

## Interpreting Verbal Behavior

John W. Donahoe  
Neuroscience and Behavior Program  
and  
Department of Psychology  
University of Massachusetts at Amherst

Skinner distinguished between two complementary aspects of science – experimental analysis and interpretation (Skinner, 1957, 1974; see also Donahoe, 1993; Donahoe & Palmer, 1989, pp. 25-129, Donahoe & Palmer, 1994; Palmer & Donahoe, 1992). Some phenomena occur under circumstances that permit essentially all of their determining antecedents to be manipulated and all of their relevant consequences to be measured. Research seeking to uncover the necessary and sufficient conditions for reinforcement using model laboratory preparations is directed at phenomena of this sort. However, many phenomena – and almost all complex phenomena – occur under circumstances that cannot meet these requirements. Phenomena of this latter sort are typically products of complex contingencies involving many biobehavioral processes acting concurrently over prolonged periods of time. Even if experimental analysis had successfully characterized all of the constituent processes, uncertainty would continue to enshroud the origins of these phenomena: The initial conditions and the sequence in which the selecting contingencies acted would remain at least partially unknown. To maintain that complex phenomena cannot be subjected to experimental analysis does not preclude them from being incorporated within science. However, for interpretation to be compelling, other methods must supplement experimental analysis.

Of all human behavior, verbal behavior is the most complex and presents the greatest challenge to interpretation. If verbal behavior is largely the province of interpretation then we may ask, "What methods of interpretation should be employed?" Although a number of approaches have merit, two are emphasized here: (a) supplementation of behavioral observations with observations of neural processes and (b) simulations of verbal behavior that are informed by experimental analyses of both behavior and neuroscience.

### **Interpretations Using Neuropsychological and Neural Observations**

Adult verbal behavior is the product of an extensive history of selection by both the ancestral and individual environments; i.e., by natural selection and by reinforcement. As such, the likelihood that most adult verbal behavior will ever be subjected to experimental analysis is vanishingly small. Faced with the insufficiency of experimental analysis, Skinner (1957) interpreted verbal behavior by appealing to covert behavior – such as some autoclitics – as mediators of observable environment-behavior relations. However, Skinner (1974) anticipated that future science would uncover the neural processes that implemented these mediating processes and, in that way, make our understanding of human behavior more nearly com-

plete. Neural processes and covert behavior are both products of the selection history but neural processes – unlike covert behavior – are potentially observable. Consider the following behavioral observations from a person suffering from an aphasia following brain damage. When asked whether the ceiling was "up" or "down," he could not respond appropriately. But, when asked whether he felt "up" or "down," he did so with ease (for other examples, see Donahoe, 1991; Donahoe & Palmer, 1994, pp. 309-312). From a behavior-analytic perspective, these observations are not troublesome. Topographically identical verbal responses (e.g., "down") can be members of different verbal operants. The same verbal response may be a constituent of one tact under the control of extra-organismic stimuli specifying spatial location and of a different tact under the control of intra-organismic stimuli specifying emotional state. (Note: The dependent measures in neuropsychological research may continue to be exclusively behavioral; only the independent variables need be neural.) The naturalness of the behavior-analytic interpretation contrasts with the difficulties these observations pose for normative cognitive/linguistic psychology: They must explain how the same item in the "lexicon" is "retrievable" under some circumstances but not others.

Observations at the neural level supplement the behavioral interpretation of aphasia. In the preceding case, damage to right-hemisphere sensory-association cortex impaired the first verbal operant but spared the second. Even a few years ago, attempts to correlate behavioral deficits in humans with neural damage had to await post-mortem examination. Now, the increasing precision of imaging techniques allows the measurement of neural damage and neural activity concurrently with behavioral observations. For example, functional magnetic-resonance imaging (fMRI) and positron-emission tomography (PET) can assess the locus of neural activity while the subject engages in a verbal task (e.g., Raichle, 1997). Recent behavior-analytic research using PET scans with autistic children, carried out by Julie Schweitzer at Emory University (personal communication), illustrates the productive interplay between behavioral and neural observations. Of course, care must be taken when interpreting structure/function correlations. A given verbal relation may endure when a usually important structure is damaged because other structures now mediate the relation. Contrariwise, a verbal relation may fail because the damaged structure interferes with the functioning of an intact structure that would otherwise mediate the relation. Modern imaging techniques often permit these alternatives to be disentangled by measuring not only damage to a structure but also activity in undamaged structures. These measures of neural activity can then be compared to activity in corresponding brain structures in control subjects.

