

*THE INTERPRETATION OF COMPLEX HUMAN BEHAVIOR:
SOME REACTIONS TO PARALLEL DISTRIBUTED PROCESSING,
EDITED BY J. L. McCLELLAND, D. E. RUMELHART,
AND THE PDP RESEARCH GROUP¹*

JOHN W. DONAHOE AND DAVID C. PALMER

UNIVERSITY OF MASSACHUSETTS/AMHERST AND SMITH COLLEGE

The publication, in two volumes, of *Parallel Distributed Processing* (McClelland, Rumelhart, & the PDP Group, 1986; Rumelhart, McClelland, & the PDP Group, 1986) has been heralded as a major potential contribution to the understanding of complex human behavior. For example, the review published in *Contemporary Psychology* begins: "These two volumes may turn out to be among the most important books yet written for cognitive psychology. They are already among the most controversial. . . . This new paradigm challenges the fundamental assumption underlying the currently dominant symbolic paradigm; namely, that mental processes can be modeled as programs running on a digital computer" (S. E. Palmer, 1987, p. 925). Even those who have expressed principled reservations concerning certain aspects of the PDP approach "believe this realm of work to be immensely important and rich" (Minsky & Papert, 1988, p. vii). (See Bechtel, 1985, and Miller, 1986 for other reactions to the PDP approach.²)

Parallel distributed processing is a term used to encompass a number of related models of cognition featuring networks of units (loosely analogous to collections of neurons) whose interconnections are modified through a feedback mechanism (loosely analogous to reinforcement). These models are fundamentally different from typical models of cognitive psychology in that they are selectionist rather than essentialist in flavor (cf. Donahoe, 1984). That is, the functionality of connections among the units is the result of selection by the environment rather than design by the theorist.

Behavior analysts—as students of behavior, including complex human behavior—should closely examine the approach presented in *Parallel Distributed Processing* (hereafter *PDP*) for several reasons: (a) the PDP approach is explicitly critical of many of the same constructs of mainstream cognitive psychology that are regarded as unhelpful in behavior analysis. (b) The general approach represented by *PDP* falls within the same broad conceptual framework—historical science—that encompasses behavior analysis. (c) Some of the specific accounts of complex behavior proposed in *PDP* functionally parallel the corresponding accounts advanced by behavior analysis.

To list but a few criticisms of cognitive constructs, *PDP* questions whether the cumulative effect of experience is usefully described as the storage of memories (e.g., McClelland, Rumelhart, & Hinton, 1986, p. 31), whether regularities in verbal behavior may be taken as unambiguous evidence of control by syntactic rules (e.g., McClelland, Rumelhart, & Hinton, 1986, p. 32), and whether discontinuities in the development of complex behavior require the postulation of critical periods (Munro, 1986, pp. 471-472). (References to material in *PDP* conform

¹ Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, Massachusetts: MIT Press; McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, Massachusetts: MIT Press.

Preparation of this manuscript was supported in part by NSF Grant BNS-8409948 to the University of Massachusetts/Amherst. Reprints may be requested from the first author at the Program in Neuroscience and Behavior, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003 or from the second author at the Department of Psychology, Smith College, Northampton, Massachusetts 01060.

² See also Reese, H. W. (1986, May). *Computer simulation and behavior analysis: Similarities and differences*. Paper presented at the meeting of the Association for Behavior Analysis, Milwaukee, Wisconsin.

to the convention of citing the author or authors of the chapter followed by the relevant page numbers.) Thus the PDP approach makes common cause with behavior analysis in questioning the utility of the storage metaphor, syntactic rules, and critical periods in understanding complex behavior (cf. Skinner, 1974). But common objections to such notions are perhaps the least important reasons for behavior analysts to explore the PDP approach.

HISTORICAL SCIENCE AND THE ORIGINS OF COMPLEXITY

More compelling than the sharing of common criticisms is the fact that the PDP approach and behavior analysis are both instances of historical science and, as such, assume a common stance in their efforts to understand complexity. For present purposes, two salient features of historical science are recognized. First, historical science views complex phenomena as the *by-products of the action of lower level processes* and not as the directed outcome of processes operating at the level at which the complex phenomena are observed. (As the term *level* is used here, it denotes not only more grossly differing levels such as the physiological and behavioral, but also finer gradations such as the microbehavioral and macrobehavioral. See Campbell, 1974a, for a discussion of reductionism in historical science.)

As an illustration from the best known of the historical sciences—evolutionary biology—the complex characteristics that are observed in existing species are interpreted as the result of the cumulative action of genetic processes, summarized by principles such as natural selection, and not as the expression of higher order processes such as special creation. Cosmogony is another historical science, one which seeks to understand the origin of the universe as the cumulative product of physical processes summarized by principles such as those of Newtonian mechanics. Within behavior analysis, the lower level processes of which complex behavior is the product are those associated with selection by reinforcement (e.g., Donahoe & Wessells, 1980; Skinner, 1953, 1974). (For extended discussions of historical science in evolutionary epistemology see Campbell, 1974b; in biology

see Mayr, 1982, and Sober, 1984; in behavior analysis see Skinner, 1966a, 1981, and Donahoe, Palmer, & Carlson, in press.)

A second characteristic of historical science is that the basic processes, whose cumulative effects yield complex phenomena, act in a particular order upon a given set of initial conditions—but *both the initial conditions and the order of action of the processes are imperfectly known*. Returning to the example of evolutionary biology, the processes responsible for selection acted on the chemical compounds of the lifeless environment of the early earth and, subsequently, on the products of prior selections. However, a precise specification of the environment of the early earth and of the order in which selection processes acted on that environment is not perfectly known. Such uncertainties are central to the nature of historical science because their existence imposes inescapable limitations on the account of complexity that is possible: Even if the principles describing the basic selection processes were known with certainty, uncertainty in their implications would remain because of incomplete knowledge of the initial conditions and of the order in which the various processes operated.

In a mature historical science, the principles that summarize the action of basic processes are *sufficient* to account for complexity, but they cannot be shown to be necessary for its occurrence. For example, the historical science of evolutionary biology may be competent to describe how a given species might have evolved but be unable to demonstrate that only that particular species could have arisen. With slightly different initial conditions acted upon in a slightly different order, the same selecting processes could have produced a different species (cf. Dawkins, 1986). The test of the adequacy of an historical science is its ability to provide a plausible account whereby a wide variety of complex phenomena could have been produced by the action of a small set of basic processes. In summary, because of limitations in knowledge of the complete history of selection, the power of historical science must often be evaluated by the sufficiency of its account of complexity and not by the necessity of that account. (Although this limitation of historical science has not been generally acknowledged within cognitive psychology, it has been clearly ar-

ticated by some workers, most notably Anderson, 1978, 1983.)

The Analysis and Interpretation of Behavior

Behavior analysis has explicitly recognized the inescapable difficulties in understanding complex behavior because of its status as one of the historical sciences. This recognition is most apparent in Skinner's (1957, 1974) distinction between the *experimental analysis* and the *interpretation* of behavior. In the experimental analysis of behavior, (primarily) laboratory methods are used to manipulate environmental variables with the goal of identifying the functional relations between those variables and behavior (e.g., Skinner, 1935, 1938, 1966b). The functional relations—especially those expressing the effects of the contingencies of reinforcement—provide, in turn, the basis for establishing “simplifying uniformities” (Skinner, 1966b, p. 216). The endpoint of experimental analysis is the statement of these “uniformities” in the form of principles that summarize the action of the basic processes identified by the science.

Interpretation begins where experimental analysis leaves off. That is, interpretation begins with principles derived from experimental analysis and, then, explores the implications of those principles for the understanding of complex behavioral phenomena. As with other historical sciences, if the processes described by these principles are sufficient to produce the observed complex behavior (i.e., the occurrence of the behavior is consistent with the principles) then the behavior is said to be understood (interpreted) in terms of the principles. (Of course, there are also important reciprocal relations between interpretation and behavior analysis: “The interpretation of human affairs is a rich source of suggestions for experiments” Skinner, 1966b, p. 216.)

