Psychological Review
1987, Vol. 94, No. 2, 176–191

0033-295X/87/$00.75

Modeling the Neural Substrates of Associative Learning and Memory: A Computational Approach

Mark A. Gluck and Richard F. Thompson
Stanford University

We develop a computational model of the neural substrates of elementary associative learning, using the neural circuits known to govern classical conditioning of the gill-withdrawal response of *Aplysia.* Building upon the theoretical efforts of Hawkins and Kandel (1984), we use this model to demonstrate that several higher order features of classical conditioning could be elaborations of the known cellular mechanisms for simple associative learning. Indeed, the current circuit model robustly exhibits many of the basic phenomena of classical conditioning. The model, however, requires a further assumption (regarding the form of the acquisition function) to predict asymptotic blocking and contingency learning. In addition, if extinction is mediated by the nonassociative mechanism of habituation—rather than the associative process postulated by Rescorla and Wagner (1972)—then we argue that additional mechanisms must be specified to resolve a conflict between acquisition and maintenance of learned associations. We suggest several possible extensions to the circuit model at both the cellular and molecular levels that are consistent with the known *Aplysia* physiology and that could, in principle, generate classical conditioning behavior.

Significant progress has been made in recent years in identifying and characterizing neuronal substrates of learning and memory, due in large part to the model biological system approach developed initially by Pavlov (1927) and by Lashley (1929). Lashley (1950) states the essence of the approach most simply in the following passage: "In experiments extending over the past thirty years, I have been trying to trace conditioned reflex paths through the brain or to find the locus of specific memory traces" (p. 455). The basic approach is to select an organism capable of exhibiting behavioral phenomena of learning and memory and whose nervous system possesses properties that make neurobiological analysis feasible (Alkon, 1980; Chang & Gelperin, 1980; Cohen, 1980; Goldman-Rakic, 1984; Hawkins & Kandel, 1984; Ito, 1982; Kandel & Spencer, 1968; Kandel, 1976; Mishkin, 1978; Sahley, Rudy, & Gelperin, 1981; Squire, 1982; Thompson et al., 1976; Thompson et al., 1984; Thompson & Spencer, 1966; Tsukahara, 1981; Woody, 1982). A chief advantage of model systems is that the facts gained from biological and behavioral investigations for a particular preparation are cumulative and tend to have synergistic effects on theory development and research (Thompson, 1986; Thompson, Donegan, & Lavond, in press).

Each approach and model preparation has particular advantages. The value of certain invertebrate preparations as model systems results from the fact that certain behavioral functions are controlled by ganglia containing relatively small numbers of large, identifiable cells, cells that can be consistently identified across individuals of the species (Alkon, 1980; Davis & Gillette, 1978; Hoyle, 1980; Kandel, 1976; Krasne, 1969). Knowing the architecture of the system—the essential neural circuits—means the neurons exhibiting plasticity can be identified and studied. With intact vertebrate model biological systems, these goals are more difficult to attain, but here, too, recent progress has been substantial (Cohen, 1980; Goldman-Rakic, 1984; Kapp, Gallagher, Applegate, & Frysinger, 1982; Mishkin, 1978; Schneiderman, McCabe, Haselton, & Ellenberger, in press; Squire & Zola-Morgan, 1983; Thompson, 1986; Thompson et al., in press).

A critical feature of the model biological system approach is circuit analysis, which involves tracing the neuronal pathways and systems that generate the learned response. The essential memory trace circuit for a given instance of associative learning may be defined as the necessary and sufficient circuitry for the particular instance of learning and memory, from sensory neurons to motor neurons. The memory trace itself, the essential neuronal plasticity that codes the learned response, is presumably contained within some subset of the essential memory trace circuit.

When an essential memory trace circuit has been defined in sufficient detail as a biological system, it becomes necessary to determine if the circuit will in fact generate the phenomena of

This research was supported by Office of Naval Research Grant N00014-83K-0238, a grant from the Sloan Foundation to R. F. Thompson, and a National Science Foundation Fellowship to M. A. Gluck.