Neural observations obtained from model preparations with nonhuman animals can also contribute to the interpretation of verbal behavior. For example, the biobehavioral processes involved in equivalence relations (Sidman & Table, 1982) are thought to play a central role in verbal behavior (e.g., Barnes & Hampton, 1997; Devaney, Hayes, & Nelson, 1986; Dugdale & Lowe, 1990; Hayes, 1989; Hayes & Hayes, 1992; Horne & Lowe, 1996). As one component of an equivalence relation, the sample stimulus and its associated comparison stimulus are interchangeable after matching-to-sample training. That is, after the sample stimulus S1 becomes discriminative for responding under the control of the comparison stimulus S2, the functions of S1 and S2 may be reversed. That is, S2 may provide the context for responding differentially to S1, thereby demonstrating symmetry. Findings that meet the criteria for equivalence relations have been interpreted to mean that a stimulus-stimulus relation forms between the sample and comparison stimulus (Sidman & Tailby, p. 22). But, the intra-organismic processes mediating the relations between environmental events cannot be observed at the behavioral level, only their presumed

behavioral effects. To make the processes mediating such relations observable, it is necessary to look inside the organism. In one such experiment, electrophysiological recordings indicated that training on an arbitrary matching-to-sample task using complex visual stimuli changed synaptic efficacies between neurons in a specific way: Following training, some cells in visual-association cortex could be activated by *either* the sample or the comparison stimulus. Without reinforcement for responding to these particular pairs of stimuli, cells that responded to either stimulus were not found (Sakai & Miyashita, 1991).

### Interpretations Using Neural-network Simulations

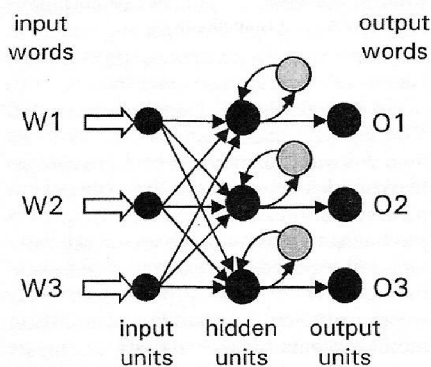
If the interpretation of verbal behavior is to make the most of the integration of behavior analysis and neuroscience, special tools of interpretation are required (Donahoe & Palmer, 1989). Ordinary language is sequential and discrete, whereas the biobehavioral processes that mediate complex behavior are overlapping and continuous. A subset of scientific verbal behavior – adaptive neural networks – has been specifically devised to capture the parallel, interacting, and dynamic character of multiple, time-varying processes (cf., Galbicka, 1992). General descriptions of neural networks have been given elsewhere (e.g., Donahoe, 1997; Donahoe, Burgos, & Palmer, 1993; cf. McClelland & Rumelhart, 1986; Bechtel & Abrahamsen, 1991), and will not be repeated here. Stated most generally, neural networks consist of interconnected units that reveal the emergent effects of fundamental biobehavioral processes acting over time. Two such effects are considered here – language acquisition and syntactic distinctions.

