JOURNAL OF THE EXPERIMENTAL ANALYSIS OF BEHAVIOR 1989, 52, 427-440 NUMBER 3 (NOVEMBER)

CONNECTIONIST MODELS OF CONDITIONING: A TUTORIAL

E. JAMES KEHOE

UNIVERSITY OF NEW SOUTH WALES

Models containing networks of neuron-like units have become increasingly prominent in the study of both cognitive psychology and artificial intelligence. This article describes the basic features of connectionist models and provides an illustrative application to compound-stimulus effects in respondent conditioning. Connectionist models designed specifically for operant conditioning are not yet widely available, but some current learning algorithms for machine learning indicate that such models are feasible. Conversely, designers for machine learning appear to have recognized the value of behavioral principles in producing adaptive behavior in their creations.

Key words: connectionist models, neural networks, operant conditioning, respondent conditioning, stimulus control, mixed schedules, concurrent schedules

Models containing networks of neuron-like units have become prominent in the study of both cognitive psychology and artificial intelligence (e.g., Anderson & Rosenfeld, 1988; Feldman, 1985; Rumelhart & McClelland, 1986). These networks contain a large number of units and purportedly compute the complex functions underlying phenomena such as word recognition, visual pattern perception, and coordinated motor action. In the study of human cognition, connectionist models represent a dramatic departure from conventional theories that assume grammar-like manipulations of symbolic information (Estes, 1988; Gluck & Bower, 1988). Conversely, the nongrammatic and nonsymbolic features of connectionist models appear to make them suitable for the study of nonhuman conditioning and cognition (Barto, Sutton, & Anderson, 1983; Kehoe, 1986a, 1988; Klopf, 1982, 1988).

This article will provide a tutorial overview of connectionist models in three parts. The first describes the general principles of connectionist models. The second part illustrates the application of connectionist principles to the modeling of compound-stimulus effects in respondent conditioning, which, however, are shared with operant conditioning. The third part will sketch connectionist algorithms for machine learning that mimic key features of operant conditioning. These algorithms serve a twin purpose here. On the one hand, they demonstrate the feasibility of connectionist models for operant conditioning. On the other hand, they illustrate how behavioral principles are being recognized and incorporated into the development of artificial intelligence.

Preparation of the manuscript was supported by Grant A28315236 from the Australian Research Council. The author extends his gratitude to Peita Graham-Clarke for her assistance in collecting data and to Renee Napier in preparing this manuscript. Correspondence concerning this article should be sent to E. James Kehoe, School of Psychology, University of New South Wales, Kensington, New South Wales 2033, Australia.

BASIC FEATURES OF CONNECTIONIST MODELS

Although connectionist models are constructed of neuron-like units and have been applied to the neural circuitry of respondent preparations (e.g., Gelperin, Hopfield, & Tank, 1985; Gluck & Thompson, 1987; Hawkins & Kandel, 1984; Zipser, 1986), these models have been constrained only weakly by the known architecture and functioning of real nervous systems. There is nothing to prevent them from being applied at a strictly behavioral level. Stripped of their surplus meaning, connectionist models can be viewed as a class of quantitative models, albeit very elaborate models, subject to the conventional criteria for testing any model. The extensive use of computer simulations requires that these models be fully specified in terms of their own inner workings and their generation of behavioral outputs. Hence, it is possible to construct and test a connectionist model that makes absolute, as well as ordinal, predictions about behavior.

As a class, connectionist models postulate a set of interacting subsystems that share two basic features:

1. The basic subsystem is a relatively lean

computational unit that can be described by two equations, namely an *activation rule* and a *learning rule.* The activation rule combines inputs to the unit and generates an output. The learning rule alters the strengths of active inputs by altering variables commonly designated as *connection weights.* As will be shown, many variations on operant and respondent principles are under active consideration as learning rules.

2. Interactions between the units entail the transmission of *activation levels* from the output of one unit to the input of another. Frequently, the output of a unit is postulated to be an all-or-none firing. However, any real number may be used. The input level to the receiving unit is usually the product of the current activation level and the current connection weight at the receiving unit.

Along with the basic features concerning the units and their interactions, connectionist models typically do not contain any executive subsystem. That is to say, many connectionist models are aimed at yielding "purposeful" or "cognitive" behavior on the basis of mechanistic interactions among the units. Behavior that might appear to follow a "rule," "hypothesis," or "strategy" is supposed to emerge from the interactions among the units, none of which contain a representation of the global rule, hypothesis, or strategy. Instead, each unit can be said to "know" only its current connection weights and current activation levels.

The fundamental ideas for connectionist modeling arose from speculation concerning neural functioning in the 1940s. Specifically, there were two key developments, namely the *linear threshold unit* and *synaptic facilitation.*

Linear Threshold Unit

The source of modern activation rules lies in the work of McCulloch and Pitts (1943). They contended that a neuron could act as a logic gate that fires in an all-or-none fashion if the sum of its inputs exceeds a certain threshold. Figure 1 depicts a generalized linear threshold unit that has four key features. First, on the left side of the unit are the input variables that are characterized as input activation levels (xi) and weighted connections (Vi). Either variable can assume any real value. Commonly, however, the activation levels are assumed to be binary ($xi = 0, 1$) and the weights are limited to fall between -1 and $+1$. Second, the total input level in any time unit (t) is the sum of the active input weights ($\sum [Vi\ xi]$). Third, the output of the unit is a binary activation level ($yj = 0, 1$). Fourth, the activation rule involves a comparison of the total input level with a threshold value (Tj). One common representation of the activation rule is:

$$yi = 1 \text{ if } (\sum [Vi\ xi] - Tj) > 0, \quad \text{otherwise } yi = 0. \tag{1}$$

By manipulating the connection weights or the threshold values, it is possible to produce common logic functions. For example, an AND gate can be constructed in the following fashion. Assume that a unit has two inputs ($x1$, $x2$), each with a connection weight of .50 ($V1 = V2 = .50$), and that the unit's threshold is .75 ($Tj = .75$). Under the McCulloch–Pitts activation rule, both inputs would have to be active ($x1 = x2 = 1$) for the total input level to exceed the threshold and thereby trigger the unit. The same unit can be converted to an OR gate either by lowering the threshold to a value less than .50 or by raising the input weights to values greater than .75. Finally, for a complete logic system, the NOT operator can be constructed by inverting the activation rule so that a total input level that exceeds threshold turns off the unit.

Synaptic Facilitation

Whereas the McClulloch–Pitts unit is fixed in its functioning, Hebb (1949, p. 50) proposed a simple rule for altering the weight of a neural connection. Effectively, Hebb applied the ancient law of contiguity at a neural level. According to Hebb, the weight of a connection gains in value if presynaptic activity is contiguous with postsynaptic activity. For example, take a neuron with two input connections. One has a large weight capable of triggering the neuron, and the other input has little or no weight. If there are simultaneous inputs, then the heavily weighted input will cause postsynaptic activity, and the weaker input will provide the presynaptic activity. In mathematical terms, the change in connection weight (dVi) at time t is represented as a product of the presynaptic activity (xi) and postsynaptic activity (yj) (Sutton & Barto, 1981):

$$dVi = c\ xi\ yj \tag{2}$$

where c is a rate parameter ($0 < c \leq 1$).

