Chapter 1
Introduction

We cannot know exactly when it happened, but some time after the dawn of humankind our hominoid ancestors must have wondered— as we do—about the mystery of their inner world. The idea of a body-inhabiting spirit that animates a lifeless form with feeling and cunning, that reveals the myriad elements of the world around, is a strong and recurring theme in myth. Throughout human history there has been an abiding need to comprehend this primal mystery. Plato proposed a rational soul located in the head as the source of all our mental functions. However inadequate Plato's view might be as an explanation of mind, who can doubt that however it happens, it happens in secret places behind the eyes and between the ears in that quirky-shaped lump of tissue.

Since the late 1800s, the principal effort to understand the human mind has shifted from the domain of the philosopher to that of the psychologist, who carries on this exploration as a scientific task (Boring, 1929). More recently, groups within disciplines ranging from linguistics to computer science have focused on the problem of cognition (mindlike activity) and call their joint endeavor "cognitive science." Among cognitive scientists, however, the study of human cognition is chiefly the concern of the cognitive psychologist.

The cognitive science community at large commonly asserts that neuroscience has little to contribute to cognitive theory because it has not yet produced a substantial body of knowledge relevant to cognitive processes. Yet to the extent that cognitive science can provide a valid explanation of how human cognition is accomplished, it must conform to the limiting as well as the enabling properties of the human brain. No account can be a serious explanatory model of human cognition if it cannot pass this test. My central thesis here is that we now have sufficient knowledge of the physiology of nerve cells and the structure of the brain to advance the theoretical formulation of putative brain mechanisms that can account for the basic competencies of human cognition and, at the same time, satisfy a reasonable demand for plausibility within the constraints of neurophysiology

and neuroanatomy. This is not to say that the models we construct will necessarily allow direct verification or disconfirmation by the current tools of neuroscience but that they can be assessed as being consistent with known biological processes, as opposed to contradicting established findings or demanding new physiological properties that strain current understanding.

Of course, it might be argued that if no model can be found to account for the important facts of human cognition without straining the credulity of neuroscientists, then perhaps neuroscience is not keen enough, has misinterpreted its own evidence, or just needs to look further. In any event, should there be a fit between the demands of a cognitive model and the constraints of neuroscience, then our confidence in at least the cognitive science side of the endeavor will grow, and such a model might suggest new foci for direct exploration in the vast physical search space confronting neuroscientists.

Discourse and debate often founder on unacknowledged differences in belief about (or sense of) the meaning of terms common among the discussants. This is particularly true in consideration of "cognitive processes" because of their relatively cryptic and complex nature. I must, therefore, acknowledge at the outset that the models I present do not aim to explain directly such ineffable matters as the felt qualities of a breathtaking sunset or why I *believe* my *intention* to open the door causes me to do so. They do, however, attempt to explain how a range of cognitive tasks that are of fundamental importance can be accomplished by the biological machinery of the human brain, and it is not unreasonable to assume that they might well offer useful clues to the more subtle aspects of human experience.

The theoretical strategy that guides this work is constructionist in flavor. The belief that neuroscience has not provided enough of the right kind of facts about the brain to be of much help to the cognitive scientist is true in a limited sense. That is, to the extent that the cognitive scientist frames proposed models only in terms of interactions among functional stages or "black box" computational modules, one finds that neuroscience is as yet unable to elucidate the micropatterns of intercellular connectivity or the dynamics of information processing in the brain with sufficient scope and precision to provide evidence either for or against proposals at this level of explanation.

My approach is to start with first principles. Beginning with the principal cognitive tasks that are unquestionably solved by humans and those properties of nerve cells that are relevant to their role as primitive components in a variety of possible information-processing mechanisms, the problem is to construct model neuronal mechanisms and systems with the minimal properties required to perform

the selected tasks (Trehub 1975a, 1977, 1978, 1983, 1987). In the neuronal-constructionist approach, the initial constraints of neurophysiology and neuroanatomy are not merely suggestive in guiding model development, they are crucial, and their application distinguishes a neuronal model from other kinds of connectionist models that are said to have a neural "flavor" and are often called neural models (Anderson et al. 1977, Feldman 1985, Feldman and Ballard 1982, Grossberg 1988, Hopfield and Tank 1986, Kohonen 1977, Rosenblatt 1962, Rumelhart and McClelland 1986; see also Smolensky 1988).