The distinction between the experimental analysis and the interpretation of behavior is a fundamental one in behavior analysis, and the failure to appreciate that distinction has led to misunderstandings vis-à-vis other approaches to complex behavior. Perhaps the most unfortunate instance of this misunderstanding occurred with Skinner's *Verbal Behavior* (1957). At the outset of that work, Skinner states: “The present extension to

verbal behavior is thus an exercise in interpretation rather than a quantitative extrapolation of rigorous experimental results” (p. 11). This effort to interpret verbal behavior was widely viewed by cognitivists as an inappropriate application of inadequate laboratory principles to complex behavior (e.g., Chomsky, 1959). Attempts to understand other complex behavior have been similarly received, although they too follow the path of interpretation—the means whereby historical science accounts for complex phenomena. For example, in *About Behaviorism*, Skinner (1974) notes: “Much of the argument goes beyond the established facts. I am concerned with interpretation rather than prediction and control” (p. 21). “. . . Our knowledge . . . is limited by accessibility, not by the nature of the facts. . . . As in other sciences, we often lack the information necessary for prediction and control and must be satisfied with interpretation, but our interpretations will have the support of the prediction and control which have been possible under other conditions” (p. 194). “We cannot predict or control human behavior in daily life with the precision obtained in the laboratory, but we can nevertheless use results from the laboratory to interpret behavior elsewhere” (p. 251).

Whatever one thinks of experimental-analytic principles, the effort to interpret complex behavior in terms of laboratory-based principles is entirely in keeping with the practice of other historical sciences, and not a peculiarity of behavior analysis. It is the practice in evolutionary biology where the evolution of a peacock's tail feathers is traced to the action of natural selection on primordial hair cells. It is the practice of cosmogony where the development of the solar system is traced to the action of processes described by Newtonian mechanics on a swirling cloud of interstellar dust particles. The uncertainty in these accounts of complexity is an inherent and inescapable characteristic of historical science; it cannot be circumvented by any alternative formulation because the uncertainty arises from irremediable lapses in our knowledge of the initial conditions and the selection history (Donahoe et al., in press).

PDP as Historical Science

A reading of *PDP* makes is abundantly clear that behavioral complexity is viewed

as the cumulative product of lower level processes. In this respect, *PDP* falls squarely within the conceptual framework of historical science. The point is made in many places, often in the context of discussions of such molar constructs as rules or schemata. Both of these constructs are central notions within mainstream cognitive psychology. "The idea of parallel distributed processing [is] that intelligence emerges from the interactions of large numbers of simple processing units . . ." (Rumelhart, McClelland, & the PDP Group, 1986, p. ix). And later, in the same vein, "the apparent application of rules could readily emerge from interactions among simple processing units rather than from application of any higher level rules" (Rumelhart & McClelland, 1986, p. 120). "Many of the constructs of macrolevel descriptions such as schemata, prototypes, rules, productions, etc. can be viewed as emerging out of interactions of the microstructure of distributed models" (Rumelhart & McClelland, 1986, p. 125). "Schemata are not 'things.' There is no representational object which is a schema. Rather, schemata emerge at the moment they are needed from the interaction of large numbers of much simpler elements all working in concert with one another. . . . In the conventional story, schemata are stored in memory. . . . In our case, *nothing stored corresponds very closely to a schema*. What is stored is a set of connection strengths which, when activated, . . . generate states that correspond to instantiated schemata" (Rumelhart, Smolensky, McClelland, & Hinton, 1986, pp. 20–21). "The fact that our microstructural models can account for many of the facts about the representation of general and specific information . . . makes us ask why we should view constructs like logogens, prototypes, and schemata as anything other than convenient approximate descriptions . . ." (Rumelhart & McClelland, 1986, p. 127). The effort within *PDP* to account for complexity in terms of simpler processes has not escaped other cognitivists. A reviewer remarked, "The importance of PDP models stems mainly from their unexpected emergent properties. . . . The schemata themselves are not explicit data structures at all, but rather implicit structures distributed over the mass of interconnections among the units. They exist explicitly only

theoretically . . ." (S. E. Palmer, 1987, pp. 926–927; cf. Skinner, 1977).

Although both PDP and behavior analysis conform equally to the view of historical science that complexity emerges from the action of lower level processes, the approaches depart fundamentally in the means whereby those processes are identified. In behavior analysis, as in other historical sciences, the processes that lead to complexity are the direct result of independent experimental analyses. That is, the validity of these processes is not dependent on the extent to which they are sufficient to interpret complex behavior, although their competence to do so contributes to their validity. In the PDP approach, unlike historical science generally, the basic processes are typically inferred from the complex behavior that they seek to interpret. Furthermore, the description of these processes is more often constrained by logical and mathematical rather than experimental-analytic considerations (e.g., Rumelhart & McClelland, 1986, p. 133; see also Klopff, 1988).

Thus far, we have described the salient features of historical science and the means whereby complex phenomena are interpreted in such sciences. We have seen that behavior-analytic and PDP approaches are both instances of historical science and, therefore, have a number of characteristics in common, such as their denial of the utility of many molar constructs from normative cognitive psychology. However, we have claimed that PDP fails to conform to the practice in other historical sciences that interpretation is based solely upon principles established by independent experimental analyses. In the remainder of this paper, we present the PDP approach in sufficient detail to permit this claim to be examined, and then make several suggestions toward an integration of the PDP approach with behavior analysis.

ADAPTIVE NETWORKS AND INTERPRETATION

Three complementary strategies of interpretation have been implemented in behavior analysis—verbal interpretation, organismic interpretation, and formal interpretation (Donahoe et al., in press). In the *verbal interpretation* of complex behavior, the impli-

cations of experimental-analytic principles are pursued using the conventions of ordinary language. What distinguishes verbal interpretation from mere speculation is that, like all interpretation in historical science, it appeals only to processes that have been identified in prior experimental analyses. Verbal interpretation is by far the most commonly employed interpretative strategy in behavior analysis, with *Science and Human Behavior* (Skinner, 1953), *Verbal Behavior* (Skinner, 1957), and *About Behaviorism* (Skinner, 1974) providing notable examples.

Although verbal interpretation is a useful method for understanding complex phenomena, it has distinct disadvantages. Very often, especially with more complex phenomena, a number of processes are involved—acting simultaneously and in many different sequences—and a purely verbal account cannot effectively keep track of them all. Although verbal interpretation will always play an important role in behavior analysis and other historical sciences, particularly during their earlier phases of development, more precise methods are desirable.

In *organismic interpretation*, complex behavior (or some aspect of it) is observed in one organism and simulated in another organism by exposing that organism to a sequence of conditions thought to produce the complex behavior in the first organism. As examples, the sequence of responses described by Kohler (1925) as demonstrating “insight” with chimpanzees has been functionally reproduced with pigeons (e.g., Epstein, Kirshnit, Lanza, & Rubin, 1984), behavioral interchanges between chimpanzees said to reveal “communication” (Savage-Rumbaugh, 1984, 1986) have been simulated with pigeons (Epstein, Lanza, & Skinner, 1980), and some aspects of generalization by children of number inflections in nouns (Berko, 1958) have been simulated with pigeons (Catania & Cerutti, 1985). In each of these organismic interpretations, the test organism has been the focus of a particular sequence of selection processes all of which had been identified and directly studied in prior experimental analyses. It should be noted that most human experimentation—whether conducted within the behavior-analytic tradition or otherwise—falls into the category of organismic inter-

pretation. This is so because, with very few exceptions, the processes contributing to complex behavior are not all under the control of the experimenter. Although research with human subjects may very carefully control the variables *within* the experiment, the differing preexperimental selection histories of the subjects *outside* the study cannot be completely controlled or even described.

In the third interpretative strategy, *formal interpretation*, logical and/or mathematical techniques are used to explore the implications of experimental-analytic principles for complex behavior. Included here would be efforts to use computer simulations of basic reinforcement processes to interpret choice behavior (e.g., Hinson & Staddon, 1983; Shimp, 1969), and it is this strategy of interpretation that the PDP approach most resembles.

