How to Use Part I

Part I provides background material, summarizing a set of concepts established for the formal study of neurons and neural networks by 1986. As such, it is designed to hold few, if any, surprises for readers with a fair background in computational neuroscience or theoretical approaches to neural networks considered as dynamic, adaptive systems. Rather, Part I is designed for the many readers—be they neuroscience experimentalists, psychologists, philosophers, or technologists—who are sufficiently new to *brain theory and neural networks* that they can benefit from a compact overview of basic concepts prior to reading the road maps of Part II and the articles in Part III. Of course, much of what is covered in Part I is also covered at some length in the articles in Part III, and cross-references will steer the reader to these articles for alternative expositions and reviews of current research. In this exposition, as throughout the *Handbook*, we will move back and forth between *computational neuroscience*, where the emphasis is on modeling biological neurons, and *neural computing*, where the emphasis shifts back and forth between biological models and artificial neural networks based loosely on abstractions from biology, but driven more by technological utility than by biological considerations.

Section I.1, "Introducing the Neuron," conveys the basic properties of neurons, receptors, and effectors, and then introduces several simple neural models, including the discrete-time McCulloch-Pitts model and the continuous-time leaky integrator model. References to Part III alert the reader to more detailed properties of neurons which are essential for the neuroscientist and provide interesting hints about future design features for the technologist.

Section I.2, "Levels and Styles of Analysis," is designed to give the reader a feel for the interdisciplinary nexus in which the present study of brain theory and neural networks is located. The selection begins with a historical fragment which traces our federation of disciplines back to their roots in cybernetics, the study of control and communication in animals and machines. We look at the way in which the research addresses brains, machines, and minds, going back and forth between brain theory, artificial intelligence, and cognitive psychology. We then review the different levels of analysis involved, whether we study brains or intelligent machines, and the use of schemas to provide intermediate functional units that bridge the gap between an overall task and the neural networks which implement it.

Section I.3, "Dynamics and Adaptation in Neural Networks," provides a tutorial on the concepts essential for understanding neural networks as dynamic, adaptive systems. It introduces the basic dynamic systems concepts of stability, limit cycles, and chaos, and relates Hopfield nets to attractors and optimization. It then introduces a number of basic concepts concerning adaptation in neural nets, with discussions of pattern recognition, associative memory, Hebbian plasticity and network self-organization, perceptrons, network complexity, gradient descent and credit assignment, and backpropagation. This section, and with it Part I, closes with a cautionary note. The basic learning rules and adaptive architectures of neural networks have already illuminated a number of biological issues and led to useful technological applications. However, these networks must have their initial structure well constrained (whether by evolution or technological design) to yield approximate solutions to the system's tasks—solutions that can then be efficiently and efficaciously shaped by experience. Moreover, the full understanding of the brain and the improved design of intelligent machines will require not only improvements in these learning methods and their initialization, but also a fuller understanding of architectures based on networks of networks. Cross-references to articles in Part III will set the reader on the path to this fuller understanding. Because Part I focuses on the basic concepts established for the formal study of neurons and neural networks by 1986, it differs hardly at all from Part I of the first edition of the *Handbook*. By contrast, Part II, which provides the road maps that guide readers through the radically updated Part III, has been completely rewritten for the present edition to reflect the latest research results.

I.1. Introducing the Neuron

We introduce the *neuron*. The dangerous word in the preceding sentence is *the*. In biology, there are radically different types of neurons in the human brain, and endless variations in neuron types of other species. In brain theory, the complexities of real neurons are abstracted in many ways to aid in understanding different aspects of neural network development, learning, or function. In *neural computing* (technology based on networks of "neuron-like" units), the artificial neurons are designed as variations on the abstractions of brain theory and are implemented in software, or VLSI or other media. There is no such thing as a "typical" neuron, yet this section will nonetheless present examples and models which provide a starting point, an essential set of key concepts, for the appreciation of the many variations on the theme of neurons and neural networks presented in Part III.

An analogy to the problem we face here might be to define *vehicle* for a handbook of transportation. A vehicle could be a car, a train, a plane, a rowboat, or a forklift truck. It might or might not carry people. The people could be crew or passengers, and so on. The problem would be to give a few key examples of form (such as car versus plane) and function (to carry people or goods, by land, air, or sea, etc.). Moreover, we would find interesting examples of co-evolution: for example, modern highway systems would not have been created without the pressure of increasing car traffic; most features of cars are adapted to the existence of sealed roads, and some features (e.g., cruise control) are specifically adapted to good freeway conditions. Following a similar procedure, Part III offers diverse examples of neural form and function in both biology and technology.