COMPOUND-STIMULUS CONTROL IN CONDITIONING

An issue common to research in conditioning, cognition, and artificial intelligence concerns arbitrary mappings from stimulus input patterns to response output patterns (e.g., C. W. Anderson, 1986; J. R. Anderson, 1985, pp 73–134; Kehoe & Gormezano, 1980). Broadly known as the *representation problem,* it arises whenever the output mapping is not a linear combination of the separate inputs. It is possible to concoct a nonlinear discrimination with as few as two inputs, specifically, the exclusive-OR problem (XOR), in which the learner must respond to each of two inputs separately but *not* to their joint occurrence (Barto, 1985; Rumelhart, Hinton, & Williams, 1986, p. 319). It is impossible to generate the appropriate reaction, namely no response, to the joint stimulus inputs by a summation of the responses attached to the two separate inputs (Minksy & Papert, 1969; see also C. W. Anderson, 1986, p. 27; Rumelhart et al., 1986, p. 319).

Within both respondent and operant conditioning, there are a number of schedules for studying representation problems. An empirical example of the XOR problem is the negative patterning schedule, in which the subject is trained with a mixture of three types of trials: (a) one stimulus that is reinforced (A+), (b) a second stimulus that also is reinforced (B+), and (c) a compound of A and B that is never reinforced (AB−) (Pavlov, 1927, p. 144; see Bellingham, Gillette-Bellingham, & Kehoe, 1985; Kehoe & Graham, 1988; Rescorla, 1972, 1973; Woodbury, 1943). More elaborate nonlinear mappings can be studied in biconditional discriminations, in which four stimuli (A, B, C, and D) and the reinforcer are combined into four symmetric compounds, for example, AC+, AD−, BC−, BD+. If each of the compounds is presented equally often, then the subjects must use each compound as a unit to make the appropriate discriminative response, because each of the individual stimuli is presented equally often with and without the reinforcer (e.g., Heinemann & Chase, 1975; Saavedra, 1975).

It is possible to convert a nonlinear problem into a linear problem by postulating a special input that is triggered only by the joint occurrence of the two basic stimulus inputs. However, for systems with a large number of basic inputs, there would an explosive proliferation of prewired special inputs to cover each possible combination (Kehoe & Gormezano, 1980, p. 375). In the literature of connectionist modeling, considerable effort has been devoted to solving nonlinear discriminations by synthesizing representations of compound inputs as the need arises (e.g., C. W. Anderson, 1986; J. A. Anderson, 1973, 1977; Rumelhart & McClelland, 1986). This effort has been centered on networks containing two or more layers of adaptive units. In a prototypical network, units in the first layer, the so-called *hidden units,* receive basic sensory inputs. In turn these hidden units send projections to a second layer containing *output units.* Generally, the hidden units acquire a set of weights such that each one is activated only by a specific combination of inputs. In turn, the connection weights between the hidden units and the output units designate the response pattern to be controlled by the combined inputs.

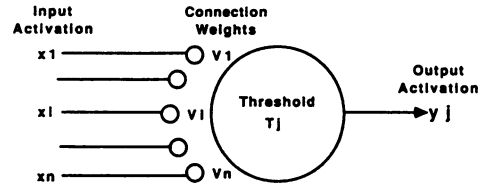

Fig. 1. A prototypical linear threshold unit showing the key variables, namely input activation levels (*xi*), connection weights (*Vi*), threshold (*Tj*), and the output activation level (*yj*).

Along with the development of the layered architecture, there has been considerable effort devoted to the development of learning rules that "tune" units to specific combinations of inputs (C. W. Anderson, 1986; Barto, 1985; Rumelhart et al., 1986). The Hebbian rule has not appeared suitable for tuning units to combinations of events, because it allows virtually unlimited increments in connection weights. Moreover, because the rule uses the product of the input, presynaptic activity (*xi*) and the output, postsynaptic activity (*yj*), any increases in the number of inputs would tend to amplify the increments in the weights of the inputs. Consequently, theoretical attention has turned to learning rules that reduce the size of increments to each weight as the number of inputs increases. As it so happens, the learning rule used by Rescorla and Wagner (1972) is

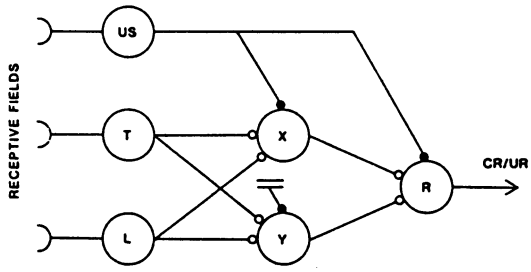

Fig. 2. A simple but prototypical layered network for configural learning. It contains three sensory inputs (T, L, US) and three adaptive units (X, Y, R) (Kehoe, 1988).

a good example of a wider class of such rules, known variously as the Widrow–Hoff rule (Sutton & Barto, 1981), the delta rule (Rumelhart et al., 1986), and the least mean squares rule (Gluck & Bower, 1988).

Whereas the Hebbian rule uses a product of input activity and output activity, the newer rules use the difference between the total input activity and the output activity. Hence, as the total input activity approaches the maximal level of output activity, increments in the input weights will diminish in size. Using the difference in levels effectively means that the inputs will divide the available weight among themselves, limiting each of them to a relatively low value. Where there is a threshold for triggering an adaptive unit, a division of the total weight among competing inputs would ensure that no input by itself would gain sufficient connective weight to be able to trigger the unit. In this way, the Rescorla–Wagner learning rule can cause a unit to become tuned to combined inputs.

A Layered Network Model

Figure 2 depicts the architecture of a small but representative network suitable for compound-stimulus effects in respondent conditioning (Kehoe, 1986a, 1988). The network contains three "sensory" inputs, one each for a tone CS (T), a light CS (L), and an unconditioned stimulus (US). The outputs from T and L each project to two hidden units (X, Y). In turn, the X and Y units project to a response generator unit (R). The US projects a fixed input to the X, Y, and R units.

The learning rule for each of the three adaptive units (X, Y, R) is essentially the same as that used by Rescorla and Wagner (1972), but is now applied to individual adaptive units rather than to the whole organism:

$$dVij = aj(Lj - V'ij) \qquad (3)$$

where $dVij$ is the change in connection value of the ith input to the jth unit, aj is the rate parameter for the target unit of the connection ($0 < aj < 1$) (on non reinforced trials, aj is reduced by the parameter B0 ($0 < B0 < 1$)), Lj is the total connection value that can be supported by the US on any given trial (Lj = 1.0 on reinforced trials; Lj = 0.0 on nonreinforced trials), and $V'ij$ is the net sum of the connection values of all currently eligible inputs to the jth unit.

The activation rule follows closely that of the logical threshold unit. That is to say, an all-or-none output depends on whether or not the net sum of the active input weights exceeds the value of a noisy threshold. At the beginning of training, all the units are activated only by the US input, but otherwise the units are not biased toward any particular sensory input or any combination of sensory inputs.