Not all computer simulation is an instance of formal interpretation. Often, the goal of computer simulation is to devise a program whose output simply mimics some aspect of complex behavior, but in which the instructions of the program do not implement experimental-analytic principles. Strictly speaking, simulations of this type fall within the rubric of artificial intelligence. However, it must be noted that unless the processes implemented in PDP or other types of computer simulations have some independent experimental foundation, the distinction between simulations in artificial intelligence and in cognitive psychology is chiefly that the former are constructed by computer scientists and the latter by cognitive psychologists.

In comparison to verbal interpretation, formal interpretation has the advantage of being precisely stated in the instructions of the program and of being able to keep track of many simultaneously interacting processes. In comparison to organismic interpretation, formal interpretation has the advantage of implementing a selection history that might require many months or years in a living organism. A disadvantage of computer simulation as a means of formal interpretation is that the initial conditions from which the computer program begins are often imperfectly specified. That is, information comparable to the evolutionary history of the organism is unavailable. It is in this respect—a shared evolutionary history—that organismic simulations

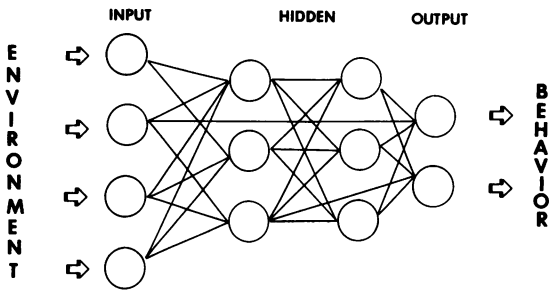


Fig. 1. A network composed of input units in contact with the environment, hidden units that are not in direct contact with the environment, and output units that constitute the behavior of the network. If a unit is activated, it will probabilistically activate all those units to which it is connected. The possible pathways between units are indicated by lines, with activation propagating along a pathway in only one direction (here, from left to right).

have an advantage over computer simulations of complex behavior (Epstein, 1984).

The PDP Approach

PDP describes a subset of a more general approach to interpretation known as adaptive network theory. (For surveys of this field, see Nilsson, 1965, and Minsky & Papert, 1988, pp. 247–287.) Adaptive network theory is described by the authors of PDP as “neurally inspired” (Rumelhart, Hinton, & McClelland, 1986, p. 75). In keeping with this inspiration, a network consists of a number of units, analogous to neurons or groups of neurons, that are interconnected (see Figure 1).

It is useful to distinguish among three types of units—input units, output units, and so-called “hidden” units. Input units are in direct contact with the environment of the network; output units constitute the behavior of the network; hidden units lie between the input and output units. Hidden units are analogous to interneurons. The environment activates one or more of the input units to provide an input pattern to the network. The activated input units then probabilistically activate the hidden units to which they are connected and these, in turn, activate some of the output units. The activated output units define the output pattern obtained from the network. The extent to which a unit activates subsequent units in the network depends upon the strength of the connection, or connection weight, between the units. Connection weights

are analogous to the synaptic efficacies with which one neuron activates the neurons with which it is in proximity. That is, the larger the connection weight, the greater the probability that activity in a “presynaptic” unit will activate a “postsynaptic” unit.

In typical simulations using adaptive networks, the initial connection weights are assigned small, randomly determined values. Thus, the network is most often a *tabula rasa*, although this is not a requirement for such simulations. The competence of the network to simulate some complex environment–behavior relation is demonstrated if the connection weights can be modified such that an input pattern from the environment reliably activates an appropriate output pattern and does not activate inappropriate output patterns. In behavior-analytic terms, complex behavior is simulated when the network has formed a discrimination.

With adaptive networks, the connection weights are modified as a result of the “experience” of the network. That is, if an input pattern activates an appropriate output pattern, the weights tend not to change. However, if the obtained output pattern does not correspond to the appropriate output pattern, then the weights are changed in proportion to the difference between the obtained pattern and the appropriate pattern. A large difference produces a large change in the connection weights; a small difference produces a small change in the weights. It is clear, then, that adaptive networks simulate complex behavior through a selection process (i.e., “learning”) and that the selection process is a function of the consequences scheduled for the output of the network. In behavior-analytic terms, complex environment–behavior relations in adaptive networks are the product of selection by reinforcement.

The PDP approach primarily exploits one method, the generalized delta rule (Rumelhart, Hinton, & Williams, 1986, pp. 318–362; Stone, 1986, pp. 444–459), to implement the selection process for changing connection weights. As mentioned previously, changes in the connection weights occur when there is a discrepancy between the output pattern produced by an input pattern and the output pattern that is appropriate for that input pattern. Adjustment of the connection weights is relatively straightforward when there are

direct connections between the input and output units (i.e., no hidden units). (An adaptive network with only input and output units is called a perceptron; Rosenblatt, 1962.) The connection weights are unchanged along paths that are active when appropriate output patterns occur; the weights are decreased along paths that are active when inappropriate patterns occur. Eventually, as shown by the perceptron convergence theorem, there are some classes of environment-behavior relations for which appropriate weights will be found (Minsky & Papert, 1969). However, when there are hidden units—which are required to represent some classes of environment-behavior relations—then the method for adjusting the connection weights throughout the network is less clear. Minsky and Papert (1988) have put the matter succinctly. “. . . Until recently, [adaptive network theory] has been paralyzed by the following dilemma: Perceptrons could learn anything that they could represent, but they were too limited in what they could represent. Multilayered networks were less limited in what they could represent, but they had no reliable learning procedure” (p. 256).

The generalized delta rule extends the basic thrust of the perceptron convergence theorem—the reduction of the discrepancy between the obtained and appropriate output—to multilayer networks (i.e., to networks containing hidden units). The goal is to select connection weights throughout the network that minimize the discrepancy between the obtained and the desired output patterns. The generalized delta rule accomplishes this goal by activating the network with an input pattern and then, working backward from the obtained output pattern, adjusting the connection weights of increasingly earlier layers of the network. The adjustments are proportional to the discrepancy between the obtained and the appropriate output pattern. Because the discrepancy is “passed back” to the connection weights of earlier units in the network, the generalized delta rule is said to “back-propagate” the discrepancy. Back-propagation exploits the known mathematical relation between the derivative of a function, here changes in discrepancy with changes in the connection weights, and derivatives of components of that function, here changes in the discrepancy with changes in the output

of the network and changes in the output of the network with changes in the weights. This is the so-called chain rule of composite functions (see Rumelhart, Hinton, & Williams, 1986, pp. 322–328). Although the generalized delta rule for multilayered networks, unlike the convergence theorem for perceptrons, does not guarantee that a solution will be found for all solvable input-output (environment-behavior) relations, it is reported that “our analyses and results have shown that, as a practical matter, the [back-]propagation scheme leads to solutions in virtually every case” (Rumelhart, Hinton, & Williams, 1986, p. 361).

The *PDP* volumes contain applications of adaptive networks to a variety of content areas. To give some sense of the approach, consider the following example of the use of an adaptive network to simulate the formation of a discriminative stimulus class (Goldiamond, 1962), what, in the vernacular, is referred to as a concept (McClelland & Rumelhart, 1986, pp. 170–215). Suppose that a child sees a number of dogs that differ somewhat from one another, but that share certain features. Many dogs are brown, but there are occasional black dogs or white dogs; most dogs have a tail, but some do not; and so on. In the adaptive network used to simulate these conditions, there were 16 features with each of the features represented by an input unit. An input unit was activated if the feature was present and was not activated if the feature was absent. Suppose further that the child’s parents say “dog” and reinforce the child’s verbal response, “dog,” in the presence of canine input patterns of the features of particular dogs, while other responses are reinforced in the presence of noncanine input patterns. In the simulation, there were eight output units, with a particular pattern of activation of the units corresponding to the verbal response, “dog.” Other output patterns corresponding to other verbal responses (e.g., “cat”) might be reinforced in the presence of different input patterns.