Widespread adoption of a computational information-processing paradigm since the mid-1970s has resulted in a surge of experimentation and theory in cognitive psychology. But while the new computational enthusiasm sparked a wealth of empirical studies and algorithmic models, it failed to stimulate a comparable intellectual probing into the kinds of biological mechanisms required to carry out the cognitive computations. In short, although it was suggested how abstract computations could account for some natural observations and various experimental results, it was not explained how the biological person could carry out the necessary cognitive computations. I take the latter explanatory burden to be at least as much the responsibility of the cognitive psychologist as it is of the neuroscientist. Other cognitive scientists whose primary interests might relate to the properties of abstract automata or machine instantiations of algorithms in the cognitive domain do not share this burden.

Levels of Explanation

David Marr (1982) proposed and laid great stress on the independence of three levels of explanation at which any information-processing machine must be understood: (1) the level of computational theory at which the computational goal and its logic are formulated, (2) the level of representation and algorithm at which the representations for input and output and the algorithm for transformation are formulated, and (3) the level of hardware (or biological) implementation at which the physical means for realizing the representations and algorithms are formulated. Marr has had a notably strong influence on the field of cognitive psychology, and the psychologist might say in response to a charge of neglecting the brain that the computational theories and algorithmic models should be all worked out before worrying about biological implementation. Indeed, Rumelhart and McClelland (1985) write, "However, Marr certainly does not propose that a theory at the computational level of

description is an adequate psychological theory. As psychologists, we are committed to an elucidation of the algorithmic level." Yet they also assert, "We believe that psychology is properly concerned with all three of these levels." And in a later publication McClelland, Rumelhart, and Hinton (1986) state that one reason for the appeal of parallel distributed processing (PDP) models is their "'physiological' flavor."

It appears that while cognitive psychology is not prepared to dismiss the physiology of the brain as irrelevant, it is committed to what it perceives as an orderly development of theory in the spirit of Marr's independence of explanatory levels: first develop models that provide understanding at the levels of computational and algorithmic theory and then (perhaps) address the issue of their implementation in the human brain. I believe, however, that a less hierarchically ordered strategy of theory development can be more productive in achieving the goal of understanding human cognition, and I believe that the failure of psychology to acknowledge explicitly the critical role of biological constraints on human information processing is in part responsible for the frustratingly slow incremental development of a coherent body of theory and knowledge (Newell 1973). Adoption of the independent-levels paradigm provides a principled basis for many more candidate models (and the empirical studies they suggest) than would otherwise be justified and tends to obscure the distinction between those proposals that are consonant with biology and those that are not. An optimal search strategy for explanatory models of human cognition calls for the virtual parallel application of multiple-level constraints instead of a search for "sufficient" models within the bounds of each assumed "independent" level of explanation.

Figure 1.1 illustrates the difference between an independent-levels search (top) and an interdependent-levels search (bottom). The first circle at the top of the illustration represents the domain of logically possible computational objectives that might lead to a solution for the problem. The right circle is the domain of possible and distinct algorithmic operations, subroutines, and full routines that can satisfy each adequate set of computational objectives. The third circle is the domain of plausible computational mechanisms in the human brain. In this scheme, the algorithmic level is the most troublesome because it is essentially unbounded. Since algorithms are wholly symbolic systems, governed only by formal rules, any specified set of algorithms can be emulated by a different set with equivalent computational effect. If we try to reduce the size of the domain of possible algorithms by simply applying a selection rule that accepts only algo-
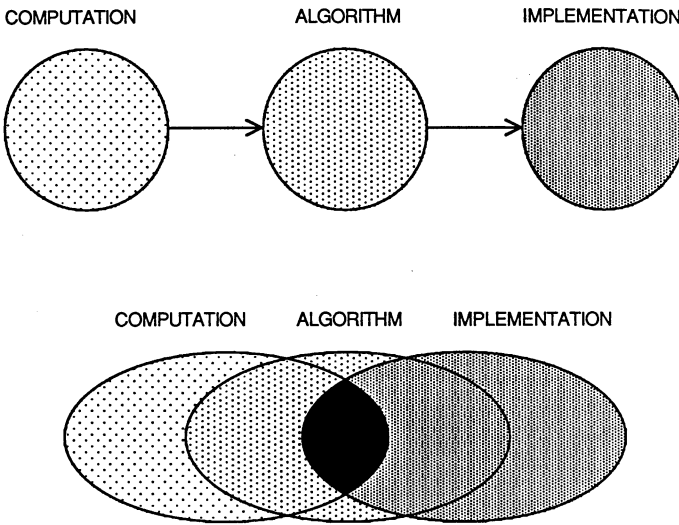
COMPUTATION   ALGORITHM   IMPLEMENTATION

