This is particularly important when we consider the apparent goal directedness of cognitive systems. This could, as Smolensky implies (section 7), be modeled teleologically at the conceptual level (i.e., the program could explicitly refer the current state of the system to a specific future state). But as a general form of explanation, teleology in the pure sense leaves a lot to be desired (see George & Johnson 1985) and the mind, as an evolved entity, must be viewed at the subconceptual level as teleonomic (Staddon 1983, p. 6) rather than teleological.[1] An important strength of the subconceptual approach is that it could explain the "emergence" of, for example, goal directedness at the conceptual level in terms of subconceptual components, the gradual and piecemeal evolution of which would be relatively easy to understand.

There are two corollaries of an evolutionary point of view which are relevant to the construction of subconceptual models. The first is that a neurally consistent subconceptual model of a complex cognitive process could (should?) be built by elaborating models of simpler processes (mirroring evolution). The second is that at each stage of additional complexity the current model should be viable in terms of the constraints facing the simpler models (mirroring the development of the nervous system).

Construction *de novo* of a subconceptual model of human cognitive processes would be difficult even if the cognitive processes were better defined than they now are. But our knowledge of the structure of the nervous system, which Smolensky decides as being irrelevant for cognitive modeling, shows that the brain develops by additions to relatively less changing, phylogenetically older parts of the system. Whatever the changes in the emergent properties of the network at each addition, the *subconceptual* descriptions of the functioning of the phylogenetically or developmentally early parts of the system are unlikely to change greatly. While available data on the dynamic behaviour of human neurones during cognition are such as to deter model building, the same is not true of simpler organisms or more peripheral parts of the brain. Thus neural data can be critical for subconceptual (but not conceptual) models. In particular, subconceptual models of simple neural systems (where dynamics may be determinable) should be seen as valid precursors to subconceptual models of more complex neural systems.

The subconceptual architecture of simple systems (locally unlike their conceptual architecture) could also provide useful models for the subconceptual architecture of complex systems. For example, once the principle and the general teleonomy ("purpose") of lateral inhibition are understood for the receptor level of a sensory system this not only transfers to the receptor levels of other sensory systems but also to the higher levels of the original system. Of course, Smolensky argues (section 2.3) that perception is exceptional in the extent to which neural analogy is useful. But consider a second exceptional case – motor control. Current neural models of the cerebellum (e.g., Pellionisz 1979) border on the subconceptual. The control of motor activity can be viewed as dependent on an "intended movement vector" (Llinas & Simpson 1981) which is, in effect, a motor command specified with respect to a sensory coordinate space. The feedback inherent in this idea is virtually identical to that embodied in the stretch reflex but the end result is a simple form of goal direction. As with models of sensory processing, then, models of movement generation can be built by adding onto lower order models control processes which have a very similar architecture at the neural or subconceptual level but which generate markedly different consequences at the conceptual level.

The unexpected emergent powers of appropriate parallel recursive neural or subconceptual networks, and the extent to which nature seems in practice to repeat the same tricks at different phylogenetic levels, suggest that it would not be unreasonable to try to relate subconceptual to neural architecture directly at this time. Smolensky describes a network which produces the effect of a schema by settling into a particular final state on the basis of a specific input and differential connections between nodes of the network. A simpler network using similar connections can solve the global stereopsis problem and settle into a final solution in a way very reminiscent of one's own perceptions on viewing global stereoptic images (see Frisby 1979, p. 150). It is reasonable to search for the neural instantiation of the equivalent of the latter network; when found, its exact form may well have important implications for subconceptual models such as Smolensky's schematising one.

There is an implied scientific connection here between the neural and subconceptual levels: The subconceptual model can incorporate neurally general features and then, through its capacity to demonstrate a solution to a particular problem at the conceptual level, can drive the neurophysiologist to look for the neural equivalent of its components or general dynamics – dynamics which would have had no meaning in the absence of the subconceptual model. Thus, Hebb's (1949) subconceptual speculations about association led physiologists to look for "Hebbian" synapses – and to find an approximate equivalent in the hippocampus. This in turn has led to an account of the hippocampus as a distributed memory system which depends on the known properties of particular subconceptual associative systems (McNaughton & Morris 1987).