The goal of the simulation was to determine whether the connection weights in the network could be changed so that only canine input patterns evoked the output pattern corresponding to the verbal response, “dog.” The effect of reinforcement was simulated by adjusting each connection weight in proportion to the discrepancy between the output pattern

on that trial and the “dog” output pattern. The network was repeatedly exposed to 50 different canine input patterns and 100 input patterns corresponding to other stimuli. After each exposure, the connection weights were adjusted so that when one of the canine input patterns occurred, the discrepancy between the obtained output pattern and the “dog” pattern was reduced. The cumulative effect of these adjustments was that the network “recognized” canine input patterns; that is, canine patterns applied to the input units of the network caused the “dog” pattern to appear on the output units.

Three characteristics of the discriminative stimulus class simulated by the adaptive network are worthy of special comment. First, the stimulus class had “fuzzy boundaries” (Rosch & Mervis, 1975). That is, no single canine feature needed to be present in an input pattern in order for the “dog” output pattern to be activated. For example, suppose that one of the input features corresponded to *has four legs*. Although the connection weights linking the unit activated by this feature with the output pattern corresponding to “dog” might be relatively strong, input patterns not containing this feature would also be able to evoke the “dog” response. Thus a dog unfortunate enough to have lost a leg through an accident, but possessing other canine features, would still be called a dog.

Second, the values of the connection weights by which input units were linked to output units depended on the particular examples of canine and noncanine inputs that were used to train the network and on the order in which the inputs were applied. Thus, the strengths of the weights in the network were path-dependent—that is, determined by the details of the selection history of the network. Networks, like living organisms, reflect their unique selection histories. To illustrate, the pathways activated by a feature such as *has four legs* would have large connection weights in a network that had been trained to distinguish dogs from fire hydrants, but would have smaller and more complexly arrayed weights in a network that had been trained to distinguish dogs from cats. Fire hydrants do not have legs, but cats do; hence, the feature, *has four legs*, would be helpful in distinguishing dogs from hydrants but not from cats. (Note that path dependence is inherent in all his-

torical-science approaches to complexity, and is not unique to the PDP and behavior-analytic approaches. For example, production systems—whatever their other characteristics—also have this property, as persuasively demonstrated in accounts of the acquisition of verbal behavior, e.g., Anderson, 1983.)

Third, by the end of training, the network was capable of responding more strongly with the “dog” output pattern to some new input patterns than to any of the input patterns to which it had been exposed. The new input patterns to which the network responded more vigorously might be described as the more “typical” canine input patterns. Thus, the network responded *as if* it had formed a prototype of the canine inputs, where a prototype may be thought of as the most common combination of input features (Posner & Keele, 1968). Note that the prototype is not a “thing” that is stored at some place within the network; it is not an “ideal representation of reality” that is waiting to be retrieved by the stimulus. Networks, and the living organisms whose functioning they are intended to simulate, act *as if* there were prototypes, but what exist are sets of connection weights and synaptic efficacies, respectively. Responding *as if* there were a prototype is simply how a trained network or an experienced organism functions after training.

The PDP Approach as Interpretation in Historical Science

To qualify as an interpretation in historical science, an account must draw upon only principles derived from findings established through prior experimental analyses. Interpretations are consumers, not producers, of principles. How well does the PDP approach conform to this criterion?

To begin, the PDP approach—as a “neurally inspired” effort at interpretation—must draw upon experimental analyses of the neurosciences as well as of behavior. Experimental analyses of neuroscience and of behavior are complementary undertakings, both of which contribute to understanding the functioning of the organism. As Skinner (1938) has noted: “What is generally not understood by those interested in establishing neurological bases is that a rigorous description at the level of behavior is necessary for the demonstration of a neurological correlate” (p. 422). “. . .

I am not overlooking the advance that is made in the unification of knowledge when terms at one level of analysis are defined ('explained') at a lower level" (p. 428). Although our focus is upon the PDP approach as a means for exploring the implications of findings from the experimental analysis of behavior, a few comments on the relation of PDP to the neurosciences are in order.

PDP and the neurosciences. Reaching an accommodation between the PDP approach and the neurosciences has been the subject of considerable discussion and controversy both within the *PDP* volumes (Crick & Asanuma, 1986, pp. 333–371; McClelland & Rumelhart, 1986, pp. 327–331; Norman, 1986, pp. 531–546) and elsewhere (e.g., Smolensky, 1988). A major focus of contention has been the lack of correspondence between the structure of PDP networks and the structure of the nervous system (cf. Segal, 1988). In general, the interconnections postulated within networks have not been closely guided by neuroanatomical findings (Crick & Asanuma, 1986, pp. 370–371; Minsky & Papert, 1988, p. 266). Only one inconsistency is examined here, but it is one that pertains to the central issue in adaptive networks—the means by which reinforcers adjust the connection weights within networks.

In order for the nervous system to implement back-propagation as a means of adjusting the connection weights, large numbers of specific back-connections from the output units to units in all of the earlier layers of the network are required. In fact, there is no neuroanatomical evidence that such rich back-connections exist. On the contrary, the neural systems mediating selection by reinforcement appear to be nonspecific systems that project diffusely within the brain areas they serve (for reviews, see Carlson, 1986, and Olds & Fobes, 1981). How can these diffuse, reinforcer-activated systems alter the specific connections between the input and output units that mediate a particular environment–behavior relation?

An adaptive network in which a diffuse reinforcement system is capable of adjusting the connection weights is shown in Figure 2. We call a network of this architecture a *selection network*. Suppose that the environment places a given pattern on the input units of the selection network. The activated input

units are indicated by filled circles in Figure 2. Dependent on the values of the connection weights prior to the input, an activated input unit will then probabilistically activate some of the hidden units with which it is connected and they, in turn, will activate some of the output units. Paths that might be activated by the input pattern shown in Figure 2 are indicated by the heavier lines. Some activated paths are "dead ends" in the sense that they are not constituents, in this instance at least, of paths that activate output units upon which the reinforcer is dependent. However, other paths do lead to the critical output unit(s), and their activation causes the reinforcer to occur. In Figure 2, the critical output unit is shown as a filled circle. When this output unit is activated, causing the reinforcer to be presented, a diffuse signal is sent throughout the network. The effect of this diffuse signal is to strengthen all of those connections *that happen to be active at that moment*. Some of the active connections will be part of "dead-end" paths, and the strengthening of these paths may not benefit the appropriate functioning of the network. However, the set of active paths must necessarily also include some paths that activate the crucial output unit(s) because, without their involvement, the reinforcer would not have occurred. Over time, it is the connection weights of these latter pathways that will be strengthened most reliably.

Adaptive networks and behavior analysis. There are a number of striking consistencies between the behavior-analytic account of the acquisition of environment–behavior relations and the interpretation of those relations by means of selection networks. First, an operant is conventionally defined as a class of responses, all of which have a partially common effect on the environment (cf. Reynolds, 1968, p. 17; Skinner, 1935). In an adaptive network with a diffusely projecting reinforcement system, selection produces a number of different output patterns but all of them include activation of the crucial output unit(s) upon which the reinforcer is dependent. The class of output patterns activated by the input pattern is analogous to the class of responses that constitute the operant.

Second, with operant as contrasted to respondent conditioning, the critical response class is said to be emitted rather than elicited.

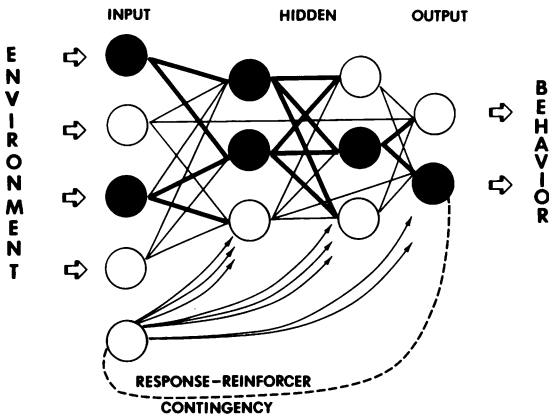


Fig. 2. A selection network consisting of input, hidden, and output units. The solid circles represent units that are activated as the result of the environmental input at that moment. Because of an environmentally mediated contingency between a particular output of the network and a reinforcer, the occurrence of the designated output pattern causes the reinforcing stimulus to be presented. The reinforcing stimulus activates an input unit having diffuse projections throughout the network. The diffuse projections increase the strength of a connection to the extent that the connection is from an active unit and the connection terminates on an active unit.