COMPUTATION   ALGORITHM   IMPLEMENTATION

Figure 1.1
Top: An independent-levels search in theory development. *Bottom:* An interde-pendent-levels search.

rithms with the fewest number of computational steps, we face two problems; the number of steps required for a given algorithm may be quite different depending on its biological implementation, and, more important, any chosen algorithm may be composed of elemen-tary operations or subroutines that the brain does not compute.

A better way to reduce the size of the domain is to prune, during the course of composing an algorithm, each assumption that is not credible on the evidence of neurophysiology and neuroanatomy. This violates the neat separation of explanatory levels and is, in fact, the strategy for theory development represented by the interdependent-levels search shown at the bottom of figure 1.1. In this paradigm, search is always confined to the much smaller domain established by the parallel constraints of all three levels. For expository or didactic purposes, any adequate model resulting from this approach can be analytically decomposed into the three independent levels of descrip-tion even though their logical independence was not honored during the construction of the model.

Competence and Style

Few psychologists would disagree that the central theoretical goal of cognitive psychology is to explain how humans are able to know and

understand their physical-social environment and act productively and creatively. But instead of attempting to provide an explanation for cognitive competence, psychologists simply assume that competent cognitive mechanisms exist and model them in terms of abstract nodes, modules, and processing stages with appropriate functional labels (canonical feature detector, window of attention) attached (see for example Biederman 1987, Kosslyn 1980, Kosslyn et al. 1990). At the same time, their principal theoretical effort is directed at determining the combination of properties (e.g., serial operations, parallel operations) in an unexplicated (nominally empowered) central cognitive apparatus that might force its output to conform with the curves of performance obtained in experimental studies. This prevailing approach cannot explain basic competence, but it can explain cognitive style and provide empirical evidence useful in assessing explicit models of competent brain mechanisms.

A proper model of a competent mechanism has four important characteristics: (1) it is described in sufficient detail to establish its biological plausibility; (2) logical analysis and computer simulation demonstrate that it does perform the task(s) required of it; (3) its biological instantiation can reasonably be expected to accomplish the task(s) within a time period commensurate with the normal ecological demands on the organism modeled (Feldman 1981, Trehub 1987); and (4) its interface with other mechanisms in its system is biologically plausible. Among models that have already demonstrated basic explanatory competence, secondary aspects, such as parsimony and style of performance, are also considered. For example, if two competing models of associative learning are judged equally competent and parsimonious but display different acquisition curves, the model with performance that better matches the data of relevant laboratory experiments should be preferred. In addition, the ability of a model to explain cognitive epiphenomena should be considered in choosing among models equally competent on primary tasks. If the explanation of a known visual illusion, for example, were to "fall out" of a model formulated to account for other phenomena, we would take this as additional evidence in its favor.

On these grounds, then, I have chosen initially to explore models that might provide solutions to some of the fundamental tasks for the human cognitive brain. If, after basic competence can be demonstrated, it should turn out that the behavior of a given putative mechanism or system of mechanisms also conforms to experimental findings in the domain of behavioral style and cognitive epiphenomena, then we will have increased confidence that our theory building is on the right track.

Issues in the Design of a Cognitive Model

Any investigator attempting to formulate a model of human cognition must make initial design choices about scope and process. As he or she confronts the demand for competence over an increasing variety and complexity of tasks, these basic choices will necessarily be augmented. An evolving model architecture will include modular components that were developed initially and at the same time will be constrained by the physical principles instantiated in these earlier modules. Thus, the generality, competence, and operating characteristics of an articulated model will be strongly influenced by basic design choices (in practice, often implicit and unexamined) made at the earliest stages of model building. Some of these issues are particularly important in planning and analyzing the design of a cognitive system.