Here, we start with the observation that a brain is made up of a network of cells called neurons, coupled to receptors and effectors. Neurons are intimately connected with glial cells, which provide support functions for neural networks. New empirical data show the importance of glia in regeneration of neural networks after damage and in maintaining the neurochemical milieu during normal operation. However, such data have had very little impact on neural modeling and so will not be considered further here. The input to the network of neurons is provided by *receptors*, which continually monitor changes in the external and internal environment. Cells called *motor neurons* (or *motoneurons*), governed by the activity of the neural network, control the movement of muscles and the secretion of glands. In between, an intricate network of neurons (a few hundred neurons in some simple creatures, hundreds of billions in a human brain) continually combines the signals from the receptors with signals encoding past experience to barrage the motor

neurons with signals that will yield adaptive interactions with the environment. In animals with backbones (vertebrates, including mammals in general and humans in particular), this network is called the *central nervous system* (CNS), and the brain constitutes the most headward part of this system, linked to the receptors and effectors of the body via the spinal cord. Invertebrate nervous systems (neural networks) provide astounding variations on the vertebrate theme, thanks to eons of divergent evolution. Thus, while the human brain may be the source of rich analogies for technologists in search of "artificial intelligence," both invertebrates and vertebrates provide endless ideas for technologists designing neural networks for sensory processing, robot control, and a host of other applications. (A few of the relevant examples may be found in the Part II road maps, **Vision**, **Robotics and Control Theory**, **Motor Pattern Generators**, and **Neuroethology and Evolution**.)

The brain provides far more than a simple stimulus-response chain from receptors to effectors (although there are such reflex paths). Rather, the vast network of neurons is interconnected in loops and tangled skeins so that signals entering the net from the receptors interact there with the billions of signals already traversing the system, not only to yield the signals that control the effectors but also to modify the very properties of the network itself, so that future behavior will reflect prior experience.

The Diversity of Receptors

Rod and cone receptors in the eyes respond to light, hair cells in the ears respond to pressure, and other cells in the tongue and the mouth respond to subtle traces of chemicals. In addition to touch receptors, there are receptors in the skin that are responsive to movement or to temperature, or that signal painful stimuli. These external senses may be divided into two classes: (1) the proximity senses, such as touch and taste, which sense objects in contact with the body surface, and (2) the distance senses, such as vision and hearing, which let us sense objects distant from the body. Olfaction is somewhere in between, using chemical signals "right under our noses" to sense nonproximate objects. Moreover, even the proximate senses can yield information about nonproximate objects, as when we feel the wind or the heat of a fire. More generally, much of our appreciation of the world around us rests on the unconscious fusion of data from diverse sensory systems.

The appropriate activity of the effectors must depend on comparing where the system should be—the current target of an ongoing movement—with where it is now. Thus, in addition to the external receptors, there are receptors that monitor the activity of muscles, tendons, and joints to provide a continual source of feedback about the tensions and lengths of muscles and the angles of the joints, as well as their velocities. The vestibular system in the head monitors gravity and accelerations. Here, the receptors are hair cells monitoring fluid motion. There are also receptors to monitor the chemical level of the bloodstream and the state of the heart and the intestines. Cells in the liver monitor glucose, while others in the kidney check water balance. Receptors in the *hypothalamus*, itself a part of the brain, also check the balance of water and sugar. The hypothalamus then integrates these diverse messages to direct behavior or other organs to restore the balance. If we stimulate the hypothalamus, an animal may drink copious quantities of water or eat enormous quantities of food, even though it is already well supplied; the brain has received a signal that water or food is lacking, and so it instructs the animal accordingly, irrespective of whatever contradictory signals may be coming from a distended stomach.