That is, the environment acting on the organism permits the response to occur. At other times or under other organismic conditions, the same environment might occasion other responses. Similarly, an input pattern to a network does not so much elicit the required output as the network *permits* that output to occur. Depending upon the initial state of the network and its selection history, the connection weights at that moment determine the range and probability of the specific environment-behavior relations that it can mediate.

Third, by whatever means the designated output (operant) occurs, it is those means that are strengthened. Thus, conditioning with both selection networks and living organisms is essentially superstitious in nature (Skinner, 1948). At heart, operant conditioning is a procedural arrangement whereby the presentation of the reinforcer is dependent on the occurrence of a specified response, whereas respondent conditioning is an arrangement whereby the reinforcer is dependent on a specified stimulus that already evokes a particular response (cf. Donahoe, Crowley, Millard, & Stickney, 1982). Both of these pro-

cedures may be simulated with a selection network having a diffuse projection system activated by the reinforcer, and initial work indicates that such networks are competent to learn a variety of environment-behavior relations (e.g., Barto & Arandan, 1985).

Fourth, there are important similarities between the behavior of organisms and the output of adaptive networks *after* selection has occurred. After an organism's behavior has been subject to differential conditioning in which a response has been reinforced in the presence of one environment and a different response during a second environment, intermediate test stimuli (following intradimensional training) or compound test stimuli (following interdimensional training) occasion a *mixture* of the two discriminated operants (e.g., Bickel & Etzel, 1985; Donahoe & Wessells, 1980, pp. 176-196). That is, the test stimuli evoke only the responses that have been reinforced in the presence of the training stimuli, and not new responses. After an adaptive network has been differentially conditioned, an analogous phenomenon may occur. When a new input pattern is applied to the network, the output pattern tends toward one of the two trained output patterns. This property of adaptive networks is the result of what are called attractor dynamics (Jordan, 1986; Sejnowski, 1986, p. 389; Smolensky, 1986, pp. 424-429).

ADAPTIVE NETWORKS AND THE EXPERIMENTAL ANALYSIS OF BEHAVIOR

The accounts of differential conditioning provided by adaptive networks and experimental-analytic findings are strikingly and persuasively congenial. Because, from the behavior-analytic perspective, complex behavior is the cumulative product of extensive differential conditioning, adaptive networks provide a potentially powerful means for formal interpretation.

Some of the possible contributions of behavior-analytic findings to adaptive-network interpretations may be illustrated by returning to the central issue of reinforcement. Regardless of whether the connection weights in the network are adjusted by back-propagation or by a nonspecific projection system

as we have proposed, the event that triggers the adjustment procedure must be well specified. According to the back-propagation procedure, the weights are adjusted as a function of the discrepancy between the obtained output pattern and the appropriate output pattern. Thus, in order for back-propagation to occur, the environment must provide detailed information about the appropriate output pattern. This precondition is often denoted by saying that back-propagation requires a "teacher." Quite reasonably, biologically oriented critics of the PDP approach have commented that this assumption is inconsistent with the conditions under which most complex behavior is acquired (e.g., Minsky & Papert, 1988, p. 264; Segal, 1988, p. 1107). More often, complex behavior—such as verbal behavior—is discovered, not instructed. However, because the environment usually does not provide an all-knowing "teacher," some of these same critics have mistakenly concluded that selection by consequences cannot play a major role in modulating synaptic efficacy in the nervous system. (Much of the difficulty arises from a failure on the part of both adaptive-network theorists and their critics to distinguish between contingency-shaped and rule-governed behavior, but that important matter will not be pursued here; see Skinner, 1974).

The belief that naturally occurring contingencies are insufficient to bring about complexity has been a long-standing, but misconceived, impediment to selectionist accounts of complexity in the historical sciences. This mistaken belief was reflected in the criticisms of Darwin's contemporaries who argued that, although artificial selection by animal husbandrymen could produce progressive changes, natural contingencies could not do so without the intervention of a Designer. The same mistaken belief is reflected in the present-day criticism of linguists that the verbal environment is too impoverished for "language" to be acquired without the intervention of a Language Acquisition Device (Chomsky, 1980; see also Pinker & Mehler, 1988).

Within behavior analysis, the problem of accounting for the emergence of complex environment-behavior relations from the operant level of simpler behavior and in the absence of an all-knowing teacher falls within the extensive literature on shaping, chaining,

and fading. In shaping, the response topography necessary for the occurrence of a reinforcer progressively approximates the topography ultimately required for the reinforcer. With adaptive networks, this arrangement may be simulated by progressively changing the criterion output pattern. In chaining, responding alters the stimuli available to the organism with the result that these new stimuli serve as conditioned reinforcers with respect to the responses that produce them and as discriminative stimuli with respect to subsequent responses in the sequence. With adaptive networks, this arrangement may be simulated by having the output pattern affect the subsequent input pattern to the network. In this way, a feedback loop is implemented that is mediated by the environment rather than by recurrent connections within the network itself. We are unaware of any work within the PDP framework that exploits the literature on shaping and chaining.

Fading is also critically important to the acquisition of complex environment-behavior relations within the behavior-analytic framework. In fading, the complex stimuli that ultimately control behavior are progressively approximated as training proceeds. With adaptive networks, fading may be simulated by progressively changing the input pattern to the network in the direction of the final input pattern. There has been only some very preliminary work on fading with adaptive networks. In the one case known to us, the input patterns that were to be discriminated were initially very dissimilar, but were progressively shifted toward the highly similar patterns required by the final discrimination. Using fading, the adaptive network achieved criterion performance with only 25% of the trials that were required when the highly similar input patterns were used throughout training (Jacobs, 1988)!

Some Shared Orienting Attitudes

We have seen that both behavior analysis and adaptive-network theory deny the functionality of the molar constructs of information processing and assert the centrality of the guidance and selection of complex behavior by the environment. Moreover, experimental analysis appears to have much to offer adaptive-network theory in the simulation of selection; namely, findings concerning the behavioral and

neural processes responsible for reinforcement. Likewise, adaptive-network theory appears to have much to offer experimental analysis; namely, a means for integrating these findings so that their implications for the interpretation of complex behavior may be pursued more precisely. These specific mutual benefits are possible because both the behavior-analytic and adaptive-network approaches share a number of general orienting attitudes toward the origins of complex behavior. We conclude this paper by describing three of these shared orientations—the preeminence of the environment, the ubiquity of multiple causation, and the disenfranchisement of consciousness.

Preeminence of the environment. Perhaps, the characteristic that most clearly distinguishes the behavior-analytic approach from information-processing approaches to complex behavior is the effort to ferret out the environmental antecedents of complex behavior and to resist the temptation to attribute behavior to inferred organismic events that are beyond the reach of the environment. Although the organism is the locus of environmental action, it is the environment, and not the organism, that is the initiator and shaper of behavior (Skinner, 1974). Adaptive-network theory takes a similar stance; complex behavior occurs when the environment activates the input units of the network. Rumelhart, Hinton, and McClelland, in their summary of “eight major aspects of a parallel distributed processing model,” list “. . . an *environment* (sic) within which the system must operate” (1986, p. 46). The reliance upon the environment (i.e., upon “experience”) as the origin of complexity is especially clear in the treatment of verbal behavior from a PDP perspective: “the greater the amount of experience, the more independent the system should be from its start state and the more dependent it should be on the structure of its environment. . . . To the extent that stored knowledge is assumed to be in the form of explicit, inaccessible rules of the kind often postulated by linguists, . . . it is hard to see how it could ‘get into the head’ of the newborn” (Rumelhart & McClelland, 1986, p. 142). (See D. C. Palmer, 1986, for an extended discussion of this point.) Adaptive-network theory, in agreement with behavior analysis and only a few other contemporary views (e.g., Gibson, 1979; see also Costall, 1984),

regards the environment as providing a richer source for shaping complex behavior than is appreciated by the information-processing approach (see Rumelhart, Hinton, & McClelland, 1986, p. 54; Rumelhart, Smolensky, McClelland, & Hinton, 1986, p. 39 ff).