System Input-Output

Is the input to the model system to be the low-level output of a selected sensory modality (visual, auditory, somasthetic) or composed of the preprocessed categories that are assumed to have been extracted from a modality in some unexplicated fashion? Existing semantic models, for example, typically assume a high-level network of abstract nodes or else vectors of activation levels over an assembly of units, where each node or vector represents a word or concept in an established lexicon (Anderson and Bower 1973, Fahlman 1979, Hinton 1981). It is highly unlikely that any single model of human cognition can be comprehensive, and simplifications of this kind are inevitable. But where tactical simplification evades a fundamental cognitive problem, the fact of the problem should be acknowledged and some indication of its possible solution given.

The neuronal models that I propose, in addition to addressing issues of high-level cognition, attempt to deal with the important computations required to interpret the first level of information extracted by the sensory transducers from complex external and internal environments. A fundamental problem common to all modalities centers on the critical process of pattern classification (Harnad 1987, Margolis 1987). In the models presented in the following chapters, patterns are defined as the distribution of cell discharge over physically indexed arrays of neurons. The signaling of a particular distribution of activity in an array by the discharge of a single physically indexed cell in a detection set (or by N equivalent cells) constitutes pattern classification. All of the axonal branches of any given detection cell, taken together, may be thought of as a single discrete

information channel or line. Thus, pattern classification is an array-to-line mapping.

If each possible sensory pattern were mapped to a distinct line, the number of lines required would grow in a combinatorial explosion as the number of cells in the sensory array increased. Such a scheme would quickly exceed the capacity of even the human brain despite its many billions of neurons. On the other hand, if the system could not map patterns having very small arbitrary differences onto differ-ent lines, important environmental distinctions might escape detec-tion. For some tasks, it might be advantageous or simply innocuous to lump many physically different inputs into a single category (gen-eralization); for others, a failure to make a fine distinction (discrimina-tion) might carry a serious penalty. Thus, a fundamental design issue for any model of human cognition centers on the capacity of the system to adjust the delicacy of its sensory classifications according to the ecological requirements of its host.

A number of connectionist models are designed to map input vec-tors onto output vectors by processes of iterative relaxation involving widespread positive and negative feedback among cells or by direct feed-forward algorithms, which do not depend on relaxation (Rumel-hart and McClelland 1986). These models are among a class of com-putational schemes that transform a set of input patterns into a smaller set of patterns represented by differing activity distributions over a fixed population of cells constituting the output of the system. Models of this kind, commonly called parallel distributed processing (PDP) models, are said to have distributed representation as opposed to local representation, as is the case when an input pattern is identi-fied by activity on an indexed line.

From a strictly computational point of view, it might appear that there is an advantage to mapping from array to pattern instead of from array to line. For any fixed number N of output units (or lines), there will be many more possible discrete output states for the pat-terns composed by the activity levels of all N units than for the activ-ity of only one of N units. In the case of a whole pattern, the potential number of distinct outputs increases as a combinatorial function of N; in the case of a single line, the potential number of distinct outputs increases only as a linear function of N. Thus, it might seem that an array-to-pattern model would enjoy the economy of fewer units required to make any given number of classifications when compared to an array-to-line model. But there are other factors to take into account.

First, can the algorithms that map input patterns onto output pat-terns utilize the full potential range of output states? Simulation stud-

ies have shown that the proportion of N possible output vectors that can be used in the array-to-pattern paradigm without severe degradation of system performance is only a small fraction of the full range (Hopfield 1982, 1984; Ratcliff 1990). Second, if the system is one that must select an action conditional on the environmental input, then each output of an array-to-line model represents a distinct physical channel that can be directly connected to the subsystem that generates the action required. Thus, a simple action potential from the axon (line) representing the class of the input pattern can trigger or modulate the appropriate response. In contrast, the output of an array-to-pattern model represents a complex signal that must be decoded and then selectively routed before the appropriate action can be triggered (figure 1.2). The requirement of a separate decoding stage for pattern classification and signal routing makes the array-to-pattern model a weaker computational candidate than the array-to-line model unless there is an overriding advantage on other grounds.

Relaxation and Feed-Forward PDP versus Comb-Filter Processing of Input
Another important design issue concerns the nature of the computations required to map input to output. One common approach in connectionist models is to generate each output state by the computational process of iterative relaxation (Rumelhart and McClelland 1986). In such models, each processing unit U_j is directly connected to all other processing units through a pattern of activation-transfer weights W_{ij} typically ranging from -1 to $+1$. The activity level of any given unit U_j at a given time t is some specified function of its activity level at $t - 1$ and the sum of all the outputs from other units after each is multiplied by its transfer weight W_{ij} on U_j. If there is an appropriate pattern of transfer weights, then after an indeterminate number of time steps, the profile of activity over the population of output units (output vector) will settle into a state desired by a predetermined criterion of goodness. When a stimulus is applied to the system (for example, an input vector of features), we can think of the activity over all units as undergoing a process of dynamic reorganization, finally converging (relaxing) on a stable output pattern (a *local* energy minimum).

