one's reflection. Since walking is virtually effortless, unlike, say, doing algebra, which many people do find a strain, one might conclude from casual observation that walking is a computationally easy task—easier, at least, than doing algebra. The preconception that walking is computationally rather trivial is, however, merely an illusion. It is easy enough for toy manufacturers to make a doll that puts one foot in front of the other as long as she is held by the child. But for the doll to walk as we do, maintaining balance as we do, is a completely different task. Locomotion turns out to be a complicated matter, the ease implied by introspection notwithstanding.

Another computational issue of critical importance in generating hypotheses in computational neuroscience concerns the time available for performing the computation. From the point of view of the nervous system, it is not enough to come up with solutions that merely give the correct output for a given input. The solutions must also be available within milliseconds of the problem's presentation, and applications must be forthcoming within a few hundred milliseconds. It is important that nervous systems can routinely detect signals, recognize patterns, and assemble responses within one second. The ability of nervous systems to move their encasing bodies appropriately and swiftly was typically selected at every stage of evolution, since by and large natural selection would favor those organisms that could flee or fight preda-

tors, and catch and cache prey. *Ceteris paribus*, slow nervous systems become dinner for faster nervous systems. Even if the computational strategies used by the brain should turn out not to be elegant or beautiful but to have a sort of evolutionary do-it-yourself quality, they are demonstrably very fast. This tiny response-time rules out as just too slow many kinds of ostensibly elegant computational architectures and clever computational principles. This point is all the more significant when it is considered that events in an electronic computer happen in the nanosecond (10^{-9}) range, whereas events in neurons happen in the millisecond (10^{-3}) range.

A related consideration is that organic computers such as brains are constrained in the amount of space available for the essential elements—cell bodies, dendrites, axons, glial cells, and vascularization—and the cranial capacity is in turn limited by the mechanisms of reproduction. In mammals, for example, the size of the pelvic cavity of the mother constrains head size of offspring, and therefore brain size of offspring. What this all means is that the length of wiring in nervous systems must also be limited—evolution cannot just help itself to indefinite lengths of connecting wire but must make every centimeter count. In a human brain, for example, the total length of wiring is about 10^8 meters and it has to be packed into a volume of about 1.5 liters. The spatial configuration of sense organs and muscles on the body and the relative position of the afferent and efferent systems will also be relevant to the computational genre that has been selected in the evolution of nervous systems (figure 1.3). One strategy the brain uses to economize on wire is to map the processing units so that neighboring units process similar representations. Another strategy involves sharing wire, meaning that the same wire (axon) can be used in coding a large range of representations (Mead 1989). The computational genre adopted for a nervous system will, therefore, be constrained not only by temporal factors but also by spatial factors.

Computation is also limited by power consumption, and on this matter too the brain is impressively efficient. For example, a neuron uses roughly 10^{-15} joules of energy per operation (e.g., one neuron activating another at a synapse). By contrast, the most efficient silicon technology currently requires about 10^{-7} joules per operation (multiply, add, etc.) (Mead 1989). Using the criterion of joules per operation, the brain is about 7 or 8 orders of magnitude more power efficient than the best of the silicon chips. A direct consequence of their energy efficiency is that brains can perform many more operations per second than even the newest supercomputers. The fastest digital computers are capable of around 10^9 operations per second; the brain of the common housefly, for example, performs about 10^{11} operations per second when merely resting.

Finally, there are constraints imposed by the materials of construction. That is, cells are made out of proteins and lipids, they have to rely on mitochondria for their energy supply; nervous systems must have the substances and dispositions necessary for growth and development, and they must exploit such features as the membrane properties of cells and the available chemicals in order to function as an organic computer. Additionally, the nervous system