I believe, therefore, that subconceptual models of cognition and schemata, should use an architecture related to current models of perceptual systems and that subconceptual models of goals, intentions, and similar ideas should use architecture related to current models of motor control. A corollary of this is that larger subconceptual models should have features which are consistent with the general architecture of both sensory and motor neural models. Subconceptual understanding of cognition may even turn out to be a natural consequence of designing a single formal system which encompasses our current general understanding at the neural level of both perception and motor programming. Available neural information may be highly suggestive even when a direct link to sensory or motor systems is not possible, as in the specific instance of subconceptual models of working memory and the known dynamic properties of the hippocampus.

ACKNOWLEDGMENT

Preparation of this paper was supported by a Senior Fellowship from the New Zealand Medical Research Council.

NOTE

1. Teleonomy represents purpose in the sense that the eye has the purpose of seeing (Monod 1974, p. 20). That is, the form of the eye is limited by the laws of optics and the progressive selective advantage conferred by small changes in the quality of light sensitivity. However, teleonomy at no point implies teleology. The eye is merely a consequence and not a goal of evolution.

The psychology of connectionism

Dominic W. Massaro

Program in Experimental Psychology, University of California, Santa Cruz, CA 95064

Electronic mail: psych20@ucscd.ucsc.edu

Smolensky (1988) offers a well-reasoned overview of connectionism that helps place it in proper perspective. He settles into a level of description between the symbolic and the neural – at what is in my view an information-processing level. This is exactly where connectionism belongs; nor is it without much company at this level. An intermediate level of description is what psychology has been seeking since Fechner, Helmholtz, Donders, and James. In this light, connectionism is an alternative to symbol systems and modularity, but not to the information-processing paradigm. One goal of this commentary is to evaluate several properties of connectionism with respect to the information-processing paradigm currently in use in psychology, and to mention a few precedents of the connectionist approach. Connectionism gives me an appreciation of a continuity of theory development in psychology during the last century. I will conclude with some limitations of current connectionist models as psychological models of mind and behavior.

Evaluating connectionism. Connectionism shares several attributes with information processing even though the primitives and currency of connectionist models are often presented as different from more traditional information-processing models. For example, connectionism has been referred to as parallel distributed processing. However, neither parallel nor distributed appears to be a necessary or sufficient characteristic of either paradigm. Information processing and connectionism contain both parallel and sequential processing. Parallel processing refers to the simultaneous occurrence of two or more processes. Sequential processing exists when one process is dependent on the output of another. Traditional information-processing models have assumed that feature analysis of letters occurs in parallel, and letter recognition is dependent on the output of the feature analysis (e.g., Selfridge's 1959 seminal pandemonium model). In information-processing models, certain processes are assumed to be sequential; for example, a short-term memory search might not begin until the test item is recognized (Sternberg 1975). There is also sequential processing in connectionist models in that top-down activation of lower-level units might not occur until lower-level units activate higher-level units (Anderson et al. 1977).

Local versus distributed processing does not appear to be a characteristic that distinguishes information processing and connectionism. In connectionism, local representation corresponds to the case in which information about a pattern is stored in the connections of a single unit reserved for that pattern. Distributed representation refers to the information about a pattern being stored in the interconnections of a large number of processing units. The local versus distributed distinction in connectionism parallels the earlier distinctions in information processing between template matching and feature analysis (Neisser 1967) and between feature and node theories of memory representation (e.g., Schvaneveldt & Meyer 1973). Template matching places all the information about a pattern within a single representation whereas feature analysis distributes the information across several representations. Local representation of some concept exists when all of the information about the concept is represented in a single location, whereas a distributed representation would be distributed across several locations or not located in one place. Selfridge's (1959) pandemonium model also qualifies as distributed because a letter is represented in terms of the features that make it up. Such a distributed model might also represent structural relations among the features (Oden 1979). Local connectionist models have been proposed, and others contain certain local characteristics such as the local representation of the letters in NETtalk (Sejnowski & Rosenberg 1986). It has been proven that a class of distributed models can be mimicked by local models, which perhaps blurs the local-distributed distinction (Smolensky 1986). It should be stressed that I am not discrediting the local-distributed distinction; the thesis is that the distinction is not new and does not differentiate connectionist and information-processing paradigms of inquiry.