An alternative approach utilizes a direct feed-forward filter design and can be characterized as a one-pass processing mechanism. In a system of this kind, each unit α_i of the input vector connects directly to each output unit U_j through a transfer weight W_{ij}. In one version of this approach, at the start of processing (t_0), all output units are either at zero activity or have uniform low-level activity, and all trans-

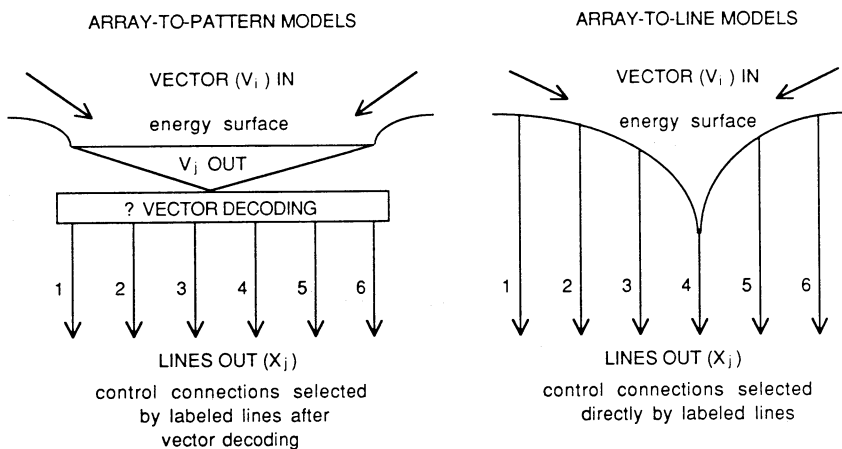ARRAY-TO-PATTERN MODELS   ARRAY-TO-LINE MODELS

Figure 1.2
Illustration of the differences between array-to-pattern (A/P) and array-to-line (A/L) models. In A/P, input vectors are mapped to output vectors (patterns of unit activity over a population of cells). The output patterns must then be remapped to labeled lines (decoded) for selective control. In A/L, input vectors are directly mapped to the labeled line (discrete filter within a comb filter system) having the highest correlation with the input vector (line 4 in the figure on the right). If the pattern of activity represented by the input vector changes sufficiently, the energy minimum will shift from line 4 to a different line with a higher correlation with the new vector. This will be the newly activated labeled line.

fer weights range from near zero to some higher positive value. At $t + 1$, the activity level of each output unit is a function of the sum of the products of the activity in each unit of the input array and its particular transfer weight W_{ij} on the unit of the output array. If there is an appropriate pattern of transfer weights at any time during the presentation of a stimulus, the output unit desired by a predetermined criterion of goodness will have a higher level of activity than any other output unit. This model can be thought of as a parallel-access comb filter (analogous to an array of independent narrow pass-band filters in electronic signal processing) in which the distribution of transfer weights on each output unit determines its selective tuning for particular input patterns (Trehub 1967, 1975a, 1987).

After considering the relative merits of relaxation, feedforward PDP, and comb filtering for cognitive systems, I decided that comb filtering should be the preferred approach. It does not require additional decoding; it seeks the absolute energy minimum, which is, in fact, the output unit having the highest correlation with the current input vector (figure 1.2); it does not exhibit the limitations in memory

efficiency and discrimination that characterize PDP models (Ratcliff 1990, Hopfield 1982). The level of activity in each unit of a comb-filter scheme is independent of any of its neighbors, so the number of arbitrary input-output memories that can exist simultaneously in storage is exactly the number of processing units—filters/output lines—in the system (because of the independence of processing units, subtle distinctions between mappings will not be confounded or distorted in memory, a clear advantage from the standpoint of system competence); it generates a useful output within a single synaptic delay plus the time required for integration to threshold (typically, 1–5 milliseconds); and it makes the minimal assumption that all activation-transfer weights in the filtering cells have a positive sign (excitatory) and that each processing cell in the extended model maintains its activity as either an excitatory or an inhibitory neuron, thus conforming with the constraints of real mammalian synapses and neurons. These considerations, together with additional factors related to efficient learning, imagery, the recall of specific events, and symbol manipulation, convinced me to develop cognitive models based on comb-filter mechanisms.