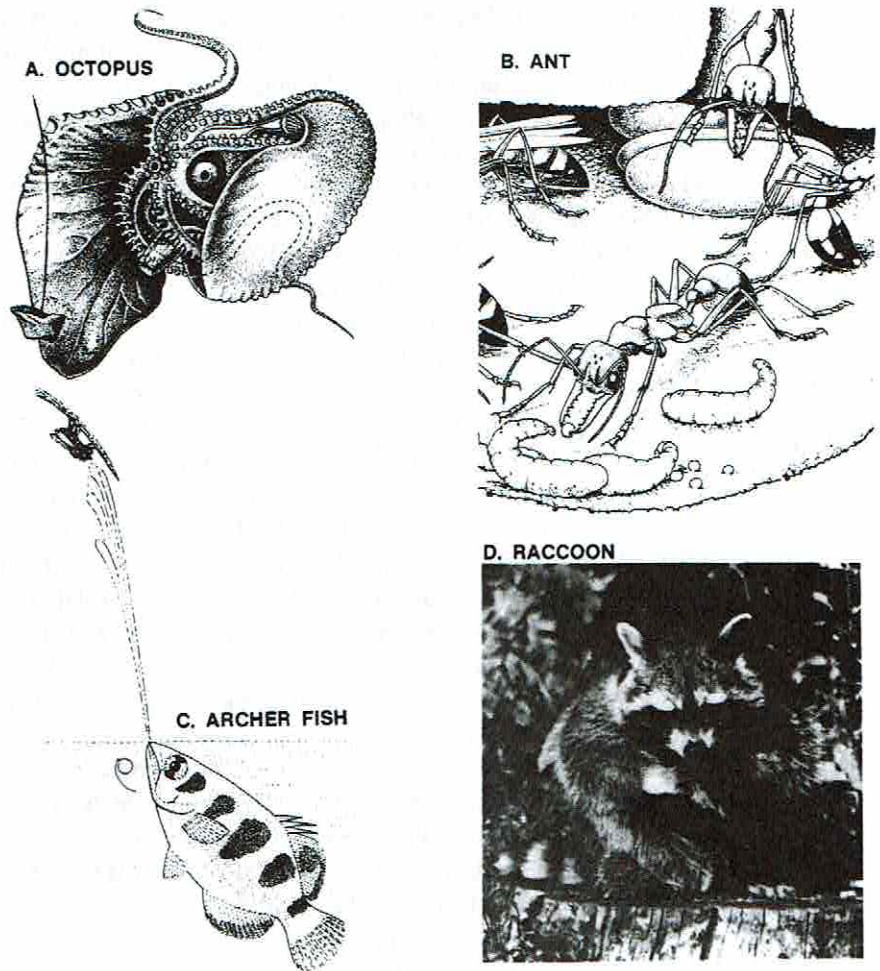


Figure 1.3 Evolutionary specializations of manipulation. The octopus (A) can manipulate objects with its tentacles, which are modified limbs; the ant (B) moves things with its pincers, which are modified jaws (mandibles). The archer fish (C) can use its mouth and pharynx to shoot droplets of water at airborne insects, an elementary form of tool use. The raccoon (D) performs dextrous manipulations of foodstuffs with its handlike paws. (From Shepherd 1987.)

needs a constant supply of oxygen and a reliable supply of nutrients. Evolution has to make what it can out of proteins, lipids, membranes, amino acids, etc. This is not altogether unlike the engineering make-do game where the given materials are limited (a finite number of popsicle sticks, rubber bands, and paper clips), and the task, for example, is to build a weight-supporting bridge. Indeed, John Allman (1990) has suggested that brain expansion in homeotherms was spurred by the need to engage in intense prey-catching in order to keep the home fires burning, as it were. In the competition for large amounts of fuel, homeotherms with sophisticated neural machinery that upgraded prey-catching and predator avoidance would have had an advantage.

Two conceptual ideas have structured much of how we tend to conceive of problems in computational neuroscience. First is the notion of *levels*, and the