Testing a Network Model

By comparison to other current network models, the present network is miniscule. However, as a quantitative model for behavior, it has a fearsome number of free parameters. Each of the three adaptive units has (a) its learning rate parameter (aj), (b) its nonreinforcement parameter (B0), and (c) its mean threshold value (Tj). In order to test the model rigorously and forestall a mere curve-fitting exercise, I have undertaken a two-pronged strategy that is applicable to any connectionist model. First, on a between-group basis, I sought to discover whether a single set of parameter values could generate the acquisition curves under a variety of schedules (Kehoe, 1988). Second, on a within-subject basis, I have conducted transfer experiments to discover whether parameter values fitted to responding under one training schedule can accurately predict performance under a different schedule.

Between-group tests. As the empirical basis for the initial between-group test of the model's scope, Figure 3 shows three instances of *configural learning,* in which the subject shows differential responding to a compound versus its two components (Kehoe & Gormezano, 1980). The top panel shows the behavioral data, and the bottom panel shows simulations

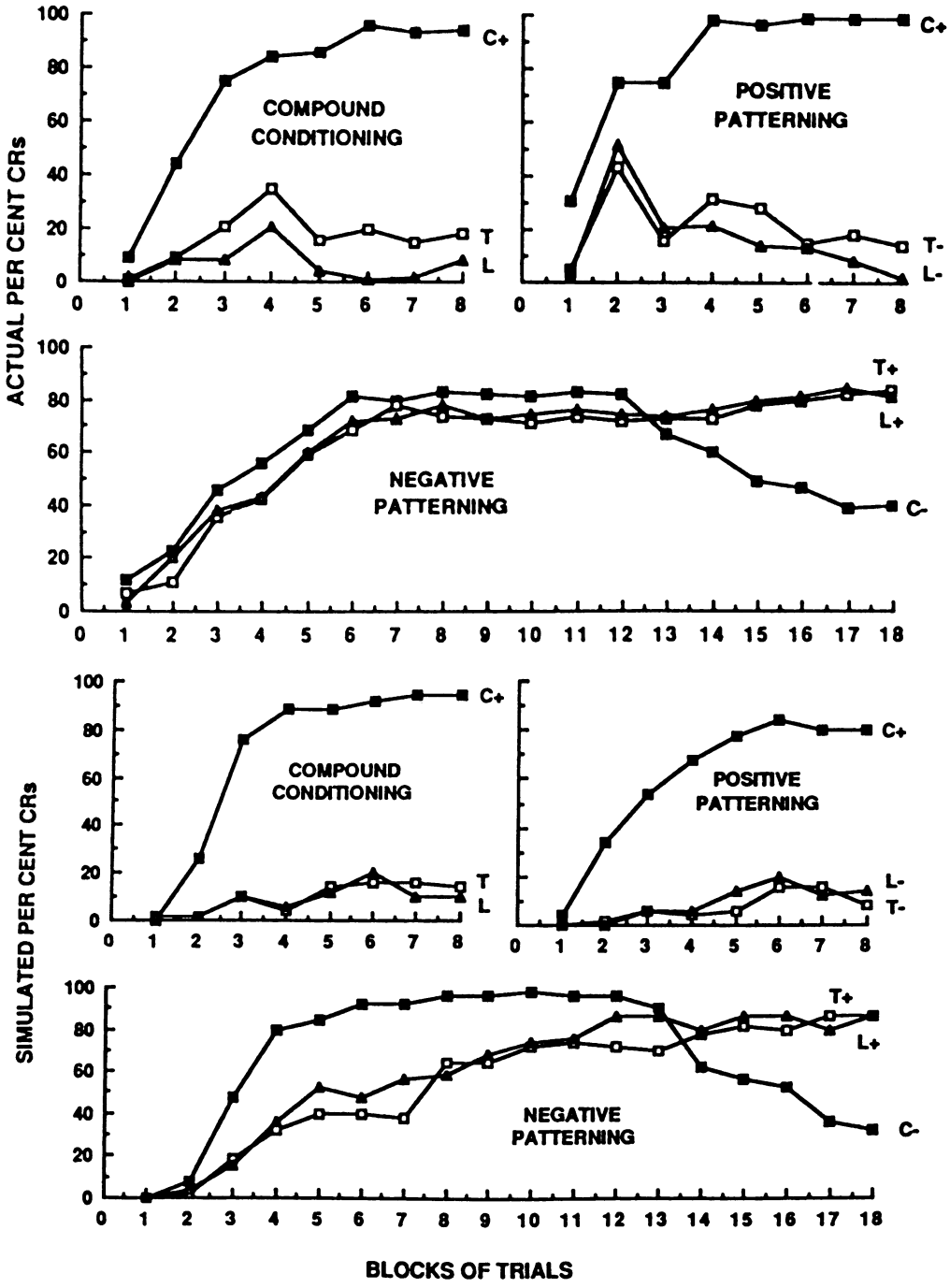

Fig. 3. Behavioral data and simulations for three examples of configural learning, namely compound conditioning, positive patterning, and negative patterning (Kehoe, 1988).

generated by Kehoe's (1986a, 1988) model. In each panel, the lower set of curves represents the course of acquisition in the nonlinear, negative patterning schedule (Bellingham et al., 1985). The upper curves represent the differentiation seen in *compound conditioning,* which entails only reinforced presentations of the compound (TL+) (Kehoe, 1986b; Kehoe & Schreurs, 1986a; Wickens, Nield, Tuber, & Wickens, 1970), and *positive patterning,* which involves reinforced presentations of the compound (TL+) intermixed with an equal number of unreinforced presentations of the separate components (T−, L−) (Bellingham et al., 1985). In compound conditioning, occasional tests with the individual components were needed to expose their separate response-evoking capacities, but control experiments have demonstrated that a few test trials do not alter the results discernibly (Kehoe & Schreurs, 1986b). Although group curves are shown, they reflect the individual subjects. In compound conditioning and negative patterning, approximately 75% of the animals in any one experiment show differentiation of the compound from the components. In positive patterning, virtually 100% of the animals show appropriate differentiation.

According to the model, all three instances of configural learning demand that the network synthesizes a representation of the compound by allowing the X or Y unit to be triggered only by the joint occurrence of the two CS inputs. According to the Rescorla–Wagner learning rule, if two inputs to either the X, Y, and/or R units occur simultaneously, then those inputs compete for the available connection weight supported by the US input. Thus, in compound training, simultaneous T and L inputs to, say, the X unit will tend to divide the available connection weight between them. If only the summated T-X and L-X connections are strong enough to trigger the X unit, then the X unit effectively constitutes a representation of the compound (Kehoe, 1986a, 1988).

To discover whether a single set of parameters would reproduce all three configural learning phenomena, two groups of parameters were manipulated, namely the mean threshold value of each unit (Tj) and the learning rate parameter for each unit (aj). The value of B0 was held constant at .33 for all three units. An extensive search of the parameter space was conducted. That is to say, for each combination of parameter values, a simulation was conducted for compound conditioning, positive patterning, and negative patterning. The results of the simulations were then compared to the behavioral data using the total sum of squares.