Representation
At any given moment, a cognitive computational system can be characterized by the internal state that maps an input onto an output. This state can be thought of as the memory of the system. In connectionist models, memory resides in their complex distribution of transfer weights. These complex distributions define an internal set of standards against which arbitrary input stimuli are matched for appropriate routing or classification. The internal standards can be taken as representations of particular kinds of external events that are significant for the system and require a selective response. We may think of such representations as the concepts held by an individual. Since the kinds of events to which the system must respond may vary in a number of particulars yet still maintain their general significance, the question arises of how best to structure the internal representations for adaptive response despite individual variation within significant classes of stimuli.

What kind of computational process should we adopt to provide an appropriate repertoire of concepts? One approach opts for the construction and storage of generalized prototypes where each prototype is a unitary representation of a stimulus class, consisting of some average pattern of features obtained by weighting each feature of a stimulus according to the frequency of its occurrence among the exemplars of its class (Rosch and Mervis 1975, Stillings et al. 1987).

There is a serious problem with this kind of model, however. If the brain were to store its concepts in memory only as unitary averaged prototypes, it would be unable to recognize individuality among particular exemplars of a concept. We recognize a face not only as a face but also as a young face, an old face, a face that is happy or sad. These distinctions are not trivial and could not occur by matching sensory input against an averaged prototypical standard. On the other hand, if we were to maintain in memory a representation for each sensory entity experienced, then the notion of concept would have little meaning, and the storage capacity of the brain would soon be exceeded.

There is a solution to this dilemma, however: store exemplars and establish new classification standards only when the existing conceptual structure fails a real or imagined test of its current utility or when a stimulus is experienced as a novel one in a significant context. In this scheme, the number of representations that one maintains grows as needed by the individual or demanded by the physical and social environment. Clearly, if representations are built in such a fashion, then there must be critical interactions among conceptual structure, motivation, attention, imagination, and learning. The behavior of the cognitive brain model that I propose depends on just such interactions.

Figure 1.3 illustrates the distinction between representation by averaged prototypes and representation by significant exemplars. Imagine the stippled and black regions in the top box as enclosing the point locations, in a multidimensional space, of all experienced variations of two different classes of stimuli. An exhaustive representation within the brain of all instances of each class is neither physically possible nor desirable, so some procedure must be employed before storage to reduce the vast number of stimulus instances depicted at the top. The box at the bottom left shows all exemplars of the two kinds of entities collapsed into two distinct multidimensional points (prototypes) by a process of weighted averaging. Let us call the stippled point A and the black point B. Given an input stimulus, the system will classify the stimulus as either A or B by determining its multidimensional distance from A and B and selecting the class having the lesser distance. The box at the bottom right shows significant exemplars of each kind as subsets of multidimensional points with individual characteristics. In this case, the multidimensional distance of a stimulus from each of the points in the representational space will not only determine its class membership but will also categorize it as having the properties of its closest exemplar.

THEORETICAL MULTIDIMENSIONAL LOCATIONS
FOR ALL INSTANCES OF TWO KINDS

LOCATIONS OF AVERAGED PROTOTYPES   LOCATIONS OF SIGNIFICANT EXEMPLARS
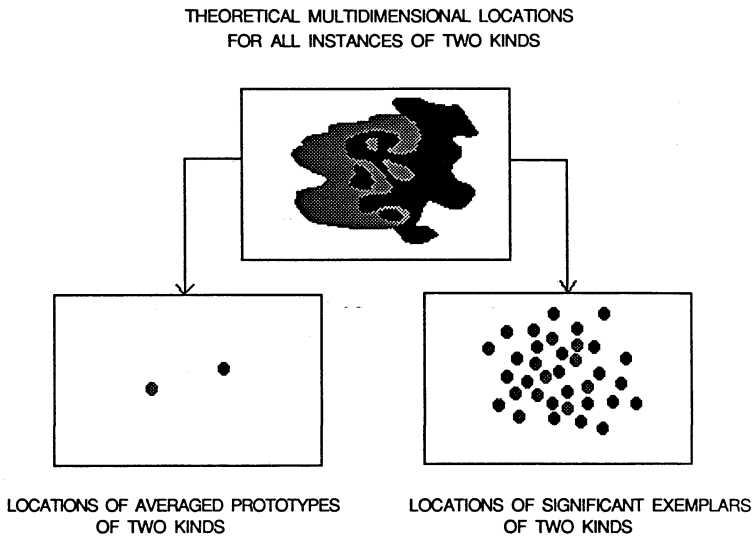OF TWO KINDS   OF TWO KINDS

Figure 1.3
Illustration of the difference between the averaged-prototypes paradigm and the significant-exemplars paradigm for the representation of concepts.