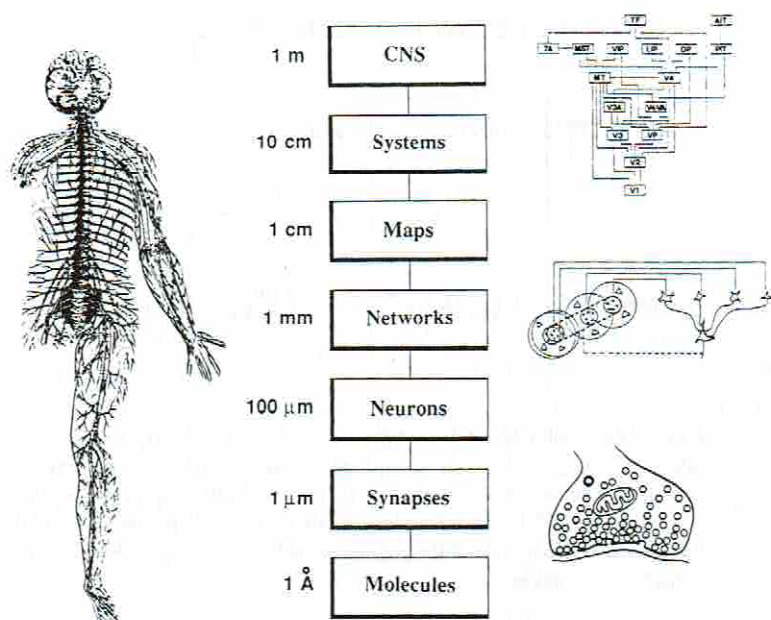


Figure 1.4 Schematic illustration of levels of organization in the nervous system. The spatial scales at which anatomical organizations can be identified varies over many orders of magnitude. Icons to the right represent structures at distinct levels: (top) a subset of visual areas in visual cortex (van Essen and Maunsell 1980); (middle) a network model of how ganglion cells could be connected to simple cells in visual cortex (Hubel and Wiesel, 1962), and (bottom) a chemical synapse (Kandel and Schwartz, 1985). (From Churchland and Sejnowski 1988.)

second concerns the *co-evolution* of research on different levels. In the brain, there is both large-scale and small-scale organization, and different functions take place on higher and lower levels (figure 1.4). One sort of account will explain how signals are integrated in dendrites; a different account will explain the interaction of neurons in a network, or the interaction of networks in a system.⁵ A model that captures the salient features of learning in networks will have a different face from a model that describes the NMDA channel. Nevertheless, the theories on one level must mesh with the theories of levels both higher and lower, because an inconsistency or a lacuna somewhere in the tale means that some phenomenon has been misunderstood. After all, brains are assemblies of cells, and something would be seriously amiss if neurons under one description had properties incompatible with the same neurons under another description.

That there are levels of organization is a matter of fact; co-evolution of research, on the other hand, is a matter of research strategy in light of the presumed fact. The hallmark of co-evolution of theories is that research at one level provides correction, constraints, and inspiration for research at higher and at lower levels (figure 1.5). Computational space is undoubtedly vast, and the possible ways to perform any task are probably legion. Theorizing at a high level without benefit of lower-level constraints runs the risk of exploring a part of that space that may be interesting in its own right but remote from where

	NECESSARY	NOT NECESSARY
SUFFICIENT	<i>PURE BOTTOM-UP</i>	<i>either TOP-DOWN or BOTTOM-UP</i>
NOT SUFFICIENT	<i>CO-EVOLUTION STRATEGY</i>	<i>PURE TOP-DOWN</i>

Figure 1.5 Possible research strategies for trying to understand how the brain works as they divide on the question of the importance of cellular and molecular levels in theories of brain function. Some neuroscientists may prefer the pure bottom-up strategy; some psychologists and philosophers prefer the pure top-down strategy; probably no one falls into the upper right box, but we added it to round out the possibility-table; the co-evolutionary strategy (lower left) is the one adopted in this book.

the brain's solutions reside. Thus microlevel constraints and the testing of hypotheses against microlevel facts are of central importance.

On the other hand, research from neuropsychology, neuroethology, and psychophysics, as well as experimental psychology generally, provide the detailed characterization of what needs to be explained at lower levels. Without a scientific delineation of cognitive and other psychological capacities, lower-level research has only incomplete and confused ideas about the very capacity whose mechanisms are the focus of research. Thus, for example, in studying the neuronal mechanisms of motion perception in visual areas MT and MST, it is wasteful not to have up-to-date information on the psychophysics of visual motion detection. Research on computational principles can profitably co-evolve with research in neuroscience and in psychology, for something new is to be learned from the brain about powerful computational methods, and neuroscience in return can absorb abstract discoveries in the theory of computing and practical discoveries in the construction of computing machines.

It is on the *network level* that we decided to concentrate discussion. We made this decision because network models can be more highly constrained by neurobiology than can high-level psychological models and because network models are more straightforwardly testable against the actual networks modeled. At the same time, we prefer models of capacities that are well studied in psychology and neuropsychology so that we can take advantage of economizing constraints from top-down research. This usually means that the capacities themselves will be rather low level; visual motion detection rather than planning, bending in the leech rather than chess playing in the human. For us, this is an advantage since it is discouraging to put effort into a model of some brain function if the model can be assessed only abstractly or aesthetically. Consequently, the models we select for more intensive discussion in this book

will be models that generally exhibit the "accessibility" features: the bending reflex in the leech, a model of the vestibulo-ocular reflex (VOR), and models of visual capacities such as stereopsis. Our particular choices are also guided by the independent virtues of these models. There are, of course, some drawbacks to "neurally close" models. In a nutshell, how realistic should they be before useful results drown in the detail? How much of the teeming detail is needed to get a reasonably accurate model? These questions will be taken up mainly in chapters 3 and 4.