The emphasis upon the environment as the shaper both of behavior and adaptive networks should not be taken to mean that either behavior analysis or adaptive-network theory denies the crucial contribution of the organism as the locus of the cumulative effects of prior environmental selection. As Skinner has noted, “The environment made its first great contribution during the evolution of the species, but it exerts a different kind of effect during the lifetime of the individual, and the combination of the two effects is the behavior that we observe at any given time” (Skinner, 1974, p. 19; see also Skinner, 1966a, 1984). Similarly, in *PDP* it is remarked, “Some have argued that since we claim that human cognition can be explained in terms of PDP networks and that the behavior of lower animals can also be described in terms of such networks we have no principled way of explaining why rats are not as smart as people.” Does this “criticism” sound familiar? The criticism is answered in much the same way that behaviorists would answer it, except for a few notable differences in technical vocabulary: “We are not claiming, in any way, that people and rats and all other organisms start out with the same prewired hardware. . . . But there must be another aspect to the difference between rats and people as well. This is that the human environment includes other people and the cultural devices that have been developed to organize their thinking processes” (Rumelhart & McClelland, 1986, p. 143; cf. Skinner, 1981).

Multiple causation. A corollary of the proposition that behavior is the product of the combined effect of the contemporary environment acting on an organism that has been changed by the cumulative effect of selection by prior environments is the following: Because both prior and present environments are complex, any given behavior—particularly complex behavior, which is the product of a prolonged history of selection—is likely to be under the combined control of many different environmental stimuli. Multiple con-

trolling variables are clearly evident in verbal behavior, where slips of the tongue and the pen often point to the complexity of the antecedents of a given response. For example, the failure of the subject and predicate to agree in number in the sentence, "The assemblage of feathers *are* beautiful," probably reflects such facts as "feathers" is closer to the verb than is "assemblage" and "feathers are beautiful" has been more commonly uttered (and reinforced) than "the assemblage is beautiful." Indeed, the pervasiveness and importance of multiple controlling variables for the understanding of complex behavior led Skinner (1957) to entitle one of the chapters in *Verbal Behavior*, "Multiple Causation."

Adaptive networks inherently implement multiple causation. That is, the output pattern produced by the network is dependent upon *all* of the input units that are activated by the contemporary environment and *all* of the connection weights within the network, which, in turn, reflect the cumulative action of *all* of the previous environments. The use of adaptive networks to simulate the effects of multiple controlling variables on complex behavior is in its infancy, as in the treatment of "sentence processing" in *PDP* (e.g., McClelland & Kawamoto, 1986, pp. 272–325). Nevertheless, certain conclusions have been reached that are congenial to a behavioral analysis.

As two examples of such conclusions, adaptive-network theory questions the status of certain apparently verbal constructions as genuine instances of verbal behavior and the status of response topography by itself as a fruitful means for studying verbal behavior. First, note the treatment in *PDP* of center-embedded sentences, one of the constructions thought by linguists to require recursive mechanisms in accounts of sentence processing. McClelland and Kawamoto comment: "... people cannot parse such sentences without the use of very special strategies, and do not even judge them to be acceptable. Consider, for example, the 'sentence': ... The man who the girl who the dog chased liked laughed. ... sentences in natural language are simply not structured in this way. Perhaps, then, the search for a model of natural language processing has gone down the garden path, chasing a recursive white rabbit" (1986, pp. 323–324). Skinner's earlier comments about

such constructions are to similar effect: "Perhaps there is no harm in playing with sentences in this way, ... but it is still a waste of time, particularly when the sentences thus generated could not have been emitted as verbal behavior" (Skinner, 1974, p. 109).

Second, behavior analysis has long rejected response topography as a sufficient basis for either an adequate experimental analysis (Skinner, 1935) or interpretation (Skinner, 1957) of behavior. Regarding verbal behavior Skinner has stated, "What is needed—and what the traditional 'word' occasionally approximates—is a unit of behavior composed of a response of identifiable form functionally related to one or more independent variables. ... In this way we may distinguish between the operant *fast* in which the controlling variable is shared by the operant *speedy* and the operant *fast* in which the controlling variable is similar to that in the operant *fixed*" (Skinner, 1957, pp. 20–21). "Those who have confused behaviorism with structuralism, in its emphasis on form or topography, have complained that it ignores meaning. ... Meaning is not properly regarded as a property either of a response or a situation but rather of the contingencies responsible for both the topography of behavior and the control exerted by stimuli" (Skinner, 1974, pp. 100–101). In the treatment of "words" by adaptive networks, a similar point is made. "We will probably all agree that there are different readings of the word *bat* in the sentences *The bat hit the ball* and *The bat flew round the cave*. ... Different readings of the same word are just different patterns of activation; really different readings, ones that are totally unrelated ... the two readings of *bat* simply have very little in common" (McClelland & Kawamoto, 1986, pp. 314–315). They have "very little in common" because the "different patterns of activation" are produced by different input patterns to the network. That is, the controlling stimuli for responses of the same topography are different and, therefore, the responses are members of different operants.

Consciousness. The orienting attitude that behavior analysis and adaptive-network theory share toward consciousness might not appear, at first glance, to be fundamental to their approaches to complex behavior. In fact, an argument can be made that it is this shared

attitude that is central to their common approach to complex behavior.

The information-processing approach, either explicitly or implicitly, is prone to the belief that behavior taken as evidence of consciousness occupies a special status with respect to other behavior. Although a number of characteristics are often mentioned as giving consciousness its special status, the one of concern here is the belief that there is a fairly direct and straightforward relation between what we think about what we are doing and what we are indeed doing. That is, consciousness is regarded as a valid, or, at a minimum, a useful indicator of the behavioral and physiological processes that intervene between the stimulating environment and the occurrence of other behavior. The experimenter's consciousness, if not the subject's, is commonly accorded this ability. Sometimes the assumption is explicit, as when computer simulations of complex behavior are designed to mimic the introspections of experts at a complex task requiring expertise (Newell & Simon, 1961); more often it is implicit, as when an experimenter introspects under the partial control of measures of the subject's behavior and infers the cognitive processes said to underlie the subject's behavior.

Behavior analysis has provided extensive verbal interpretations (Skinner, 1945, 1964, 1974) and some organic simulations (e.g., Lubinski & Thompson, 1987) of consciousness and its origins. Consciousness (i.e., verbal behavior under the control of intraorganismic events) is a defective indicator of those events for two classes of reasons. First, the social community faces uncircumventable impediments to fostering such discriminations (Skinner, 1945, 1964). Second, "There has been no opportunity for the evolution of a nervous system which would bring some very important parts of the body under that control (p. 242)... Introspection has had to use whatever [neural] systems were available, and they have happened to be systems which made contact with those parts of the body that played a role in its internal and external economy... [The verbal system] does not make contact with that vast nervous system that mediates... behavior. [It] does not because there are no nerves going to the right places... The brain plays an extraordinary role in behavior but not as the object of that

special behavior called knowing. We can never know through introspection what the physiologist will eventually discover with his special instruments (pp. 238–239)... What is felt or seen through introspection is only a small and relatively unimportant part of what the physiologist will eventually discover" (Skinner, 1974, p. 274).

Some proponents of adaptive-network theory have come to essentially the same conclusion on this point. In summarizing the PDP approach, Norman says "Introspection... is based upon observation of the outputs of a subconscious (PDP) system. As a result, introspections are only capable of accurate descriptions of system states. Because there is no information available about how the state was reached, introspection cannot give reasons for the resulting states" (1986, p. 544). He, then, advocates a close future relation between cognitive psychology and the neurosciences. Minsky and Papert (1988) concur: "What any distributed network learns is likely to be quite opaque to other networks connected to it (p. 274)... It is because our brains primarily exploit [adaptive networks] that we possess such *small* degrees of consciousness, in the sense that we have so little insight into the nature of our own conceptual machinery... What appear to us to be direct insights into ourselves must be rarely genuine and usually conjectural... Reflective thought is the lesser part of what our minds do" (p. 280).