Learning

In connectionist models, the distribution of activation-transfer weights determines the memories the system has, its concepts, and the input-output mappings it will make. Thus, the issue of how transfer weights are set, organized, and modified must have a high priority in any effort to provide a full physical explanation for cognitive processes. In other words, a satisfactory model of learning is of fundamental importance.

Among the many problems that arise in attempting to formulate a neuronal model of human learning, particularly significant are those related to the biological basis for normalization of transfer weights, the number of stimulus repetitions required to reach a satisfactory criterion of performance, and the mechanism for representing the temporal context of a learned event.

The problem of normalization arises because all stimulus patterns that must be learned may vary over a wide range in terms of the number of active units in their input vectors. Moreover, while equal levels of performance may be demanded in response to a variety of stimuli, the number of opportunities for learning may be markedly different for each stimulus. Since the magnitude and distribution of change in transfer weights during learning can be influenced by the

number of active input units evoked by a stimulus and the number of times the same stimulus is experienced, as well as its characteristic pattern, some mechanism is required to prevent system degradation by an excessive bias in favor of stimuli that are relatively large (discharge many units) or that occur relatively frequently. Normalization is important and must be accounted for within the operation of any proposed learning model (Trehub 1975a).

For some problems, many stimulus-response trials are needed before learning can reach an acceptable level of performance, and for others only a single trial is necessary. Most computational models of learning depend on processes that require many trials to reach an acceptable criterion of performance (Rumelhart and McClelland 1986). Typically, a stochastic convergence operation is assumed or incremental weight adjustments are made in an iterative search for a distribution of transfer weights that will lead to an appropriate separation of input patterns. Such models are unable to account for one-trial learning. Ideally, we seek a plausible biological system capable of one-trial learning but that also can manifest the characteristics of progressive improvement over trials.

Another important aspect of any model of learning is whether it allows for self-initiated and self-organized learning. A prevailing paradigm for learning in PDP models is the so-called generalized delta rule (Rumelhart and McClelland 1986; see also Minsky and Papert 1988), which requires the prior specification of a desired output vector. The object is to adjust connection weights throughout the network so that a desired mapping from input vector to output vector is obtained. When the network is stimulated, the activity of every output unit is compared to its desired (target) activity. If there is a discrepancy, its magnitude and sign are measured and used as criteria for adjusting all connection weights to the examined unit in a direction that reduces the observed discrepancy. Typically the magnitude of weight adjustment is proportional to the measured discrepancy. In multilayer networks, the procedure is repeated for corresponding units, working backward to the input layer. For this reason, the error correction scheme is often called back propagation.

Learning of this kind requires an external agent (teacher) to set a standard. It cannot be self-initiated and self-organized unless one assumes that there is an innate store of desired output vectors (targets) appropriate to each stimulus situation; however, in this case, learning would be quite irrelevant. The model I propose is able to learn without the intervention of an external agent but can profit by outside guidance.

A substantial part of human learning and memory is episodic in

character (Tulving 1972). Sequences of experience are interrelated and rendered meaningful by contiguity within particular time frames. Episodic representations are also essential for chaining inferences and for planning future action. Models should be able to learn and conserve, in proper temporal order, sequences of stimuli, as well as nontemporal representations of input patterns.

One final point should be emphasized: any ecologically relevant model of learning must have a mechanism that can selectively retrieve learned representations as a function of an individual's current motivation and environmental contingencies. This introduces the collateral problem of structuring effective and plausible interfaces among the learning mechanism, the various sense modalities, and the motivational system that guides human behavior.

Selecting the First Modality
The choice of which modality to model first will play a role in shaping the subsequent development of the cognitive model.