**Language acquisition.** Children are often said to acquire verbal behavior without the need for substantial reinforcement from others (e.g., Brown & Hanlon, 1970). This conclusion is overstated (e.g., Mørk, 1990), but is not the focus of the present discussion. Here, the focus is upon the ability of neural networks to also learn without immediate reinforcement from the external environment. In neural networks that simulate neural systems in living organisms, environmental reinforcers simultaneously strengthen connections along two sets of pathways: The first set mediates those environment-behavior relations that reliably precede the reinforcer. The second set arises from units in motor areas that mediate the reinforced response and projects to units in the neural reinforcing system. Speaking nontechnically, neural networks (and organisms) learn not only what to do to produce reinforcers but also how to reinforce their own activity (Donahoe & Palmer, 1994, pp. 96-99). Stated at the behavioral level, stimuli occurring prior to a reinforced response can function both

as discriminative stimuli and conditioned reinforcers (Keller & Schoenfeld, 1950).

Vocal responses are uniquely capable of capitalizing on the neural mechanisms of conditioned reinforcement. Once a child's nonverbal behavior has come under the discriminative control of a verbal stimulus, subsequent vocal responses are automatically reinforced by the reinforcing system to the extent that those responses produce verbal stimuli that approximate the discriminated verbal stimuli (Donahoe & Palmer, 1994, 317-319). When the child hears its own voice approximate a vocal discriminative stimulus, conditioned reinforcement immediately occurs for the vocal responses producing that stimulus. Overt vocal behavior is not required for conditioned reinforcement, however. It is sufficient that the environment initiates conditioned activity in motor systems for that activity to engage some of the pathways that project to the reinforcing system. Motor activity that produces overt vocal behavior may yield greater conditioned reinforcement, but covert verbal behavior can also initiate reinforcement (Donahoe, 1997, pp. 353-357). Young children emit overt vocal responses even when they are alone, which exploits both environmentally and intranetwork-mediated conditioned reinforcement. Over time, the role of the child's overt vocal responses in conditioned reinforcement declines because overt behavior may be subjected to aversive contingencies that subvocal speech cannot (Skinner, 1957).

**Syntactic distinctions.** The acquisition of "syntax" has been particularly identified as a challenge to behavior-analytic interpretations of verbal behavior. With the aid of neural networks, this challenge has begun to be met. Simulation research does not yet provide an interpretation of all the verbal relations from which syntax has been inferred. However, this work does demonstrate that some behavioral findings now taken as evidence of syntax can emerge as the products of simpler processes when simulated in neural networks.



**Figure 1.** Basic architecture of a network with recurrent connections used to simulate syntactic constraints. A sequence of words (*Ws*) that formed a sentence stimulated successive input units. The input units activated the hidden units by completely connected feedforward pathways. Each hidden unit then activated a context unit (gray-filled) that re-activated its hidden unit via a recurrent connection. Because activations of the context units required some time to decay, their activations at any moment reflected the cumulative effects of the particular sequence in which the words appeared in the sentence. Not shown are the pathways that completely interconnected the hidden units with the output units. The connection weights were modified during training until stimulation of an input unit by one word activated the output unit corresponding to the next word (*O*) in the sentence. Networks with larger numbers of units were used in the actual simulations.

The competence of neural networks to mediate input-output relations that are consistent with syntactical distinctions is illustrated in the research of Jeffrey Elman (1995). It is widely recognized that brain areas that are especially important for verbal relations are replete with recurrent connections. (Recurrent connections are feedback pathways that permit the activity of a neuron to be affected by its own prior activity. A possible scenario for the evolution of such connections has been given elsewhere, Donahoe, 1991.) Figure 1 depicts recurrent connections within a network that simulates some of the verbal relations from which syntactic distinctions have been inferred. In one simulation, a sequence of words (*W1, W2, ...*) stimulated successive input units of the network. The sequence formed what is conventionally described as a sentence; e.g., *Children like ice-cream*. A number of sentences were composed using different sequences of a limited set of these and other words; e.g., *Boys eat ice cream*. For each sentence, the network learned to activate its output units such that when one word of a sentence stimulated its input unit, the network activated the output unit (*O1, O2, ...*) corresponding to the next word in the sentence. e.g., given *Children*, the network activated the output unit for *like*. For present purposes, the critical units of the network are the hidden units (see Figure 1). (These units are called "hidden" units to distinguish them from input and output units, whose counterparts in living organisms are the only ones whose states are observable at the behavioral level.) These hidden units are activated via two kinds of connections – feedforward connections from input units and recurrent connections from a second set of hidden units. The second set may be called context units. Context units are activated by the hidden units to which they