The bottom panel of Figure 3 depicts simulated acquisition curves. As can be seen, the simulations were able to capture (a) the relatively slow acquisition of negative patterning characterized by an initial rise in responding to the unreinforced compound stimulus followed by a gradual decline, and (b) the more rapid, massive differentiation between the compound and its components in both compound conditioning and positive patterning. These simulations were obtained when (a) the X unit had a higher mean threshold than the Y unit (Tx = .70, Ty = .15) and (b) the X unit had a higher learning rate than that of the Y unit (ax = .050, ay = .001). The high threshold for the X unit made it very sensitive to the effects of any competition between the T and L inputs. When T and L divided the weights evenly (VTX = VLX = .50), then X was triggered only by the compound. Conversely, the low threshold of the Y unit made it insensitive to the effects of competition between the T and L inputs. The Y unit was triggered reliably by either the T or L input at all but the lowest connection weights. The difference in the learning rates between the X and Y units gave the X unit an advantage in competing for access to the R unit. With its high learning rate, the X unit began to trigger earlier in training and thus ensured that the X-R connection started to strengthen before the Y-R connection became eligible for alteration. The R unit had an intermediate learning rate (ar = .010) and an intermediate mean threshold (Tr = .50).

To test the predictive power of the model, simulations using the same parameter values were compared to the acquisition curves obtained in other paradigms involving compound stimuli (Kehoe, 1988). In particular, the model was able to generate curves that duplicated those observed for the *conditioned inhibition* schedule (TL−, L+), the *feature positive* schedule (TL+, L−), and *stimulus compounding* (T+, L+, test TL). The model was also able to reproduce blocking and overshadowing.

Within-subject tests. A within-subject test for

the network model has been conducted by examining transfer from the conditioned inhibition schedule to the negative patterning schedule. Procedurally, conditioned inhibition (TL−, L+) and negative patterning (TL−, L+, T+) differ in only one respect, that is whether one or both of the components are presented and reinforced. However, that small operational difference yields distinctive patterns of connection weights in my network model.

Figure 4 shows the terminal connection weights for conditioned inhibition and negative patterning obtained from the simulations described above. In the case of conditioned inhibition, the X and Y units function in parallel. The T input develops inhibitory connections with both the X and Y units, whereas the L input develops excitatory connections. Hence, on TL− trials, the net sum of the inputs to both hidden units falls below the value of their thresholds, thus failing to activate the excitatory X-R and Y-R connections. On L+ trials, the excitatory L-X and L-Y connections operate in an unimpeded fashion to trigger both hidden units and ultimately the R unit.

In the case of negative patterning, the X unit becomes tuned solely to the compound and in turn acquires an inhibitory connection with the R unit. That is to say, the simulated T+ and L+ trials yield excitatory connections in the first layer, namely T-X, L-X, T-Y, and L-Y. Because of X's high threshold, only the summated T-X and L-X connections can trigger X. As a result, the X input to the R unit occurs only on TL− trials, and the X-R connection weight is driven into the negative range (VXR = −.60) in contrast to the excitatory Y-R connection (VYR = .83). Consequently, on compound trials, in which both X and Y units are triggered by the combined T and L inputs, the opposing weights of the X-R and Y-R connections largely cancel each other and preclude CRs to the compound. On component trials, the separate T and L inputs are insufficient to trigger X but are high enough to trigger Y. Consequently, only the strongly excitatory Y-R connection activates the R unit with the consequent generation of a response.

The distinctive patterns of connection weights produced by conditioned inhibition and negative patterning provided an opportunity to test the model on a within-subject basis. By conducting conditioned inhibition training and

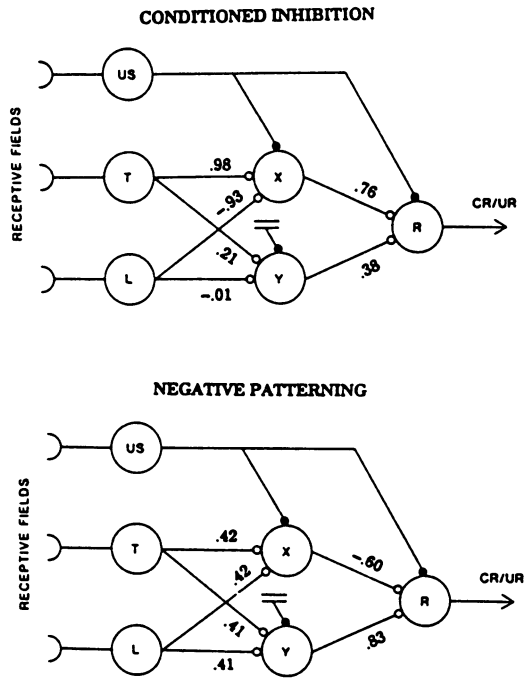

Fig. 4. The terminal connection weights for simulations of conditioned inhibition and negative patterning.

then estimating the appropriate parameters, it was possible to test whether the model could predict the course of transfer to a subsequent negative patterning schedule. Figure 5 shows the results of the behavioral experiment in the left panel and the corresponding simulations in the right panel.[1] The results for 3 of 8 animals are shown. They represent the animal that showed the largest differentiation between B+ and AB− at the termination of the conditioned inhibition schedule (Fastest S), the animal that showed the median level of differentiation (Median S), and the animal showing the least differentiation (Slowest S).

For each animal, curves were fitted to its performance during the conditioned inhibition schedule by conducting an extensive search across combinations of the learning rate parameters for the X, Y, and R units (ax, ay, and ar). For the fastest subject, the parameter values for ax, ay, and ar were .070, .005, and .003. For the median subject, the learning rates

[1] Data from Graham-Clarke, P., & Kehoe, E. J. (1988). *Atomistic and configural process: Transfer from feature schedules to patterning schedules in the rabbit nictitating membrane preparation.* Unpublished manuscript.