I have chosen to start with the visual system, for both objective considerations and intuitive preferences. If we assume that the number of neurons in each of the functional/anatomical areas of the human brain gives an indication of the relative ecological importance of that function, then vision must be taken as the most important of our sensory processes. The primary and secondary sensory areas of the human visual cortex contain more than 1.2 billion nerve cells. In contrast, the closest competing cortical areas, devoted to processing bodily sensations, contain roughly 160 million cells (Blinkov and Glezer 1968). Thus, in terms of sheer biological resources, evolution seems to have invested the human brain with at least seven and a half times as many cells for its visual computations as for its somesthetic computations. Another anatomical indication of the importance of vision is revealed when we examine the ratio of the number of cells in the primary visual area of the cortex (area 17) to the number of cells in the lateral geniculate nucleus (a lower visual center) over a variety of mammalian species: in the rabbit, the ratio is 20:1; in the macaque monkey, 145:1; in humans, 900:1 (Blinkov and Glezer 1968). And if secondary visual areas of the cortex were counted in computing these ratios, the divergence among the species would be even greater. This evolutionary progression underscores the central role of vision in human adaptation.

In explaining why people need to have so much more of their brain resources devoted to visual processing than other creatures have, it can be argued that, say, a monkey, as it swings and leaps from tree to tree, must make exquisite discriminations, judgments of distance,

clearance, and the weight-bearing properties of branches, all of which require greater visual acuity and processing speed than do normal human activities. Humans, however, must not only respond to the affordances of the visual environment but must know and understand the world revealed to them through the medium of sight. Vision is the modality that provides the richest direct source of information about the nature of our physical world. Beyond purely sensory considerations, visual experience and imagery give us the most important extensional anchors for the semantics of language (Johnson-Laird, Herrmann, and Chaffin 1984; Miller and Johnson-Laird 1976). One of the key problems in the development of a biologically plausible model for human cognition is to elucidate the relationship between vision and language and to detail the physical interactions between visual and linguistic mechanisms.

Tasks to Be Performed

A list of basic tasks that a cognitive model must be capable of performing tends to focus the design of the system and helps sharpen the test criteria for objective assessment of the model's competence. It is highly unlikely that any such list will cover all cognitive tasks, but whatever is explicitly described will aid the theorist and provide common points of reference for proponents and critics of the model.

Following are a number of general tasks that I have taken as reasonable tests of a human cognitive system. The list starts with tasks specific to the visual domain and broadens to include those related to semantic processing, reasoning, and planning. The model must be able to:

1. Parse an object or a part of an object as a stimulus entity when presented with any arbitrary object or an arrangement of objects (a scene).

2. Represent the relative location of any parsed object in a three-dimensional viewer-centered space.

3. See an object or a scene on at least one occasion and recognize it later despite substantial changes in object size, angular orientation, or position in space.

4. Search for and locate an object that has been learned if it is present in a complex scene.

5. Reconstruct an approximate image of an object or image that has been learned when it is absent.

6. Construct and learn new images by combining parts of objects and scenes recalled from its learned repertoire (memory).

7. Recognize learned patterns despite inputs that are substantially incomplete or degraded by noise.

8. Disambiguate the stimulus and sequentially recognize the constituent patterns if the model is presented with a complex pattern composed by the superposition of previously learned patterns.

9. Detect, learn, and recall spatial relationships among objects in a scene.

10. Respond, given any arbitrary input pattern, with a series of recognition indicants and their associated images recalled from its learned repertoire, which are ordered in output according to some measure of pattern similarity with the arbitrary stimulus.

11. Learn substantial sequences of visual input and later accurately recall at least parts of the image content of selected sequences in correct temporal order.

12. Learn and recall a name for each entity it has learned.

13. Organize and relate its internal representations as equivalents to subject and predicate in a propositional structure.

14. Generate sequences of related inferences, a substantial proportion of which are logically true within the terms of a complex propositional structure.

15. Image, or otherwise recall, if the model is presented with a name, its representation of the object, entity, characteristic, or relationship that the name stands for.

16. Control its behavior in accordance with its motivational needs.

17. Attach some indicant of value to any current or imaged environment (scene) or episode according to the degree to which it meets its motivational needs.

18. Plan, execute, and learn sequences of its own behavior that lead to environments or episodes that meet its motivational needs.

These requirements motivated the mechanisms and the system architecture of the brain model that I propose. They represent a basic core of processes that must be explained as we progress toward a full account. The following chapters address these requirements and present the details of a brain model that I believe provides a credible biophysical and structural foundation for understanding human cognitive competence.