are connected. Thus, the activity of hidden units at time  $t$  is affected by their own prior activity (via context units) in the preceding time interval,  $t-1$ . In this way, the activations of hidden units reflect their recent *history* of activation, which – in turn – reflects the sequence in which previous words have appeared. The net result is that a given input word activates hidden units in the cumulative context of their activations by the lingering effects of previous words in the sentence.

After training with a number of sentences, networks with recurrent connections were tested. Tests consisted of successively stimulating the input units with a potentially grammatical sequence of words from its trained "vocabulary," but in a sequence that had *not* been trained. The results of the tests indicated that the output unit activated next usually corresponded to a plausible continuation of the new sentence. For example, if *sandwiches* had been a part of the trained sentence *Boys buy sandwiches*, and the network were presented with the untrained sequence *Boys like ...*, the network might most strongly activate the output unit corresponding to *ice cream*. Thus, in some sense the trained network regarded *sandwiches* and *ice cream* as syntactically equivalent. An examination of the patterns of activation of the hidden units receiving recurrent connections indicated how equivalence was accomplished. When words stimulated their input units, those words that are conventionally categorized as "nouns" initiated a similar pattern of activation across the hidden units. In contrast, those words conventionally categorized as "verbs" initiated a different pattern, but one that was similar across all verbs. In short, the activation patterns of the hidden units clustered into conventional syntactic classes as an emergent product of the dynamics of the recurrent network. Of course, nothing in the network corresponded to "nouns" or "verbs" or to "rules" for the use of these types in a sentence. All that existed were the strengths of connections be-

tween units, and these allowed the various patterns of activation on the hidden units to emerge as the inputs are stimulated. In short, the network acted *as if* syntactic classes existed, but – in fact – they did not (cf. Rumelhart & McClelland, 1986, pp. 120). (These simulations yielded other interesting results. For example, transitive and intransitive verbs formed sub-clusters within the verb cluster and transitive verbs fostered the activation of units for nouns as direct objects whereas intransitive verbs did not. Also, nouns for animate and inanimate objects formed sub-clusters with the noun cluster.)

Recurrent networks do not claim to offer a complete interpretation of the verbal behavior from which linguists infer syntactical distinctions. Clearly, the regularities from which syntax is inferred are dependent on much more than the cumulative sequential context in which words appear (e.g., Donahoe & Palmer, 1994; Donahoe & Wessells, 1980). However, these simulations do indicate that some of these distinctions emerge from the dynamics of networks as they respond to sequences of inputs. Moreover, experimental work supports the claim that humans are sensitive to such sequential regularities. Eight-month-old infants who were exposed to sequences of phonemes for as little as two minutes subsequently responded differently to those sequences than to other sequences made up of the same phonemes appearing in a different order (Saffran, Aslin, & Newport, 1996; see also Bates & Elman, 1996).

We are at the very beginning of efforts to provide a detailed natural-science interpretation of verbal behavior. Nevertheless, previously problematic phenomena such as language acquisition and syntactic distinctions appear within reach. Fundamental biobehavioral principles arising from behavior analysis and neuroscience provide a powerful means of interpretation when they are implemented in neural networks.