Although the processing is assumed to occur among neural-like units, I agree with Smolensky's thesis that connectionist models are *not* necessarily neural models. If connectionist models were psychological models, then the appropriate test of those models would be at the level of physiology and not at the level of observable behavior (see Moore et al. 1986). However, the current state of the art precludes any tests of connectionism at the neurophysiological level. For example, Marrocco (1986), a neurobiologist, believes that evaluating the neural bases for phenomena such as word-superiority effects is outside the purview of current electrophysiological methods. Connectionist models have been used to explain behavior at a functional level of description by assuming processes analogous to those occurring at a physiological level. This form of explanation is not new. Thorndike (1898) rejected cognitive explanations of cats learning to escape a cage in favor of a neurological explanation. The neurological explanation actually assumed that learning resulted from connections being established between neurons. More recently, the feature detectors for letters have been viewed as instances of neural units uncovered in electrophysiological research (e.g., Lindsay & Norman 1972). Models are metaphors and connectionist models are only less obviously metaphors because the metaphor is glossed neurologically rather than psychologically (Gentner & Grudin 1985). If models are formulated in a connectionist paradigm using "neurological" terms, they might not attract the analytic scrutiny necessary for precision, systematization, and empirical evaluation.

I am impressed by the great affinity between connectionism and stimulus sampling theory (Estes 1950). The latter views behavior in terms of the association of responses and antecedent stimulus events. Association refers to a connection between a stimulus and a response. As Estes (1982) points out, psychological terms such as perception and memory were inadmissible at that time and, strictly speaking, my guess is that reference to anything inside the black box was outlawed. Thus, the connections were established and modified between observable elements such as stimuli and responses. With the "cognitive revolution," it became acceptable to refer to events in the black box, but these events remained at the same functional level as those in behaviorism. Now connectionism offers the hope that the functional level can be replaced by the neurological level, but my belief is that connectionism can be tested only at the functional level and must therefore be evaluated at that level.

In current implementations of connectionism, learning can be instantiated by changing the weights of the connections between the neural units. The same outcomes could be implemented in a variety of other ways, as acknowledged by connectionist theorists (Rumelhart et al. 1986). Rather than modifying existing connections between units, changes in performance might be supported by establishing new connections. There would be some probability of establishing a new connection between units that are active at the same time. This simple learning rule could be formalized to produce mathematical predictions equivalent to those of stimulus sampling theory (Estes 1950). What is important to note for our purposes is that the several neural possibilities do not appear to inform functional explanations of human performance. Although Thorndike (1898) remained a connectionist throughout his long and productive career, he abandoned the neurological part of his theory (Hilgard 1987). Similarly, current assumptions about neural processing in connectionist models might meet the same fate as the "ether" that was assumed to be the medium of light-wave propagation (Pribram 1986). The mathematics of light-wave propagation did not change with the demise of the "ether," and some of the mathematics of connectionism might remain with fundamental changes in the current assumptions about neural functioning.

In information processing, we ask to what extent one input system is influenced by processing from another input system or to what extent information from one stage of processing is communicated to another stage of processing. Analogously, information processing addresses the question of the extent to which separate operations must be postulated to describe the relationship between perception and action. In the classic example of signal detection theory, it was necessary to assume separate sensory and decision stages of processing as opposed to simply assuming a direct mapping between stimulus and recessing. Connectionist models, on the other hand, have blurred the distinction between sequential operations or stages of processing. They tend to assume a direct mapping of connections between input and output. The assumption of sequential stages of processing appears to be necessary for mapping categorical information into action, even in connectionist models. The primitive input features are not sufficient to determine the appropriate response. In reading English text aloud for example, visual features of the letters do not predict the pronunciation as well as the letter names do (i.e., categories).