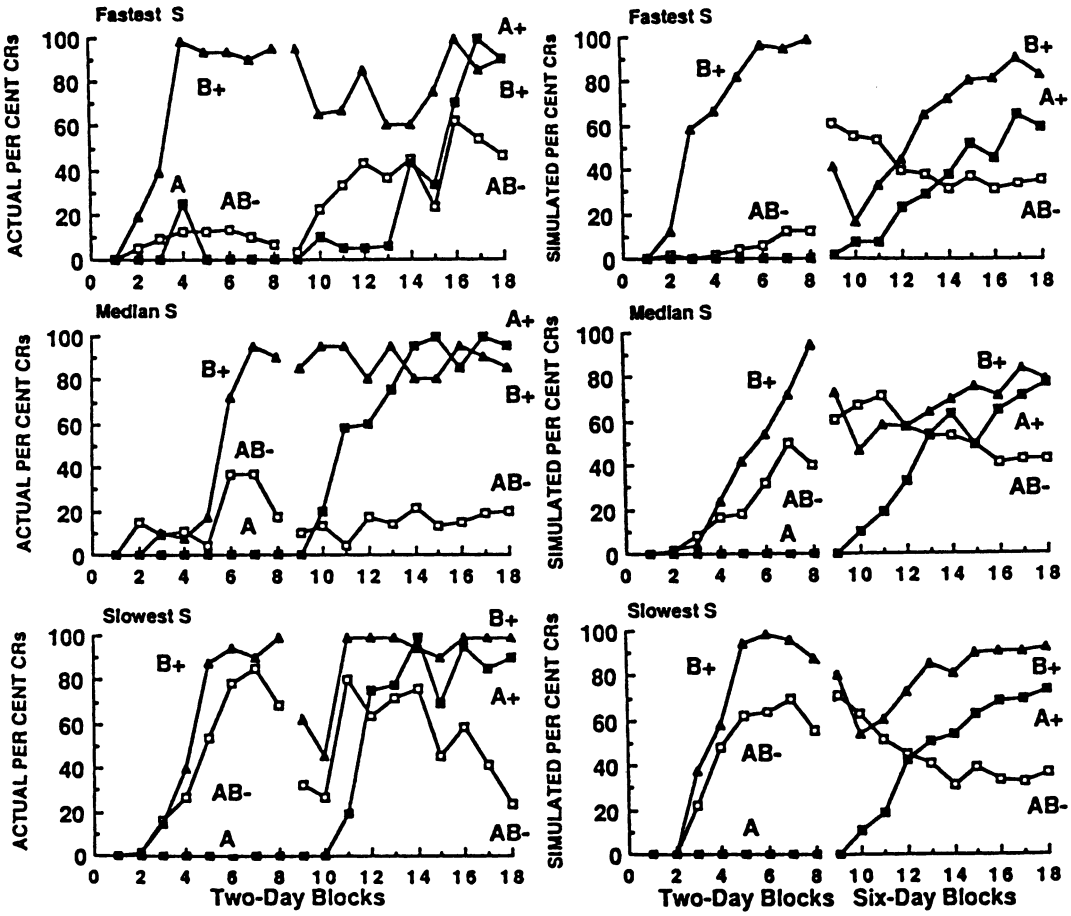

Fig. 5. Behavioral data and simulations for transfer from conditioned inhibition to negative patterning in 3 subjects.

were .010, .005, and .002, and, for the slowest subject, the learning rates were .010, .002, and .006. Other parameters were fixed uniformly for all 3 animals (Tx = .70, Ty = .15, Tr = .50; $B0x$ = .20, $B0y$ = .20, $B0xr$ = .50, $B0yr$ = .20). The estimated parameter values for each animal were then used in a simulation of transfer to the negative patterning schedule.

A comparison of the actual and simulated responding reveals a moderate correspondence in negative patterning. Both the rabbits and the model showed successful transfer from conditioned inhibition to negative patterning. However, as the reader may have noted, the abscissa for the simulated negative patterning is compressed from 2-day to 6-day blocks. That is to say, the attainment of negative patterning in the simulations required approximately three times as long as the animals did. In the simulations, a similar pattern of transfer appeared for all three parameter sets, namely (a) a transitory dip in responding to B+ immediately following the introduction of the negative patterning schedule, (b) an abrupt rise in responding to AB− followed by a gradual decline, and (c) a steady rise in responding to A+. In comparing the simulated curves to those of the actual animals, it becomes clear that there is considerable room for improvement in capturing the large individual differences among the animals. The best agreement between a simulation and actual behavior appeared in the Worst S, which showed the initial dip in responding to B+, the gradual decline in responding to AB−, and the steady rise in responding to A+. The next best agreement appeared in the Best S, which showed the dip in responding to B+ and the rise in

responding to A+, but showed a slow rise rather than a decline in responding to AB−. In the Median S, there was very little agreement between the actual and simulated curves.

The foregoing exercise clearly indicates that individualized within-subject prediction provides a rigorous test for the proposed network model, exposing potential shortcomings that did not appear in between-group assessments (Kehoe, 1986a, 1988). These shortcomings can be used to advantage in guiding further development and testing of the model. However, rather than relying on individual differences to generate a spread of parameter estimates, a delineation of the parameters can be more readily accomplished by a systematic manipulation of the schedules' independent variables, for example, the proportion of reinforced to unreinforced trials.

LEARNING ALGORITHMS AND OPERANT REINFORCEMENT

Connectionist models designed specifically for operant conditioning will have begun to appear by the time this article is published.[2] However, the application of connectionist modeling techniques to operant conditioning has lagged behind respondent conditioning. In part, this lag is historical, because Hebb's learning rule is essentially a stimulus-response contiguity principle applied at the neural level. Consequently, the modeling of respondent conditioning received a head start on operant conditioning. However, interest in learning algorithms that mimic operant conditioning has arisen in connection with the design of machine systems that learn to control the movement of a vehicle in its environment. Such machine systems are intended to control things such as wobble in communication satellites or search patterns among robot exploration vehicles.

In designing adaptive control systems, Barto, Sutton, and their colleagues have explored a variety of machine learning tasks that share key features with many operant schedules (Barto, 1985; Barto & Sutton, 1981; Barto et al., 1983; Sutton, 1984). As in any operant schedule, the actions of the control system are intended to act on the machine's surrounding environment and be sensitive to the consequences of its actions. Furthermore, Barto et al. (1983) argue that a successful control system must be able to learn when (a) the system has no foreknowledge of the environment's "dynamics," that is to say, the environment's stimulus–reinforcer and response–reinforcer contingencies, and (b) the system receives feedback on a delayed and/or infrequent basis.

In addition to the operant-like contingencies in machine learning tasks, Sutton (1984) has contended that a successful control system must be able to learn efficiently when the feedback itself supplies no explicit instructions as to the most appropriate response. Specifically, Sutton distinguishes between learning with a "critic" and learning with "instruction." In learning with a critic, the feedback indicates only whether or not the preceding response was appropriate to achieving a goal. For example, in a categorization task, feedback that entails only "right" or "wrong" would be considered learning with a critic. In contrast, the feedback in learning with instruction designates the exact nature of the desired response. Accordingly, in a categorization task, instructional feedback would entail the name of the appropriate category. If the subject makes a correct response, such feedback confirms the response. Alternatively, if the subject makes an incorrect response, the identical feedback provides an explicit corrective.

Sutton's (1984) distinction between critical versus instructional feedback identifies a further parallel between operant conditioning and the constraints on adaptive control systems. Specifically, the reinforcing events commonly used in operant contingencies appear to approximate critics rather than instructions, as defined by Sutton (1984). Conversely, the USs used in traditional respondent conditioning paradigms act more like instructions than critics. Specifically, in respondent paradigms such as Pavlov's salivary preparation and the rabbit

---

[2] Commons, M. L. (1989, June). *Models of acquisition and preference.* Paper presented at 12th Symposium on Models of Behavior: Neural Network Models of Conditioning and Action, Cambridge, MA.

Reid, A. K. (1989, June). *Computational models of instrumental and scheduled performance.* Paper presented at 12th Symposium on Models of Behavior: Neural Network Models of Conditioning and Action, Cambridge, MA.