Basic Properties of Neurons

To understand the processes that intervene between receptors and effectors, we must have a closer look at "the" neuron. As already emphasized, *there is no such thing as a typical neuron*. However, we will summarize properties shared by many neurons. The "basic neuron" shown in Figure 1 is abstracted from a motor neuron of mammalian spinal cord. From the *soma* (cell body) protrudes a number of ramifying branches called *dendrites*; the soma and dendrites constitute the input surface of the neuron. There also extrudes from the cell body, at a point called the *axon hillock* (abutting the initial segment), a long fiber called the *axon*, whose branches form the *axonal arborization*. The tips of the branches of the axon, called *nerve terminals* or *boutons*, impinge on other neurons or on effectors. The locus of interaction between a bouton and the cell on which it impinges is called a *synapse*, and we say that the cell with the bouton *synapses upon* the cell with which the connection is made. In fact, axonal branches of some neurons can have many varicosities, corresponding to synapses, along their length, not just at the end of the branch.

We can imagine the flow of information as shown by the arrows in Figure 1. Although "conduction" can go in either direction on the axon, most synapses tend to "communicate" activity to the dendrites or soma of the cell they synapse upon, whence activity passes to the axon hillock and then down the axon to the terminal arbo-

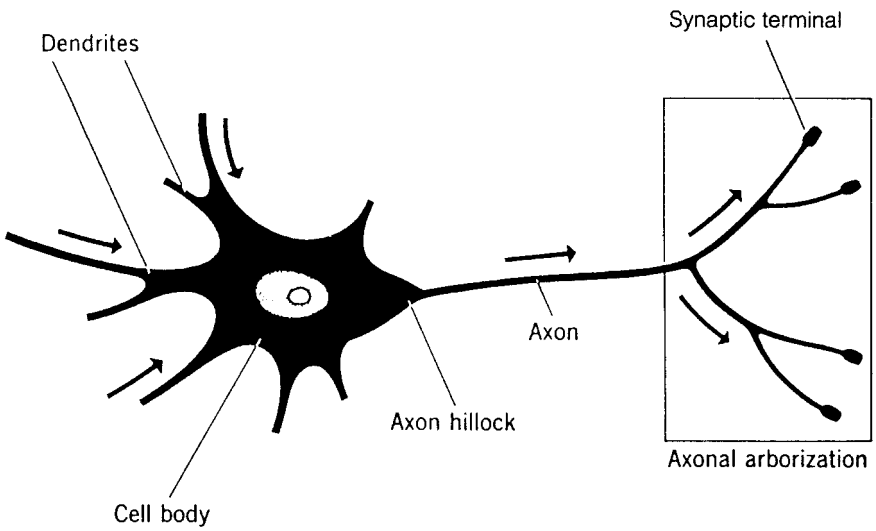

Figure 1. A "basic neuron" abstracted from a motor neuron of mammalian spinal cord. The dendrites and soma (cell body) constitute the major part of the input surface of the neuron. The axon is the "output line." The tips of the branches of the axon form synapses upon other neurons or upon effectors (although synapses may occur along the branches of an axon as well as at the ends). (From Arbib, M. A., 1989, *The Metaphorical Brain 2: Neural Networks and Beyond*, New York: Wiley-Interscience, p. 52. Reproduced with permissions. Copyright © 1989 by John Wiley & Sons, Inc.)

rization. The axon can be very long indeed. For instance, the cell body of a neuron that controls the big toe lies in the spinal cord and thus has an axon that runs the complete length of the leg. We may contrast the immense length of the axon of such a neuron with the very small size of many of the neurons in our heads. For example, amacrine cells in the retina have branchings that cannot appropriately be labeled dendrites or axons, for they are short and may well communicate activity in either direction to serve as local modulators of the surrounding network. In fact, the propagation of signals in the "counter-direction" on dendrites away from the soma has in recent years been seen to play an important role in neuronal function, but this feature is not included in the account of the "basic neuron" given here (see DENDRITIC PROCESSING—titles in SMALL CAPS refer to articles in Part III).