Although "neurally close" models of "psychophysically dissectable" capacities are our preference, we hasten to say that we recognize that other scientists have quite different preferences, and that ideas useful for the science of the mind-brain may come from widely diverse locations in the research spectrum. Not only are we cognizant of the value of modeling research outside the compass of our particular prejudices, we would be dismayed if everyone were to share our prejudices, for the co-evolutionary advice regarding methodological efficiency is "let many flowers bloom." And at this stage in the history of neuroscience, the advice is entirely reasonable. First, because it is far too early in the hunt to know where the Big Breakthroughs will come or what they will look like. Second, because ultimately what is wanted is a story, unified from top to bottom—from behavior, through systems to networks to neurons and molecules: a unified science of the mind-brain.

Mathematical models and computer simulations of the single neuron have a distinguished tradition, beginning with Hodgkin and Huxley in 1952 and continuing with highly detailed and revealing models of motor neurons (Rall, 1964), Purkinje cells (Bush and Sejnowski, 1991), hippocampal pyramidal cells (Traub et al., in press), and dendritic processing (Segev et al., in press, Koch et al. 1990). The single neuron is not the main focus of this book, however, since a good choice of accessible texts with that focus already exists.⁶ To have put single neuron modeling on center stage would have meant not only redescribing the familiar, but also bypassing some little-known but ought-to-be-known models aimed at the network level. Our discussion by no means excludes single neuron models, however, and several detailed cellular models are introduced in the context of network models in which they might find a place (chapters 5 and 6). Moreover, we emphasize the importance of single neuron models as the bedrock and fundament into which network models must eventually fit. Network models are thus considered not in isolation from single neuron models, but as having a future wherein the two enmesh.

The rationale for a primer is threefold. First, we calculated it would be useful to present and discuss the conceptual framework of the emerging discipline of computational neuroscience, accompanied by a selection of sterling or anyhow seminal examples to flesh out the ideas. Sometimes, for both neophyte and cognoscenti, it can be worthwhile to step back from the crowd of trees and have a look at the shape of the woods.

Second, there are four broad constituencies—neuroscience, psychology, computer science, and philosophy—each voicing a specific and entirely legiti-

mate demand with respect to neuromodeling, and each harboring a specific complaint about other constituencies. Having encountered these demand-complaint pairs on innumerable occasions and being convinced of the sometime fruitfulness of boundary-fuzzing, we wanted to have a go at satisfying the demands and addressing the complaints. Made concise for presentation and arrayed to display parallelism, the demand/complaint pairs are articulated below:

The neuroscientist:

1. Show me results of neuromodeling that help explain or predict experimental results.
2. They (the nonneuroscientists) do not know anything much about neuroscience even though they are doing "neural modeling."

The psychologist:

1. Show me results of neuromodeling that help explain or predict psychological functions and behavior.
2. They (the nonpsychologists) do not know anything much about the results from psychophysics and psychology even though they are modeling psychological capacities and performance.

The computer scientists:

1. Show me results of neuromodeling that help understand the nature of computation and representation or that yield new ideas about these things.
2. They (the noncomputer scientists) do not know anything much about electrical circuits, mathematical analyses, or existing theory of computation.

The philosopher:

1. Show me results of neuromodeling that are relevant to philosophical problems concerning the nature of knowledge, the self, and the mind.
2. They (the nonphilosophers) do not understand some of the useful, time-saving, and agony-saving contributions of philosophers in constraining questions about how the mind works.