Smolensky makes it clear that connectionism offers an alternative to the physical-symbol-system paradigm (see also Dorffner & Place 1986). The latter often uses symbols to embody sensory experience and rules to map experience into action. Connectionism uses activations to embody sensory experience and the modification and transmission of these activations to map experience into action. Connectionism is completely consistent with the paradigm of information processing, however. As noted in the characterization of information processing, symbols were not necessary for the representation of information in the system. In fact, information-processing theory might be considered to have a history of nonsymbolic representations including discriminability, familiarity, memory strength, and even activation and inhibition among representations. In fact, Massaro and Cohen (1987) proved that a specific connectionist model with only input and output layers is sometimes mathematically equivalent to a specific process model called the fuzzy logical model of perception (Massaro 1987).

Connectionism also contrasts with the fundamental principles of modularity (Fodor 1983; Gardner 1985). Modularity has as its foundation the same decomposition assumption as information processing. Complex behavior can be decomposed into a myriad of simpler processes, and these processes might be further decomposed into yet simpler processes. The strong claims of modularity involve the lack of communication among input systems and the impossibility of a functional "middle-level" description of cognition. Modularity assumes some isolation among processes because input processes occur relatively independently of one another. Although the modularity assumption is not inconsistent with information processing, it is inconsistent with the high interconnectivity among the units from the earliest until the latest processing in connectionist models. Both modularity and connectionism make assumptions that are considered empirical questions within the information-processing paradigm. Given this observation, the issue concerns not whether connectionism or modularity is the better description of human performance a priori but how the paradigm of information processing dictates an empirical test between these alternatives.

Constraints on psychological inquiry. It is important to consider three constraints on psychological inquiry. The first is that the phenomena of interest are highly complex, but they do not easily constrain psychological theory (Manicas & Secord 1983). For example, the acquisition of the past tense is a highly complicated affair and can be loosely described by diametrically opposed rule-based and instance-based theories. A second constraint on psychological inquiry is that scientists have a confirmation bias and seek to generate reality to support their pet theory (Tweney et al. 1981). The third constraint on psychological inquiry is that there is a plethora of sufficient theories for the phenomena of interest. As made transparent by case histories of specific phenomena (e.g., memory search) and progress in mathematical psychology, there are many ways to skin a cat or to describe a particular observed phenomenon.

Strategies for psychological inquiry. Faced with these three constraints, what are the implications for psychological inquiry? In my view, four strategies can overcome the three constraints. First it is necessary to develop highly simplified experimental studies in order to reveal fundamental regularities or laws related to the phenomena of interest; these regularities or laws cannot be tested in highly complex situations. A complex experiment or natural situation is influenced by multiple causes; the regularities of each will not be easily disentangled. The second remedy for psychological inquiry is the requirement to test between alternative models of performance using the research strategy of falsification (Popper 1959) and strong inference (Platt 1964). The investigator is required to develop opposing models and to devise an experiment that tests the differential predictions of those models. Simply accumulating evidence that is consistent with a given model is not an ideal strategy because the data might be equally consistent with a diametrically opposed model. Third, it is necessary to carry out incisive experiments and to perform a fine-grained analysis of the results to distinguish among competing models. Only by thorough, systematic, and single-minded analyses can an investigator eliminate alternative models. Only this elimination will reduce the set of models consistent with the observations of human performance.

Good scientific inquiry dictates that we should actively attempt to eliminate alternative models. Given the power of physical symbol systems and connectionism, however, the observed data often fail to be sufficient to decide among alternative models. The fourth strategy is that only discriminating models are permitted to survive. To be discriminating is to predict *only* the actual results, not the universe of possible results. The investigator therefore addresses not only observed results but also a range of hypothetical results that do not necessarily occur. Collyer (1985) has made a similar point that although a more complex model might be more accurate than a less complex model, the more complex model should not necessarily be preferred. The more complex model would be less preferable if it also predicted results that did not occur.