To understand more about neuronal "communication," we emphasize that the cell is enclosed by a membrane, across which there is a difference in electrical charge. If we change this potential difference between the inside and outside, the change can propagate in much the same passive way that heat is conducted down a rod of metal: a normal change in potential difference across the cell membrane can propagate in a passive way so that the change occurs later, and becomes smaller, the farther away we move from the site of the original change. This passive propagation is governed by the *cable equation*

$$\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}$$

If the starting voltage at a point on the axon is V_0, and no further conditions are imposed, the potential will decay exponentially, having value $V_{(x)} = V_0 e^{-x}$ at distance x from the starting point, where the length unit, the *length constant*, is the distance in which the potential changes by a factor of $1/e$. This length unit will differ from axon to axon. For "short" cells (such as the rods, cones, and bipolar cells of the retina), passive propagation suffices to signal a potential change from one end to the other; but if the axon is long, this mechanism is completely inadequate, since changes at one end will decay almost completely before reaching the other end. Fortunately, most nerve cells have the further property that if the change in potential difference is large enough (we say it exceeds a *threshold*), then in a cylindrical configuration such as the axon, a pulse can be generated that will actively propagate at full amplitude instead of fading passively.

If propagation of various potential differences on the dendrites and soma of a neuron yields a potential difference across the membrane at the axon hillock which exceeds a certain threshold, then a regenerative process is started: the electrical change at one place is enough to trigger this process at the next place, yielding a *spike* or *action potential*, an undiminishing pulse of potential difference propagating down the axon. After an impulse has propagated along the length of the axon, there is a short *refractory period* during which a new impulse cannot be propagated along the axon.

The propagation of action potentials is now very well understood. Briefly, the change in membrane potential is mediated by the flow of ions, especially sodium and potassium, across the membrane. Hodgkin and Huxley (1952) showed that the *conductance* of the membrane to sodium and potassium ions—the ease with which they flow across the membrane—depends on the transmembrane voltage. They developed elegant equations describing the voltage and time dependence of the sodium and potassium conductances. These equations (see the article AXONAL MODELING in Part III) have given us great insight into cellular function. Much mathematical research has gone into studying Hodgkin-Huxley-like equations, showing, for example, that neurons can support rhythmic pulse generation even without input (see OSCILLATORY AND BURSTING PROPERTIES OF NEURONS), and explicating triggered long-distance propagation. Hodgkin and Huxley used curve fitting from experimental data to determine the terms for conductance change in their model. Subsequently, much research has probed the structure of complex molecules that form *channels* which selectively allow the passage of specific ions through the membrane (see ION CHANNELS: KEYS TO NEURONAL SPECIALIZATION). This research has demonstrated how channel properties not only account for the terms in the Hodgkin-Huxley equation, but also underlie more complex dynamics which may allow even small patches of neural membrane to act like complex computing elements. At present, most artificial neurons used in applications are very simple indeed, and much future technology will exploit these "subneural subtleties."

An impulse traveling along the axon from the axon hillock triggers new impulses in each of its branches (or *collaterals*), which in turn trigger impulses in their even finer branches. Vertebrate axons come in two varieties, myelinated and unmyelinated. The myelinated fibers are wrapped in a sheath of *myelin* (Schwann cells in the periphery, oligodendrocytes in the CNS—these are glial cells, and their role in axonal conduction is the primary role of glia considered in neural modeling to date). The small gaps between successive segments of the myelin sheath are called *nodes of Ranvier*. Instead of the somewhat slow active propagation down an unmyelinated fiber, the nerve impulse in a myelinated fiber jumps from node to node, thus speeding passage and reducing energy requirements (see AXONAL MODELING).

Surprisingly, at most synapses, the direct cause of the change in potential of the postsynaptic membrane is not electrical but chemical. When an impulse arrives at the presynaptic terminal, it causes the release of *transmitter* molecules (which have been stored in the bouton in little packets called vesicles) through the presynaptic membrane. The transmitter then diffuses across the very small *synaptic cleft* to the other side, where it binds to receptors on the postsynaptic membrane to change the conductance of the postsynaptic cell. The effect of the "classical" transmitters (later we shall talk of other kinds, the neuromodulators) is of two basic kinds: either *excitatory*, tending to move the potential difference across the postsynaptic membrane in the direction of the threshold (*depolarizing* the membrane), or *inhibitory*, tending to move the polarity away from the threshold (*hyperpolarizing* the membrane). There are some exceptional cell appositions that are so large or have such tight coupling (the so-called gap junctions) that the impulse affects the postsynaptic membrane without chemical mediation (see NEOCORTEX: CHEMICAL AND ELECTRICAL SYNAPSES).