Certain formal analyses have also undermined consciousness as bearing a special, privileged relation to other behavior. These analyses have shown that, with the customary nonlinear activation mechanisms of adaptive networks, there can be, in principle, no invariant relation between the output of a network and the events within the network that produced that output (Smolensky, 1986, pp. 422–424). That is, there is no isomorphism between the output pattern of the network and activity within the network (i.e., unit activity). Moreover, "... if two models were started in corresponding [output] states and given corresponding inputs, they would *not* continue to stay in corresponding states" (Smolensky, 1986, p. 424). This last is a general characteristic of historical science and constitutes a strong formal argument against conventional information-processing models.

Knowledge of the behavioral output and environmental input does not sufficiently constrain conjectures about intraorganismic (or intranetwork) events to warrant inferences about such events from these observations alone. In this important conclusion, which strikes at the very heart of the information-processing enterprise, behavior analysis (Skinner, 1974) and adaptive-network theory concur. (For related discussions of difficulties in inferring antecedents from knowledge of only the outputs of a system, see Anderson, 1978, Gleick, 1987, and Churchland & Sejnowski, 1988.)

SOME CLOSING THOUGHTS

A dominant theme of our reactions to *PDP* has been that behavior-analytic and adaptive-network approaches share much conceptual common ground and have much to gain from one another in the effort to interpret complex behavior. Experimental analyses, both of behavior and neuroscience, provide the findings necessary for the construction of computer simulations that conform to the requirements of historical science. Computer simulations based on experimental analyses provide a powerful means for the formal interpretation of complex behavior, one that is superior to verbal interpretation in its ability to implement the intricacies of environmental selection in a precise and expeditious manner.

Skinner has long criticized a fascination with what he dubbed the "Conceptual Nervous System" (Skinner, 1938, p. 421), arguing instead for a science of behavior studied at its own level. His position, we believe, has been vindicated; the experimental analysis of behavior has advanced with little concern for physiological mechanisms. However, the field finds itself today, like Darwinism before it (Catania, 1987), a science whose accomplishments have outstripped its acceptance by the general public and by much of the scientific community. Skinner's verbal interpretations of complex behavior, however compelling to the prepared reader, have failed to excite much interest outside the field. Indeed, it is commonly held by those who call themselves cognitive scientists that behavioral interpretations are impoverished or even inadequate in principle (Bever, Fodor, & Garrett, 1968; Chomsky, 1959). It is our belief

that computer simulations using adaptive networks can serve an important polemical function by demonstrating, through formal interpretation, the power of behavioral principles.

However, we would be remiss if we did not indicate the antipathy toward behavior analysis that is expressed in *PDP*. Although Skinner's views on many issues predate and parallel those expressed in *PDP*, none of his work appears among the many citations in the bibliography. Moreover, in the one instance in which behavior analysis is considered, a vigorous effort is made to distance work on adaptive networks from behavior analysis. To wit, in a section entitled "Some objections [emphasis ours] to the *PDP* approach," the following statements appear: "A related claim that some people have made is that our models appear to share much in common with behaviorist accounts of behavior. While they do involve simple mechanisms of learning, there is a crucial difference between our models and the radical behaviorism of Skinner and his followers. In our models, we are explicitly concerned with the problem of intermodal representation and mental processing, whereas the radical behaviorist explicitly denies the scientific utility and even the validity of the consideration of these constructs. . . . Our models must be seen as completely antithetical to the radical behaviorist program and strongly committed to the study of representations and process" (Rumelhart & McClelland, 1986, p. 121).

Behavior analysts will recognize such criticisms as ill-informed. As has occurred all too often, the editors of *PDP* have mistakenly equated the behavior-analytic position that behavioral data do not sufficiently constrain inferences about underlying microbehavioral and physiological processes to permit fruitful conjectures about such processes with the "black-box" position that such processes are of no relevance to a science of behavior (Skinner, 1974, pp. 233-237). To the degree that "intermodal representation" and "mental processing" denote microbehavioral events or events in the real nervous system, behaviorists have no objection to the inclusion of such terms in a complete account of the functioning of the organism. To the contrary, as already noted, Skinner has long acknowledged "the advance that is made in the unification of

knowledge when terms at one level of analysis are defined ('explained') at a lower level" (Skinner, 1938, p. 428). "The physiologist of the future will tell us all that can be known about what is happening inside the behaving organism. His account will be an important advance over a behavioral analysis, because the latter is necessarily 'historical'—that is to say, it is confined to functional relations showing temporal gaps. . . . What he discovers cannot invalidate the laws of a science of behavior, but it will make the picture of human action more nearly complete" (Skinner, 1974, pp. 236–237).

What Skinner rejected as futile was the attempt to draw strong inferences about physiology from behavior or, more generally, to draw inferences about physical events taking place at a lower level from the observation of physical events occurring at a higher level. If lower level events are to be understood, then those events must be studied directly. (See Donahoe & Palmer, 1988, for a discussion of this point as it relates to the concept of inhibition.) With respect to the relation between behavior and physiology, Skinner was asserting nothing more remarkable than, for example, that physiology is not the way to study chemistry or that chemistry is not the way to study physics. As applied to the present issue, the behavior-analytic position is simply that if one is interested in the architectures of adaptive networks, which are crucial determinants of the environment-behavior relations that a network can mediate (Minsky & Papert, 1988, p. 266), then behavior-analytic findings must be supplemented by the relevant findings from neuroscience. The demands of interpretation in historical science are not met when one attempts to infer underlying processes from phenomena whose explanation is sought in terms of those very same processes. That way lies the chaos of circular reasoning.

We began with the comment that behavior analysts should regard adaptive-network theory as an ally in the effort to interpret complex behavior. A sufficient reason for taking this position, although we have indicated many better reasons, is the Machiavellian dictum, "My enemy's enemy is my friend." We may hope that adaptive-network theorists, faced with the inevitable onslaughts from mainstream cognitive psychology and linguistics,

will be more receptive to interactions with behavior analysts on the grounds that "A friend in need is a friend indeed!" The better reason for such cooperation is that adaptive-network theory, if it is to be a means for interpretation in historical science, must be guided by principles and findings from behavior analysis and neuroscience.

REFERENCES

- Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, **85**, 249–277.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Barto, A. G., & Arandan, P. (1985). Pattern recognizing stochastic automata. *IEEE Transactions on Systems, Man and Cybernetics*, **15**, 360–375.
- Bechtel, W. (1985). Contemporary connectionism: Are the new parallel distributed processing models cognitive or associationist? *Behaviorism*, **13**, 53–61.
- Berko, J. (1958). The child's learning of English morphology. *Word*, **14**, 150–177.
- Bever, T. G., Fodor, J. A., & Garrett, M. (1968). A formal limitation of associationism. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 582–585). Englewood Cliffs, NJ: Prentice-Hall.
- Bickel, W. K., & Etzel, B. C. (1985). The quantal nature of controlling stimulus-response relations as measured in tests of stimulus generalization. *Journal of the Experimental Analysis of Behavior*, **44**, 245–270.
- Campbell, D. T. (1974a). 'Downward causation' in hierarchically organized biological systems. In F. J. Ayala & T. Dobzhansky (Eds.), *Studies in the philosophy of biology* (pp. 179–186). London: Macmillan.
- Campbell, D. T. (1974b). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper* (Vol. 1, pp. 413–463). (*The library of living philosophers*, Vol. 14). LaSalle, IL: Open Court.
- Carlson, N. R. (1986). *Physiology of behavior* (3rd ed.). Boston: Allyn & Bacon.
- Catania, A. C. (1987). Some Darwinian lessons for behavior analysis: A review of Bowler's *The eclipse of Darwinism*. *Journal of the Experimental Analysis of Behavior*, **47**, 249–257.
- Catania, A. C., & Cerutti, D. T. (1985). Some nonverbal properties of verbal behavior. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 185–211). Hillsdale, NJ: Erlbaum.
- Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language*, **35**, 26–58.
- Chomsky, N. (1980). *Rules and representations*. New York: Columbia University Press.
- Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, **242**, 741–745.
- Costall, A. P. (1984). Are theories of perception necessary? A review of Gibson's *The ecological approach to visual perception*. *Journal of the Experimental Analysis of Behavior*, **41**, 109–115.
- Crick, F. H. C., & Asanuma, C. (1986). Certain aspects