Empirical tests of connectionist assumptions. Taking the tack of fine-grained tests, I tested the assumption of interactive activation (i.e., two-way connections between units) of specific connectionist models (Massaro 1988; submitted). Interactive activation was shown to be both unnecessary and inconsistent with empirical results. Both the interactive-activation model of written word recognition (McClelland & Rumelhart 1981) and the TRACE model of speech perception (McClelland & Elman 1986) have been falsified (Massaro 1988; submitted), when tested against human performance. Interactive activation is not necessary to account for the integration of multiple sources of information in language processing. In addition, interactive activation is inconsistent with the finding that the representation of bottom-up information remains independent of top-down context. This research demonstrates that it is possible to test assumptions of specific connectionist models in the same way that assumptions of information-processing models can be tested.

Superpower of hidden units. I have used the fourth strategy to assess connectionist models with hidden units. As anticipated by Minsky and Papert (1969) in their critique of perceptrons (Rosenblatt 1958), connectionist models with more than two layers of units might be too unconstrained to be informative. Models of this type might be Turing-equivalents that are capable of mimicking any computable function. As presently formulated, many of the connectionist models with two-way connections among different levels of units and connectivity among units at a given level appear to be too powerful (superpowerful). They appear to be capable of predicting not only observed results but also results that do not occur (Massaro 1986; 1988). I have demonstrated that some connectionist models simulate results that have not been observed in psychological investigations and results generated by incorrect process models of performance (Massaro 1988).

Both the frameworks of psychophysics (specifying the environmental characteristics used by subjects) and information processing (specifying the processing these characteristics undergo) are necessary for progress in the field. Apparently, connectionist models do not work out the psychophysics in any great detail. In fact, the claims they make are often compromised by the arbitrary nature of the psychophysics they assume. As an example, in the TRACE model of McClelland and Elman (1986), they acknowledge that although the features, phonemes, and word levels were central to the predictions, these may not be the actual units that are functional. They go on to state that their model is equally valid even if different units are functional. Perhaps this complaint is best summarized by a recent request transmitted by electronic mail: "I'm trying to work up a proposal for research in classificatory neural networks. The computational end is easy, but I need an application that will provide the data for experimentation." My point is that the ease of computation can preclude psychophysical inquiry.

The superpower of connectionist models with hidden units can also camouflage the observation of different stages of processing. Hidden units serve only to bypass the assumption of an intervening stage of processing in which the input is categorized before an output is selected. Consider the pronunciation of the letters of the alphabet. If the input consists of visual features of the letters and the output consists of articulatory features of the pronunciation, it is unlikely that the mapping between input and output would be linearly separable. For example, the letters *c* and *o* would have more similar input representations than the letters *c* and *z*. However, the output representations would be more similar for *c* and *z* than for the letters *c* and *o*. Some intermediate categorization is necessarily involved to map visual properties of the letters to articulatory properties of the pronunciation. In experimental situations somewhat analogous to this example, Miller (1982) found that some type of categorization occurs before input is mapped onto output. If the categorization of the letters was of interest, however, then the mapping between input and output would be linearly separable. In this case, no hidden units would be needed to solve the mapping. Categorization is a natural intervening process that is assumed in information processing but not in current connectionist models. For example, NETtalk (Sejnowski & Rosenberg 1986) used letters rather than visual features as input because the number of hidden units to solve the mapping of visual features to pronunciation would have been prohibitive. My belief is that connectionism will have to become more stage-like to solve mappings in an informative manner.

ACKNOWLEDGMENTS

The research reported in this paper and the writing of the paper were supported, in part, by grants from the National Science Foundation (BNS-83-15192), the Public Health Service (PHS R01 NS 20314), and the graduate division of the University of California, Santa Cruz. Michael Cohen, Ray Gibbs, and Alan Kawamoto provided helpful discussions on several issues discussed in the paper.