Most neural modeling to date focuses on the excitatory and inhibitory interactions that occur on a fast time scale (a millisecond, more or less), and most biological (as distinct from technological) models assume that all synapses *from* a neuron have the same "sign." However, neurons may also secrete transmitters that modulate the function of a circuit on some quite extended time scale. Modeling that takes account of this *neuromodulation* (see SYNAPTIC INTERACTIONS and NEUROMODULATION IN INVERTEBRATE NERVOUS SYSTEMS) will become increasingly important in the future, since it allows cells to change their function, enabling a neural network to switch dramatically its overall mode of activity.

The excitatory or inhibitory effect of the transmitter released when an impulse arrives at a bouton generally causes a subthreshold change in the postsynaptic membrane. Nonetheless, the cooperative effect of many such subthreshold changes may yield a potential change at the axon hillock that exceeds threshold, and if this occurs at a time when the axon has passed the refractory period of its previous firing, then a new impulse will be fired down the axon.

Synapses can differ in shape, size, form, and effectiveness. The geometrical relationships between the different synapses impinging on the cell determine what patterns of synaptic activation will yield the appropriate temporal relationships to excite the cell (see

DENDRITIC PROCESSING). A highly simplified example (Figure 2) shows how the properties of nervous tissue just presented would indeed allow a simple neuron, by its very dendritic geometry, to compute some useful function (cf. Rall, 1964, p. 90). Consider a neuron with four dendrites, each receiving a single synapse from a visual receptor, so arranged that synapses A, B, C, and D (from left to right) are at increasing distances from the axon hillock. (This is not meant to be a model of a neuron in the retina of an actual organism; rather, it is designed to make vivid the potential richness of single neuron computations.) We assume that each receptor reacts to the passage of a spot of light above its surface by yielding a generator potential which yields, in the postsynaptic membrane, the same time course of depolarization. This time course is propagated passively, and the farther it is propagated, the later and the lower is its peak. If four inputs reached A, B, C, and D simultaneously, their effect may be less than the threshold required to trigger a spike there. However, if an input reaches D before one reaches C, and so on, in such a way that the peaks of the four resultant time courses at the axon hillock coincide, the total effect could well exceed threshold. This, then, is a cell that, although very