Staddon, J. E. R. (1989, June). *Simple parallel model for operant learning with application to a class of inference problems.* Paper presented at 12th Symposium on Models of Behavior: Neural Network Models of Conditioning and Action, Cambridge, MA.

nictitating membrane preparation, the elicitation of the UR by the US effectively acts to either confirm an existing CR or to instruct the subject as to the nature of the CR to be acquired. Of course, a CR is rarely identical to the UR and, moreover, the pairing of a CS with a US can yield associations beyond those expressed in the CS–CR relation (e.g., Rescorla & Solomon, 1967; Woodruff & Williams, 1976). Nevertheless, a difference in the kind of reinforcer as well as contingency appears to be an operant–respondent distinction with important implications for the design of adaptive control systems.

According to Barto et al. (1983), learning with a critic requires mechanisms additional to those used in learning with explicit instruction. In learning with instruction, the learner is faced mainly with a problem of stimulus control, that is to say, the mapping of stimulus inputs onto the designated responses. In learning with a critic, as in operant conditioning, the learner must also discover the appropriate responses in a situation. Thus, there must be a mechanism that effectively records the learner's history of response–reinforcer relations.

Having specified the conditions imposed on an adaptive control system, Barto, Sutton, and their colleagues have examined a variety of suitable algorithms. Two of these algorithms will be described below, one relevant to free-operant contingencies and one relevant to discriminant operant contingencies. These algorithms are not meant to be theories of operant reinforcement, but rather are described here to illustrate the mathematical tools that are becoming available for modeling of operant behavior.

Mimics for Free-Operant Conditioning

Sutton (1984) has studied a family of algorithms for an artificial "nonassociative task" similar to a concurrent schedule. Sutton's aim was to discover the rules that optimize the rate of reinforcement when the learning system had two mutually exclusive and exhaustive response classes coded as (1, 0). Because these two classes were exhaustive, R1 can be construed as a discrete class (e.g., a bar press), and R0 can be construed as all other behavior. In Sutton's task, each response was followed by a positive reinforcer (+1) on a probabilistic schedule. The schedules included one in which the reinforcement probability for both responses was high (.9, .8), one in which the probabilities were intermediate (.55, .45), and one in which the probabilities were low (.2, .1) (cf. Alsop & Elliffe, 1988). In addition to the positive reinforcer contingency, there was also a punishment contingency. A response occurrence that was not followed by a positive reinforcer was followed instead by a complementary punishing event (−1). Thus, in each of the three schedules, the probabilities of punishment were inversely related to the probabilities of positive reinforcement. That is to say, there was a mixed schedule of positive reinforcement and punishment for each response class.

Using this combined schedule, Sutton (1984) explored a variety of algorithms that can be seen as versions of a symmetric reinforcement principle. In general, Sutton's algorithms all followed a similar formula, namely the change in a response strength variable (W) is a product of the response level at time t and the subsequent reinforcer's value at time $t + 1$. A simple expression of this notion would be:

$$dW = cRS^r \tag{4}$$

where dW is the change in the response strength $R1$ relative to $R0$, c is a growth rate parameter ($0 < c$), R is the response (0, 1) at time t, and S^r is the value of the reinforcer (1, −1) at time $t + 1$. As will be detailed below, the variations in the rules revolve around the exact variables that enter into each of the main terms of the equation. With regard to an activation rule, Sutton used a threshold rule, in which the current value of W was compared to a noisy threshold. $R1$ occurred if W exceeded the momentary value of the threshold, and $R0$ occurred if W fell below the momentary value. Thus, high values of W favored $R1$, and low values favored $R0$. There were two major variations in learning rules that can be termed *the absolute reinforcer rule* and *the predicted reinforcer rule.*

Sutton's absolute reinforcer rule can be expressed as follows:

$$dW = c(R - P^R)S^r, \tag{5}$$

where P^R is the probability of $R1$ relative to $R0$ at time t. Under this rule, the absolute value of the reinforcer ($S^r = +1, -1$) still applies as it does in the basic formula (Equation 4).

However, the magnitude of the change in W is modulated by the discrepancy between the current response value (1, 0) and the relative probability of $R1$ versus $R0$. For example, if $R = 1$ and the value of P^R is near 1.00, the discrepancy is relatively small. Accordingly, neither positive reinforcement nor punishment of $R1$ would have much effect on W. Conversely, if $R = 0$ and the value of P^R is near 1.00, then the discrepancy is both large and negative. In that case, a positive reinforcer (+1) would yield a large decrement in W in favor of $R0$, and a punisher (−1) would yield a large increment in W in favor of $R1$.

The predicted reinforcer rule adds what could be seen as a cognitive aspect to Sutton's reinforcement rule. Specifically, the predicted reinforcer rule adds a comparison of the present reinforcer, with the learner's expected reinforcer level based on its previous history of reinforcers. A member of this class is:

$$dW = c(R - P^R)(S^r - P^r), \qquad (6)$$

where the term $(S^r - P^r)$ represents the difference between the current reinforcer value (S^r) and the aggregated value of previous reinforcers (P^r). The value of P^r is updated on each time step according to the formula, $dP^r = b(S^r - P^r)$, where b is a growth rate parameter ($0 < b \leq 1$). Thus, in this learning rule, the effectiveness of the reinforcer is dampened if it corresponds to the net value of previous reinforcers and is amplified if it differs from the previous net value. In less quantitative terms, expected reinforcers are relatively ineffective, whereas unexpected reinforcers are more effective.

Across the simulated schedules of reinforcement, Sutton (1984) found that the second class of learning algorithms, those involving reinforcement comparison (Equation 6), converged the most rapidly upon $R1$, the response that simultaneously maximized the probability of positive reinforcement and minimized the probability of punishment. Of course, it remains to be seen whether any of the algorithms for machine learning can serve as a model for free-operant behavior under schedules corresponding to those simulated by Sutton. Conversely, it also remains to be seen whether any of the machine learning algorithms can generate the asymptotic matching relationships seen in more conventional concurrent schedules of positive reinforcement or punishment (e.g., Baum, 1974, 1979; de Villiers, 1980; Farley, 1980; Herrnstein, 1974; White & Pipe, 1987).

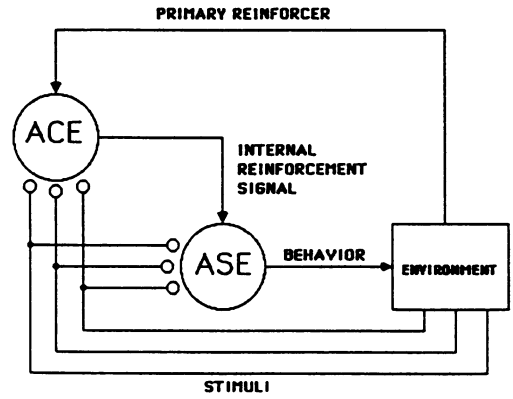

Fig. 6. A network for machine learning that mimics discriminant operant conditioning and conditioned reinforcement. The associative search element (ASE) emits operant-like actions, and the adaptive critic element (ACE) acts as a conduit for primary reinforcement and conditioned reinforcement signals to the ASE (Barto et al., 1983).