Level of analysis is not a central issue

James A. Reggia

Departments of Computer Science and Neurology, University of Maryland, College Park, MD 20742

Electronic mail: reggia@cs.umd.edu

In Smolensky's (1988) thoughtful and stimulating essay and the accompanying commentaries a major emphasis is placed on the importance of the *level of analysis* for a "proper treatment of connectionism" (e.g., section 1.3). Specifying the level of analysis at which one is working is certainly important. However, it is largely irrelevant and even misleading for identifying the fundamental differences between connectionist models and "traditional models" in cognitive science and artificial intelligence (AI) and their respective contributions. The salient distinctions (presented in more detail elsewhere; Reggia & Sutton 1988) lie in the nature of the information processing involved:

1. Connectionist networks are *self-processing*. Traditional AI-like models typically consist of a passive data structure (associative network, set of production rules, etc.) which is manipulated by an active external process/procedure (abductive inference mechanism, rule interpreter, etc.). In contrast, the nodes and links in a connectionist model are themselves active processing agents; there is typically no active external agent that operates on them, either during "problem solving" or during network modification (self-organization).

2. Connectionist models exhibit global (emergent) system behaviors derived from *concurrent local interactions* of their numerous components. The external process that manipulates the underlying data structures in traditional AI-like models typically has global access to the entire network/rule-set, and processing is strongly and explicitly sequentialized (e.g., conflict resolution in rule-based systems).

These distinctions hold regardless of "level of analysis." They are sufficiently specific to permit substantive analysis and comparison. Smolensky's "proper treatment of connectionism" almost equates connectionist models at the subsymbolic/subconceptual level with connectionist models of relevance to cognitive science in general. In contrast, from the viewpoint offered here, connectionist as well as traditional models are each simply generic classes of information processing systems that are potentially relevant at *all* levels of analysis.

To see this, it is useful to consider an example where both a connectionist model (Peng & Reggia 1989; Wald et al. 1989) and a traditional model (Peng & Reggia 1987) have been developed at the same level. The specific task involved with both of these models is to derive plausible explanatory hypotheses (diagnoses) for certain observed manifestations (symptoms) during diagnostic problem solving. Although there are many other tasks for which both classes of models have been developed, this specific example is relatively unique in that both the connectionist and the traditional models are, by definition, at exactly the same level of analysis in the sense that both are based on the same network consisting of the same nodes (representing disorders and manifestations), the same connections (representing causal associations), and the same weights (representing conditional and prior probabilities). Both models are even derived from a common theoretical foundation; the motivation for developing the connectionist version was to contrast these two paradigms. Despite being at the same level of analysis, however, the two models can be distinguished by features (1) and (2), described above. This provides a clear counterexample to the claim that level of analysis is a fundamental distinction between connectionist and traditional models.

This observation also bears on the distinction between subsymbolic connectionist models and neural network models. Smolensky and others have correctly emphasized that a subsymbolic cognitive model is not necessarily a neural model. But basing this emphasis primarily on the level of analysis confuses rather than clarifies the role of connectionism in mind and brain modeling. For example, consider the statement that "the subconceptual level of analysis is higher than the neural level" (Smolensky, 1988, p. 10, point 18b). The subconceptual level is part of the list "supraconceptual, conceptual, subconceptual, subsubconceptual . . ." that describes *nonphysical cognitive mechanisms*. Neural systems involve a corresponding list "brain, center-pathway, . . . cell group, individual neuron, subneuronal . . ." that describes *biophysical mechanisms*. Smolensky (section 4) appears to suggest that there is only one as-yet-poorly-identified level of connectionist modeling relevant to neural systems. In contrast, the viewpoint presented here is that there are many levels at which connectionist models of brain function can be conceived; how they correlate with the levels of connectionist models of cognitive mechanisms remains to be determined at all levels. For example, neurologists often analyze patients in terms of a "connectionist model" at the level of cortical "centers." The brain is viewed as consisting of centers