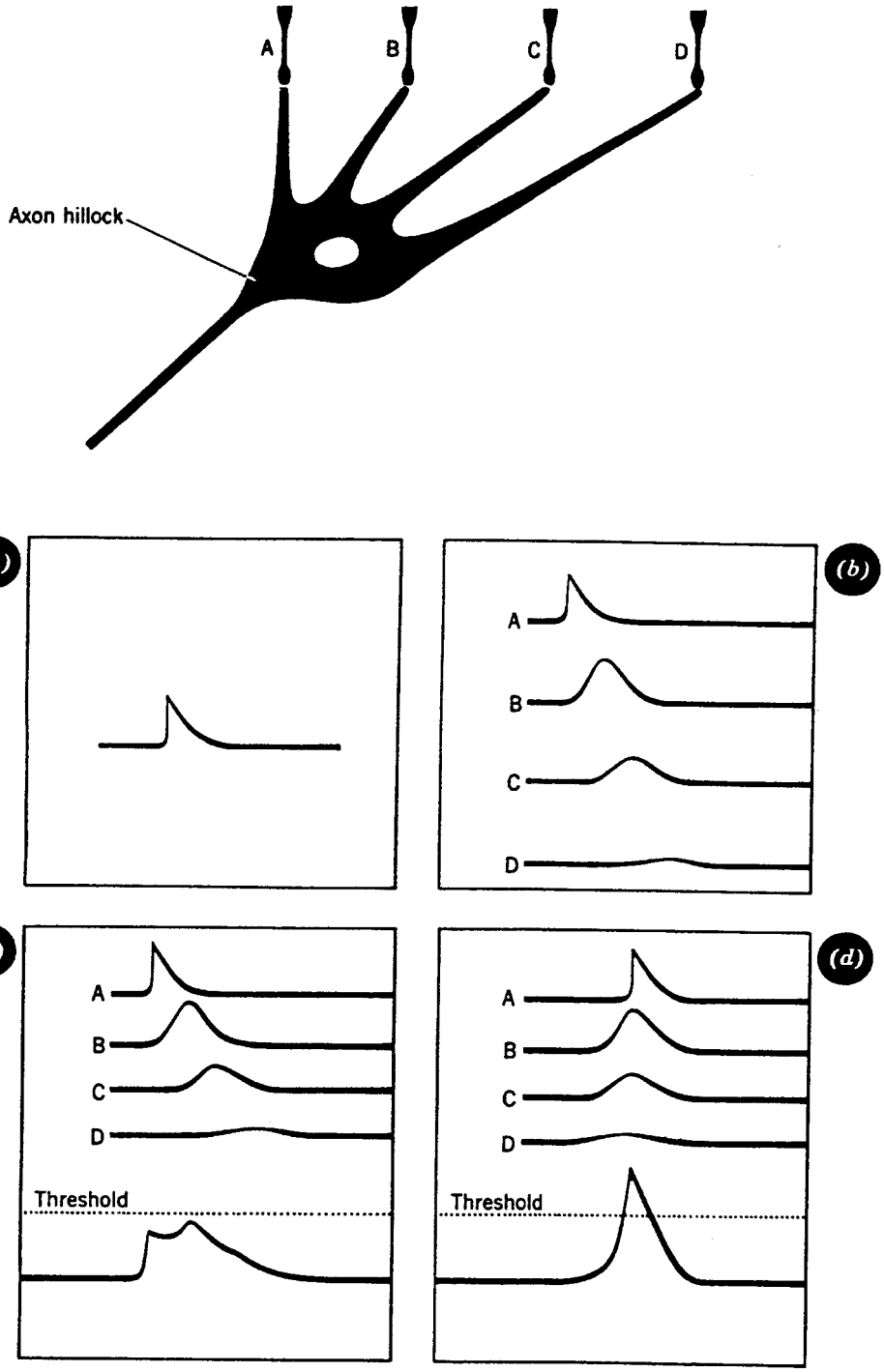

Figure 2. An example, conceived by Wilfrid Rall, of the subtleties that can be revealed by neural modeling when dendritic properties (in this case, length-dependent conduction time) are taken into account. As shown in Part *C*, the effect of simultaneously activating all inputs may be subthreshold, yet the cell may respond when inputs traverse the cell from right to left (*D*). (From Arbib, M. A., 1989, *The Metaphorical Brain 2: Neural Networks and Beyond*, New York: Wiley-Interscience, p. 60. Reproduced with permission. Copyright © 1989 by John Wiley & Sons, Inc.)

simple, can detect direction of motion across its input. It responds only if the spot of light is moving from right to left, and if the velocity of that motion falls within certain limits. Our cell will not respond to a stationary object, or one moving from left to right, because the asymmetry of placement of the dendrites on the cell body yields a preference for one direction of motion over others (for a more realistic account of biological mechanisms, see DIRECTIONAL SELECTIVITY). This simple example illustrates that the *form* (i.e., the geometry) of the cell can have a great impact on the *function* of the cell, and we thus speak of *form-function* relations. When we note that neurons in the human brain may have 10,000 or more synapses upon them, we can understand that the range of functions of single neurons is indeed immense.

Receptors and Effectors

On the "input side," receptors share with neurons the property of generating potentials, which are transmitted to various synapses upon neurons. However, the input surface of a receptor does not receive synapses from other neurons, but can transduce environmental energy into changes in membrane potential, which may then propagate either actively or passively. (Visual receptors do not generate spikes; touch receptors in the body and limbs use spike trains to send their message to the spinal cord.) For instance, the rods and cones of the eye contain various pigments that react chemically to light in different frequency bands, and these chemical reactions, in turn, lead to local potential changes, called generator potentials, in the membrane. If the light falling on an array of rods and cones is appropriately patterned, then their potential changes will induce interneuron changes to, in turn, fire certain ganglion cells (retinal output neurons whose axons course toward the brain). Properties of the light pattern will thus be signaled farther into the nervous system as trains of impulses (see RETINA).

At the receptors, increasing the intensity of stimulation will increase the generator potential. If we go to the first level of neurons that generate pulses, the axons "reset" each time they fire a pulse and then have to get back to a state where the threshold and the input potential meet. The higher the generator potential, the shorter the time until they meet again, and thus the higher the frequency of the pulse. Thus, at the "input" it is a useful first approximation to say that intensity or quantity of stimulation is coded in terms of pulse frequency (more stimulus $\approx$ more spikes), whereas the quality or type of stimulus is coded by different lines carrying signals from different types of receptors. As we leave the periphery and move toward more "computational" cells, we no longer have such simple relationships, but rather interactions of inhibitory cells and excitatory cells, with each inhibitory input moving a cell away from, and each excitatory input moving it toward, threshold.