## References

- Barnes, D. & Hampson, P. J. (1997). Connectionist models of arbitrarily applicable relational responding: A possible role for the hippocampal system. In Donahoe, J. W. & Dorsel, V. P. (Eds.) *Neural-network models of cognition: Biobehavioral foundations* (pp. 496-521). Amsterdam, Netherlands: Elsevier Science Press.
- Bates, E. & Elman, J. (1996). Learning rediscovered. *Science*, **274**, 1849-1850.
- Bechtel, W. & Abrahamsen, A. (1991). *Connectionism and the mind*. Oxford: Basil Blackwell. (26)
- Brown, R. & Hanlon, C. (1970). Derivational complexity and order of acquisition in child speech. In J. R. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.
- Devaney, J. M., Hayes, S. C., & Nelson, R. O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior*, **46**, 243-257.
- Donahoe, J. W. (1991). The selectionist approach to verbal behavior: Potential contributions of neuropsychology and connectionism, In L. J. Hayes & P. N. Chase (Eds.), *Dialogues on verbal behavior* (pp. 119-145.) Reno, NV: Context Press.

- Donahoe, J. W. (1993). The unconventional wisdom of B. F. Skinner: The analysis-interpretation distinction. *Journal of the Experimental Analysis of Behavior*, **60**, 453-456.
- Donahoe, J. W. (1997). Selection networks: Simulation of plasticity through reinforcement learning. In Donahoe, J. W. & Dorsel, V. P. (Eds.) *Neural-network models of cognition: Biobehavioral foundations* (pp. 336-357). Amsterdam, Netherlands: Elsevier Science Press.
- Donahoe, J. W., Burgos, J. E., & Palmer, D. C. (1993). Selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior*, **58**, 17-40.
- Donahoe, J. W. & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to Parallel Distributed Processing. *Journal of the Experimental Analysis of Behavior*, **51**, 399-416.
- Donahoe, J. W. & Palmer, D. C. (1994). *Learning and complex behavior*. Boston: Allyn & Bacon.
- Donahoe, J. W. & Wessells, M. G. (1980). *Learning, language, and memory*. New York: Harper & Row.
- Dugdale, N. A. & Lowe, C. F. (1990) Naming and stimulus equivalence. In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 115-138). Hove, England: Erlbaum.
- Elman, J. L. (1995). Language as a dynamical system. In R. Port & T. Gelder (Eds.), *Mind as motion* (pp. 195-224). Cambridge, MA: MIT Press.
- Galbicka, G. (1992). The dynamics of behavior. *Journal of the Experimental Analysis of Behavior*, **57**, 243-248.
- Hayes, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, **51**, 385-392.
- Hayes, S.C., & Hayes, L. J. (1992). Verbal relations and the evolution of behavior analysis. *American Psychologist*, **47**, 1383-1395. (26)
- Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and symbolic behavior. *Journal of the Experimental Analysis of Behavior*, **65**, 185-241.
- Keller, F. S. & Schoenfeld, W. N. (1950). *Principles of psychology*. New York: Appleton-Century-Crofts.
- McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.). (1986). *Parallel distributed processing* (two vols.). Cambridge, MA: MIT Press.
- Moerk, E. L. (1990). Three-term contingency patterns in mother-child verbal interactions during first language acquisition. *Journal of the Experimental Analysis of Behavior*, **54**, 293-305.
- Palmer, D. C. & Donahoe, J. W. (1992). Essentialism and selectionism in cognitive science and behavior analysis. *American Psychologist*, **47**, 1344-1358.
- Raichle, M. E. (1997). Functional brain imaging and verbal behavior. In Donahoe, J. W. & Dorsel, V. P. (Eds.) *Neural-network models of cognition: Biobehavioral foundations* (pp. 438-454). Amsterdam, Netherlands: Elsevier Science Press.
- Rumelhart, D. E. & McClelland, J. L. (1986). PDP models and general issues in cognitive science. In D. E. Rumelhart, J. L. McClelland, J. L., & The PDP Group (Eds.). *Parallel distributed processing* Vol.1 (pp. 110-146). Cambridge, MA: MIT Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926-1928.
- Sakai, K. & Miyashita, Y. (1991). Neural organization for the long-term memory of primates. *Nature*, **354**, 152-155.
- Sidman, M. & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, **37**, 5-22.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1974). *About behaviorism*. New York: Random House.