Mimics for the Discriminant Operant and Conditioned Reinforcement

In designing adaptive systems, Barto et al. (1983) have introduced algorithms that mimic the acquisition of discriminative operants and conditioned reinforcement. With regard to discriminative control, the control system is seen to be engaged in something akin to a foraging task, in which the system is learning which stimuli signal the reinforcer as well as which responses lead to the reinforcer (Barto & Sutton, 1981; Barto, Anderson, & Sutton, 1982; Sutton, 1984). With regard to conditioned reinforcement, Barto et al. (1983) argue that, in many control tasks as in many operant schedules, the primary reinforcer occurs relatively rarely and after a long sequence of actions. As in operant chain schedules, the use of conditioned reinforcers permits learning to occur in the absence of immediate primary reinforcement.

Figure 6 shows a schematic diagram for Barto et al.'s (1983) algorithm, with the labels of key features having been adapted to show their relation to conventional operant terminology. The right-hand box in Figure 6 rep-

resents the environment that provides discriminative stimuli and primary reinforcing events to the adaptive system. Each discriminative stimulus sends parallel projections to two adaptive units, designated as the associative search element (ASE) and the adaptive critic element (ACE). On the basis of the stimulus inputs and their connection weights, the ASE generates actions that operate on the environment. Subsequent primary reinforcement is delivered to the ASE via the output of the ACE, which is labeled the *internal reinforcement signal.* The ACE also detects stimulus–reinforcer relations and, as stimulus inputs become capable of triggering the ACE, it can also provide conditioned reinforcement to the ASE. Thus, the projection from the ACE to the ASE acts as a single channel for primary and conditioned reinforcement.

Inspection of Figure 6 reveals that each stimulus input can act both as a conditioned reinforcer by means of its connection weights with the ACE and as a discriminative stimulus for the operant via its connection weights with the ASE. Whereas many network models use the same learning and activation algorithms for all units, Barto et al. (1983) postulate distinctive algorithms for the ACE and the ASE. In rough terms, the ACE operates according to a stimulus–reinforcer contiguity principle like that of Rescorla and Wagner (1972). The ASE functions as an operant-like emitter according to a stimulus–response–reinforcer principle similar to Sutton's (1984) free operant-like rules. Moreover, the ASE does not rely exclusively on its weight values to generate behavior. Instead, the ASE uses a random process that is "merely *biased* by the combination of its weight values and the input patterns" (Barto et al., 1983, p. 837). According to Barto et al., the random component introduces a variety into the system's activity that is necessary for effectively searching the environment and for ultimately selecting appropriate responses through the reinforcement mechanism. In this way, Barto et al. have mimicked stimulus control over emitted responses.

As with Sutton's (1984) free operant-like algorithms, it is too early to tell whether the algorithms proposed by Barto, Sutton, and their colleagues can be adapted as models for discriminative operants and conditioned reinforcement. Nevertheless, the existence of their algorithms should provide considerable satisfaction to researchers in conditioning. Specifically, these algorithms for machine learning have explicitly incorporated key principles from operant and respondent research. In an indirect way, the use of behavioral principles in artificial systems confirms that those principles do reflect fundamental features of behavioral adaptation.

DISCUSSION

At present, connectionist modeling is most safely viewed as a set of promising quantitative tools rather than as a set of well-proven models. As with any quantitative technique, the way to success lies in identifying the ability of an application to unify previous findings and to guide experimentation along novel paths. Of course, failure can occur readily if the technique is used as just one more way to fit curves on an ad hoc basis. The large number of free parameters available in a connectionist model requires particular caution when engaging in a curve-fitting exercise. However, curve fitting can be used to test a model by determining whether a single set of parameter values can be used to fit multiple curves. As described with respect to Kehoe's (1988) model, such a test can be conducted both (a) on a between-group basis by attempting a simultaneous fit to multiple curves and (b) on a within-subject basis by determining whether parameters fitted for one schedule can predict performance in a second schedule.

Even with the potential pitfalls of curve fitting in mind, connectionist models offer considerable promise for research and theory in operant and respondent conditioning. To the extent that connectionist models become commonplace in the theoretical apparatus of cognitive psychology, neuroscience, and artificial intelligence, the use of connectionist models in conditioning will provide a valuable point of contact with those other areas. Moreover, connectionist models appear to be a natural development for both the experimental analysis of behavior and the associative learning traditions (Williams, 1987). First, connectionist models are rigorously quantitative in character. Although their total complexity grows as the number of units in a network increases, the equations characterizing each unit are rel-

atively small in number. Second, the concept of subsystems that interact through excitatory and inhibitory connections is already common in dual-process theories (e.g., Konorski, 1967; Overmier & Lawry, 1979; Rescorla & Solomon, 1967). Third, the learning rules in connectionist models are highly indebted to operant and respondent principles. Where the connectionist models provide a new twist on the old principles is in their repeated application to the units in a network rather than one application to the system as a whole. Fourth, connectionist models, particularly the models of Barto, Sutton, and their associates, have shown that is is possible to represent quantitatively the historical distinctions between operant versus respondent conditioning, revealing perhaps both their underlying similarities and differences.

From the perspective of connectionist modeling, studies of operant and respondent conditioning in animals provide a natural test bed for several reasons. Like an untutored network, animals can be introduced to training in a relatively naive state without the need for prior instructions. Likewise, the nonsymbolic nature of the inputs and outputs of connectionist models is routine in conditioning paradigms. Finally, in many conditioning procedures, it is possible to trace the course of learning on a trial-by-trial basis. For testing the workings of a network, the final solution is perhaps less interesting than observing the intermediate states of the system prior to the solution state.

In the current flush of enthusiasm for connectionist models, it is difficult to predict exactly where and how much solid achievement will occur. Nevertheless, theory and research in conditioning stand to be both contributors to and beneficiaries of those achievements. On the contributor side, the traditional rigor and simplicity of conditioning procedures provide a highly appropriate domain for the development of basic connectionist modeling techniques. On the beneficiary side, connectionist models of conditioning may help unify disparate bodies of data (e.g., Kehoe, 1988), guide experimentation into new channels (e.g., Kehoe, Schreurs, & Graham, 1987; Schreurs & Kehoe, 1987), and provide a basis for renewed interest in conditioning by workers in cognition and even more far-flung fields.

REFERENCES

Alsop, B., & Elliffe, D. (1988). Concurrent-schedule performance: Effects of relative and overall reinforcer rate. *Journal of the Experimental Analysis of Behavior,* **49,** 21–36.

Anderson, C. W. (1986). *Learning and problem solving with multilayer connectionist systems.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Anderson, J. A. (1973). A theory for the recognition of items from short memorized lists. *Psychological Review,* **80,** 417–438.

Anderson, J. A. (1977). Neural models with cognitive implications. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 27–90). Hillsdale, NJ: Erlbaum.