To discuss the "output side," we must first note that a muscle is made up of many thousands of muscle fibers. The motor neurons that control the muscle fibers lie in the spinal cord or the brainstem, whence their axons may have to travel vast distances (by neuronal standards) before synapsing upon the muscle fibers. The smallest functional entity on the output side is thus the *motor unit*, which consists of a motor neuron cell body, its axon, and the group of muscle fibers the axon influences.

A muscle fiber is like a neuron to the extent that it receives its input via a synapse from a motor neuron. However, the response of the muscle fiber to the spread of depolarization is to contract. Thus, the motor neurons which synapse upon the muscle fibers can determine, by the pattern of their impulses, the extent to which the whole muscle comprised of those fibers contracts, and can thus control movement. (Similar remarks apply to those cells that secrete various chemicals into the bloodstream or gut, or those that secrete sweat or tears.)

Synaptic activation at the *motor end-plate* (i.e., the synapse of a motor neuron upon a muscle fiber) yields a brief "twitch" of the muscle fiber. A low repetition rate of action potentials arriving at a motor end-plate causes a train of twitches, in each of which the mechanical response lasts longer than the action potential stimulus. As the frequency of excitation increases, a second action potential will arrive while the mechanical effect of the prior stimulus still persists. This causes a mechanical summation or fusion of contractions. Up to a point, the degree of summation increases as the stimulus interval becomes shorter, although the summation effect decreases as the interval between the stimuli approaches the refractory period of the muscle, and maximum tension occurs. This limiting response is called a *tetanus*. To increase the tension exerted by a muscle, it is then necessary to recruit more and more fibers to contract. For more delicate motions, such as those involving the fingers of primates, each motor neuron may control only a few muscle fibers. In other locations, such as the shoulder, one motor neuron alone may control thousands of muscle fibers. As descending signals in the spinal cord command a muscle to contract more and more, they do this by causing motor neurons with larger and larger thresholds to start firing. The result is that fairly small fibers are brought in first, and then larger and larger fibers are recruited. The result, known as Henneman's Size Principle, is that at any stage, the increment of activation obtained by recruiting the next group of motor units involves about the same percentage of extra force being applied, aiding smoothness of movement (see MOTONEURON RECRUITMENT).

Since there is no command that a neuron may send to a muscle fiber that will cause it to lengthen—all the neuron can do is stop sending it commands to contract—the muscles of an animal are usually arranged in pairs. The contraction of one member of the pair will then act around a pivot to cause the expansion of the other member of the pair. Thus, one set of muscles *extends* the elbow joint, while another set *flexes* the elbow joint. To extend the elbow joint, we do not signal the *flexors* to lengthen, we just stop signaling them to contract, and then they will be automatically lengthened as the *extensor* muscles contract. For convenience, we often label one set of muscles as the "prime mover" or *agonist*, and the opposing set as the *antagonist*. However, in such joints as the shoulder, which are not limited to one degree of freedom, many muscles, rather than an agonist-antagonist pair, participate. Most real movements involve many joints. For example, the wrist must be fixed, holding the hand in a position bent backward with respect to the forearm, for the hand to grip with its maximum power. *Synergists* are muscles that act together with the main muscles involved. A large group of muscles work together when one raises something with one's finger. If more force is required, wrist muscles may also be called in; if still more force is required, arm muscles may be used. In any case, muscles all over the body are involved in maintaining posture.

Neural Models

Before presenting more realistic models of the neuron (see PERSPECTIVE ON NEURON MODEL COMPLEXITY; SINGLE-CELL MODELS), we focus on the work of McCulloch and Pitts (1943), which combined neurophysiology and mathematical logic, using the all-or-none property of neuron firing to model the neuron as a binary discrete-time element. They showed how excitation, inhibition, and threshold might be used to construct a wide variety of "neurons." It was the first model to tie the study of neural nets squarely to the idea of computation in its modern sense. The basic idea is to divide time into units comparable to a refractory period so that, in each time period, at most one spike can be generated at the axon hillock of a given neuron. The McCulloch-Pitts neuron (Figure 3*A*) thus operates on a discrete-time scale, $t = 0, 1, 2, 3, \ldots$, where the