- of the anatomy and physiology of the cerebral cortex. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 333-371). Cambridge, MA: MIT Press.
- Dawkins, R. (1986). *The blind watchmaker*. Harlow, Essex: Longman.
- Donahoe, J. W. (1984). Skinner—The Darwin of ontogeny? *Behavioral and Brain Sciences*, 7, 487-488. (Commentary)
- Donahoe, J. W., Crowley, M. A., Millard, W. J., & Stickney, K. A. (1982). A unified principle of reinforcement: Some implications for matching. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 493-521). Cambridge, MA: Ballinger.
- Donahoe, J. W., & Palmer, D. C. (1988). Inhibition: A cautionary tale. *Journal of the Experimental Analysis of Behavior*, 50, 333-341.
- Donahoe, J. W., Palmer, D. C., & Carlson, N. C. (in press). *Complex human behavior: A biobehavioral approach*. Boston: Allyn & Bacon.
- Donahoe, J. W., & Wessells, M. G. (1980). *Learning, language, and memory*. New York: Harper & Row.
- Epstein, R. (1984). Simulation research in the analysis of behavior. *Behaviorism*, 12(2), 41-59.
- Epstein, R., Kirshnit, C. E., Lanza, R. P., & Rubin, L. C. (1984). 'Insight' in the pigeon: Antecedents and determinants of an intelligent performance. *Nature*, 308, 61-62.
- Epstein, R., Lanza, R. P., & Skinner, B. F. (1980). Symbolic communication between two pigeons (*Columba livia domestica*). *Science*, 207, 543-545.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Gleick, J. (1987). *Chaos: Making a new science*. New York: Viking.
- Goldiamond, I. (1962). Perception. In A. J. Bachrach (Ed.), *Experimental foundations of clinical psychology* (pp. 280-340). New York: Basic Books.
- Hinson, J. M., & Staddon, J. E. R. (1983). Hill-climbing by pigeons. *Journal of the Experimental Analysis of Behavior*, 39, 25-47.
- Jacobs, R. A. (1988). *Initial experiments on constructing domains of expertise and hierarchies in connectionist networks*. Department of Computer and Information Science Technical Report, University of Massachusetts, Amherst, MA.
- Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach*. Institute for Cognitive Science Technical Report No. 8604, University of California, San Diego, CA.
- Klopf, A. H. (1988). A neuronal model of classical conditioning. *Psychobiology*, 16, 85-125.
- Köhler, W. (1925). *The mentality of apes* (2nd ed., E. Winter, Trans.) New York: Harcourt, Brace, & World.
- Lubinski, D., & Thompson, T. (1987). An animal model of the interpersonal communication of interoceptive (private) states. *Journal of the Experimental Analysis of Behavior*, 48, 1-15.
- Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance*. Cambridge, MA: Belknap Press.
- McClelland, J. L., & Kawamoto, A. H. (1986). Mechanisms of sentence processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 272-325). Cambridge, MA: MIT Press.
- McClelland, J. L., & Rumelhart, D. E. (1986). On learning the past tense of English verbs. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 170-215). Cambridge, MA: MIT Press.
- McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Vol. 1* (pp. 3-44). Cambridge, MA: MIT Press.
- McClelland, J. L., Rumelhart, D. E., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models*. Cambridge, MA: MIT Press.
- Miller, L. (1986). Behaviorism and the new science of cognition. *Psychological Record*, 38, 3-18.
- Minsky, M. L., & Papert, S. A. (1969). *Perceptrons*. Cambridge, MA: MIT Press.
- Minsky, M. L., & Papert, S. A. (1988). *Perceptrons* (expanded ed.). Cambridge, MA: MIT Press.
- Munro, P. W. (1986). State-dependent factors influencing neural plasticity: A partial account of the critical period. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 471-502). Cambridge, MA: MIT Press.
- Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science*, 134, 2011-2017.
- Nilsson, N. J. (1965). *Learning machines: Foundations of trainable pattern-classifying systems*. New York: McGraw-Hill.
- Norman, D. A. (1986). Reflections on cognition and parallel distributed processing. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 531-546). Cambridge, MA: MIT Press.
- Olds, M. E., & Fobes, J. L. (1981). The central bases of motivation: Intracranial self-stimulation studies. *Annual Review of Psychology*, 32, 523-574.
- Palmer, D. C. (1986). Chomsky's nativism: A critical review. In P. N. Chase & L. J. Parrott (Eds.), *Psychological aspects of language* (pp. 44-60). Springfield, IL: Thomas.
- Palmer, S. E. (1987). PDP: A new paradigm for cognitive theory. *Contemporary Psychology*, 32, 925-928.
- Pinker, S., & Mehler, J. (Eds.). (1988). *Connections and symbols*. Cambridge, MA: MIT Press.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353-363.
- Reynolds, G. S. (1968). *A primer of operant conditioning*. Glenview, IL: Scott, Foresman.
- Rosch, E., & Mervis, C. B. (1975). Family resem-

- blances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 573-605.
- Rosenblatt, F. (1962). *Principles of neurodynamics: Perception and the theory of brain mechanisms*. Washington, DC: Spartan Books.
- Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 45-76). Cambridge, MA: MIT Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318-362). Cambridge, MA: MIT Press.
- Rumelhart, D. E., & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 110-146). Cambridge, MA: MIT Press.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations*. Cambridge, MA: MIT Press.
- Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 7-57). Cambridge, MA: MIT Press.
- Savage-Rumbaugh, E. S. (1984). Verbal behavior at a procedural level in the chimpanzee. *Journal of the Experimental Analysis of Behavior*, *41*, 223-250.
- Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York: Columbia University Press.
- Segal, M. M. (1988). Review of *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*, edited by J. L. McClelland & D. E. Rumelhart. *Science*, *241*, 1107-1108.
- Sejnowski, T. J. (1986). Open questions about computation in cerebral cortex. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 372-389). Cambridge, MA: MIT Press.
- Shimp, C. P. (1969). Optimal behavior in free-operant experiments. *Psychological Review*, *76*, 97-112.
- Skinner, B. F. (1935). The generic nature of the concepts of stimulus and response. *Journal of General Psychology*, *12*, 40-65.
- Skinner, B. F. (1938). *The behavior of organisms*. New York: Appleton-Century.
- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, *52*, 270-277.
- Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology*, *38*, 168-172.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1964). Behaviorism at fifty. In T. W. Wann (Ed.), *Behaviorism and phenomenology* (pp. 79-108). Chicago: University of Chicago Press.
- Skinner, B. F. (1966a). The phylogeny and ontogeny of behavior. *Science*, *153*, 1203-1213.
- Skinner, B. F. (1966b). What is the experimental analysis of behavior? *Journal of the Experimental Analysis of Behavior*, *9*, 213-218.
- Skinner, B. F. (1974). *About behaviorism*. New York: Knopf.
- Skinner, B. F. (1977). Why I am not a cognitive psychologist. *Behaviorism*, *5*(2), 1-10.
- Skinner, B. F. (1981). Selection by consequences. *Science*, *213*, 501-504.
- Skinner, B. F. (1984). The evolution of behavior. *Journal of the Experimental Analysis of Behavior*, *41*, 217-221.
- Smolensky, P. (1986). Neural and conceptual interpretations of PDP models. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models* (pp. 390-431). Cambridge, MA: MIT Press.
- Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, *11*, 1-23.
- Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus*. Cambridge, MA: MIT Press.
- Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 444-459). Cambridge, MA: MIT Press.