Since the demand/complaint pairs from the various constituencies are related, it seemed reasonable to try to combine our responses in an integrated text as a sort of conversation with diverse people. Moreover, since the constituencies are diverse, we wanted the book to be broadly accessible. New technical books and vast numbers of technical articles are appearing at a dizzying rate, and we judged that a less technical, more introductory text might be helpful in orienting in the midst of the technical literature. Where there are equations, we have given an English paraphrase, but in any event, they are skippable without substantial loss. References at the chapter ends permit the reader to follow up the discussion. To round out the presentation, we judged it necessary to include a brief exposure to basic neuroscience and to the foundational issues in computational theory. Thus, chapters 2 and 3 provide some background dis-

cussion on neuroscience and the science of computation. An appendix on neuroscience techniques and a glossary are also included.

The third element in the rationale was more self-oriented. The project forced us to leave intermittently the relative solace of the technical details to see what we could discern of the broader landscape and its contours. In a sense, then, the project has been an excuse to paint in broad strokes as well as an exercise in policing our implicit convictions and our covert enthusiasms. We found ourselves constantly hectoring each other with questions of this form, for many values of X and Y : what is the point of X , what does Y really mean, is X really of any use to anybody? In forcing each other to articulate answers, we often bumped up against the illusion that one's assumptions are generally well-honed, well-grounded, and entirely coherent with the rest of one's beliefs.

In chapters 4 to 7, we assume a basic background knowledge and proceed to introduce computational models. The first ones discussed are rather abstract models of visual functions that incorporate some neuronal details, but where the neural data are still unavailable, perforce they mirror the want of data. Models introduced in later chapters are increasingly realistic neurobiologically. Plasticity, introduced in chapter 5, has been modeled at many levels, from the very spare models that are virtually innocent of physiology, to higher-fidelity models of dendritic spine behavior whose grain is so fine as to include ion concentration parameters and diffusion times. In chapter 6 on sensory-motor integration, we chart the progress in Lockery's modeling of the bending reflex in the leech from a simple static model to the next increment of complexity, a model with dynamical properties, to plans—if not the finished product—for a model that includes channel properties. Likewise, the modeling of adaptation in the vestibulo-ocular reflex, though incomplete, includes the known dynamical and physiological properties of the circuits. Incorporating more cellular detail, Grillner's model of swimming in the lamprey has many physiological properties, including time constants for cellular responses and channel properties.

Obviously one intends that a model capture the salient features of reality modeled. At the same time, however, this desideratum should not equate a high degree of realism with a high degree of scientific value. Different models are useful for different purposes. At certain levels and for certain questions, abstract, simplifying models are precisely what is needed. Such a model will be more useful than a model slavishly realistic with respect to every level, even the biochemical. Excessive realism may entail that the model is too bedizened and rich to analyze or understand or even run on the computers available. For other questions, such as dendritic spine dynamics, the more realism at the biochemical level, for example, the better. But even here, the model will probably not be improved by taking into account quantum properties at the level below, or the cell's circuit cohorts on the level above. There is, of course, no decision procedure for the problem: how realistic should my model of X be, for many values of X ? Each case has to be thought out on its own, and solved with imagination and horse sense.

"Data rich, but theory poor" is a description frequently applied to neuroscience. In one obvious respect, this remains true, inasmuch as we do not yet know how to explain how brains see, learn, and take action. Nevertheless, theory in the form of computational modeling is rapidly catching up with the neurobiological data base. To be sure, there is still a great deal of experimental data that has not yet found a modeling home. Although the store is by no means exhausted, the modeling enterprise is slowed by gaps in our experimental knowledge, and these gaps need to be filled before extensive modeling can proceed. The experiments whose results are needed cannot be done by computer—they can be done only by anatomists, physiologists, biochemists, and geneticists working on real nervous tissue; by neuropsychologists studying patients with brain damage; and by psychologists studying normal humans and other animals. A lesson revealed by the modeling efforts is that there are many places where we are data poor, many questions that simply cannot be addressed in a computer model because the relevant data on which the model must rely are not yet available. Of course, assessments of wealth are essentially relative, both to where one *was*—in which case neuroscience is data rich and theory rich—and where one *wants to be*—in which case neuroscience is both data poor and theory poor.

The next few decades will be the formative years for computational neuroscience. Predicting what we shall understand of the brain by 2020 is, needless to say, a mug's game. Nevertheless, the hunch that exciting things are in store is difficult to subdue, and the thrill of discovering what we are and how we work is luring more and more students into the field. They often bring with them novel perspectives from their mother fields, as well as a bold inventiveness, a gift prized by a developing field that needs new ideas—unorthodox and otherwise. Withal, it is a remarkable time in the history of science.