Anderson, J. A., & Rosenfeld, E. (Eds.). (1988). *Neurocomputing: Foundations of research.* Cambridge, MA: MIT Press.

Anderson, J. R. (1985). *Cognitive psychology and its implications* (2nd ed.). New York: Freeman.

Barto, A. G. (1985). Learning by statistical cooperation of self-interested neuron-like computing elements. *Human Neurobiology,* **4,** 229–256.

Barto, A. G., Anderson, C. W., & Sutton, R. S. (1982). Synthesis of nonlinear control surfaces by a layered associative search network. *Biological Cybernetics,* **43,** 175–185.

Barto, A. G., & Sutton, R. S. (1981). Landmark learning: An illustration of associative search. *Biological Cybernetics,* **42,** 1–8.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13,* 834–846.

Baum, W. M. (1974). On two types of deviation from the matching law: Bias and undermatching. *Journal of the Experimental Analysis of Behavior,* **22,** 231–242.

Baum, W. M. (1979). Matching, undermatching, and overmatching in studies of choice. *Journal of the Experimental Analysis of Behavior,* **32,** 269–281.

Bellingham, W. P., Gillette-Bellingham, K., & Kehoe, E. J. (1985). Summation and configuration in patterning schedules with the rat and rabbit. *Animal Learning & Behavior,* **13,** 152–164.

de Villiers, P. A. (1980). Toward a quantitative theory of punishment. *Journal of the Experimental Analysis of Behavior,* **33,** 15–25.

Estes, W. K. (1988). Toward a framework for combining connectionist and symbol-processing models. *Journal of Memory and Language,* **27,** 196–212.

Farley, J. (1980). Reinforcement and punishment effects in concurrent schedules: A test of two models. *Journal of the Experimental Analysis of Behavior,* **33,** 311–326.

Feldman, J. A. (Ed.). (1985). [Special issue on connectionist models and their applications.] *Cognitive Science,* **9,** 1–169.

Gelperin, A., Hopfield, J. J., & Tank, D. W. (1985). The logic of *limax* learning. In A. I. Selverston (Ed.), *Model neural networks and behavior* (pp. 237–261). New York: Plenum Press.

Gluck, M. A., & Bower, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory and Language,* **27,** 166–195.

Gluck, M. A., & Thompson, R. F. (1987). Modeling the neural substrates of associative learning and memory: A computational approach. *Psychological Review,* **94,** 176–191.

Hawkins, R. D., & Kandel, E. R. (1984). Is there a cell-biological alphabet for simple forms of learning? *Psychological Review,* **91,** 375–391.

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Heinemann, E. G., & Chase, S. (1975). Stimulus generalization. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 2. Conditioning and behavior theory* (pp. 305–349). Hillsdale, NJ: Erlbaum.

Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior,* **21,** 159–164.

Kehoe, E. J. (1986a). A layered network model for learning-to-learn and configuration in classical conditioning. In C. Clifton (Ed.), *Proceedings of the eighth annual conference of the cognitive science society* (pp. 154–175). Hillsdale, NJ: Erlbaum.

Kehoe, E. J. (1986b). Summation and configuration in conditioning of the rabbit's nictitating membrane response to compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes,* **12,** 186–195.

Kehoe, E. J. (1988). A layered network model of associative learning: Learning to learn and configuration. *Psychological Review,* **95,** 411–433.

Kehoe, E. J., & Gormezano, I. (1980). Configuration and combination laws in conditioning with compound stimuli. *Psychological Bulletin,* **87,** 351–378.

Kehoe, E. J., & Graham, P. (1988). Summation and configuration: Stimulus compounding and negative patterning in the rabbit. *Journal of Experimental Psychology: Animal Behavior Processes,* **14,** 320–333.

Kehoe, E. J., & Schreurs, B. G. (1986a). Compound-component differentiation as a function of CS-US interval and CS duration in the rabbit's conditioned nictitating membrane response. *Animal Learning & Behavior,* **14,** 144–154.

Kehoe, E. J., & Schreurs, B. G. (1986b). Compound conditioning of the rabbit's nictitating membrane response: Test trial manipulations. *Bulletin of the Psychonomic Society,* **24,** 79–81.

Kehoe, E. J., Schreurs, B. G., & Graham, P. (1987). Temporal primacy overrides prior training in serial compound conditioning of the rabbit's nictitating membrane response. *Animal Learning & Behavior,* **15,** 455–464.

Klopf, A. H. (1982). *The hedonistic neuron: A theory of memory, learning, and intelligence.* Washington, DC: Hemisphere.

Klopf, A. H. (1988). A neuronal model of classical conditioning. *Psychobiology,* **16,** 85–125.

Konorski, J. (1967) *Integrative activity of the brain.* Chicago: University of Chicago Press.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics,* **5,** 115–133.

Minksy, M. L., & Papert, S. (1969). *Perceptrons: An introduction to computational geometry.* Cambridge, MA: MIT Press.

Overmier, J. B., & Lawry, J. A. (1979). Pavlovian conditioning and the mediation of behavior. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 13, pp. 1–55). New York: Academic Press.

Pavlov, I. P. (1927). *Conditioned reflexes: An investigation of the physiological activity of the cerebral cortex* (G. V. Anrep, Trans.). London: Oxford University Press.

Rescorla, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative and Physiological Psychology,* **79,** 307–317.

Rescorla, R. A. (1973). Evidence for "unique stimulus" account of configural conditioning. *Journal of Comparative and Physiological Psychology,* **85,** 331–338.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review,* **74,** 151–182.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II* (pp. 64–99). New York: Appleton-Century-Crofts.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (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. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press.

Saavedra, M. A. (1975). Pavlovian compound conditioning in the rabbit. *Learning and Motivation,* **6,** 314–326.

Schreurs, B. G., & Kehoe, E. J. (1987). Cross-modal transfer as a function of initial training level in classical conditioning with the rabbit. *Animal Learning & Behavior,* **15,** 47–54.

Sutton, R. S. (1984). *Temporal credit assignment in reinforcement learning.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review,* **88,** 135–170.

White, K. G., & Pipe, M.-E. (1987). Sensitivity to reinforcer duration in a self-control procedure. *Journal of the Experimental Analysis of Behavior,* **48,** 235–249.

Wickens, D. D., Nield, A. F., Tuber, D. S., & Wickens, C. (1970). Classically conditioned compound-element discrimination as a function of length of training, amount of testing and CS–UCS interval. *Learning and Motivation,* **1,** 95–109.

Williams, B. A. (1987). The other psychology of animal learning: A review of Mackintosh's *Conditioning and Associative Learning. Journal of the Experimental Analysis of Behavior,* **48,** 175–186.

Woodbury, C. B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology,* **35,** 29–40.

Woodruff, G., & Williams, D. R. (1976). The associative relation underlying autoshaping in the pigeon. *Journal of the Experimental Analysis of Behavior,* **26,** 1–13.

Zipser, D. (1986). A model of hippocampal learning during classical conditioning. *Behavioral Neuroscience,* **100,** 764–776.

Received October 28, 1988
Final acceptance July 3, 1989