We are indebted to many people who helped us with this article: to Nelson Donegan for many incisive and helpful comments on earlier drafts and for sharing with us his invaluable insights into animal learning theory; to Andrew Barto, Richard Granger, Robert Hawkins, Eric Kandel, Joseph Steinmetz, and Richard Sutton for their comments and suggestions on an earlier draft; and to Gordon Bower, Gregory Clark, Leon Cooper, Michael Foy, Stephen Kosslyn, Michael Mauk, Mortimer Mishkin, Misha Pavel, and Terry Sejnowski for stimulating discussions and valuable commentary. The assistance of Katie Albiston, Doug Jones, Niels Mayer, Patrick McQuillan, and Audrey Weinland is also gratefully acknowledged.

Correspondence concerning this article should be addressed to Mark A. Gluck, Department of Psychology, Stanford University, Building 420, Stanford, California 94305.

learning and memory that it is presumed to model. Even in elementary circuits, it is not always evident what the outcome of a given set of stimulus and training conditions will be at a qualitative–logical level of analysis. It would seem necessary to develop a quantitative computational model of the model biological circuit to determine more precisely what in fact the circuit can do. We report here our efforts to develop a computational model of the circuitry in *Aplysia* that exhibits elementary associative learning as identified by Kandel and associates (Carew, Hawkins, Abrams, & Kandel, 1984; Carew, Hawkins, & Kandel, 1983; Carew, Pinsker, & Kandel, 1972; Carew, Walters, & Kandel, 1981; Hawkins, 1981; Hawkins, Abrams, Carew, & Kandel, 1983; Hawkins, Castellucci, & Kandel, 1981; Kandel & Schwartz, 1982; Pinsker, Kupfermann, Castellucci, & Kandel, 1970; Walters & Byrne, 1983).

Classical Conditioning in *Aplysia*

The basic reflex studied in *Aplysia* is withdrawal of the siphon, mantle shelf, and gill to tactile stimulation of the siphon or mantle shelf. This withdrawal reflex is partly monosynaptic and can be obtained in a reduced preparation of the abdominal ganglion with sensory and motor neurons. Thus, siphon sensory neurons synapse directly on gill and siphon motor neurons and repeated activation of the sensory neurons results in habituation of the motor neuron response (Castellucci, Pinsker, Kupfermann, & Kandel, 1970). The mechanism is synaptic depression, a presynaptic process involving a decrease in transmitter release as result of repeated activation. This appears due in turn to a decreased Ca^{++} influx in the sensory neuron terminals (Klein, Shapiro, & Kandel, 1980). Sensitization, an increase in the motor neuron response to stimulation, is produced by stronger stimulation of the neck or tail (Pinsker et al., 1970).

As in the spinal flexion reflex (Thompson & Spencer, 1966), sensitization is a superimposed increase in excitability independent of habituation, in other words, dishabituation is an instance of sensitization. In *Aplysia*, sensitization is a result of a presynaptic action of interneurons on the sensory neuron terminals that results in an increased Ca^{++} influx, which is believed to be a result of activation of a cAMP-dependent protein kinase (Bernier, Castellucci, Kandel, & Schwartz, 1982; Castellucci, Nairn, Greengard, Schwartz, & Kandel, 1982; Hawkins et al., 1981; Kandel & Schwartz, 1982).

If weak stimulation of the sensory nerves (the Conditioned Stimulus, hereinafter referred to as CS) is followed by strong shock to the tail (the Unconditioned Stimulus, hereinafter referred to as US), the synaptic potential of the motor neurons to the CS is facilitated. Further, the action potential in the sensory neurons is broadened, indicating a presynaptic effect, which has been termed a pairing-specific enhancement of presynaptic facilitation (Hawkins, Abrams, Carew, & Kandel, 1983). If repeated paired trials are given, this enhancement increases above the level produced by US sensitization alone, yielding a basic phenomenon of classical conditioning, an associatively induced increase in response of motor neurons to the CS. This conditioning depends critically on the time between presentation of the CS and the US, as noted above.

The tail-shock US pathway involves interneurons that are thought to exert presynaptic action on the sensory nerve terminals. Hawkins and Kandel (1984) propose that the phenomena of conditioning in *Aplysia* result from the interplay of habituation and sensitization (in much the same way as Groves and Thompson (1970) suggested that the two processes interact) together with a third process, namely pairing specific enhancement of the excitability of the CS terminals. The mechanism is thought to be an action of the sensitization process on the CS terminal excitability temporally dependent on the occurrence of an action potential in the CS terminal, which could be characterized as a Hebb synapse (Hebb, 1949) on a terminal rather than on a neuron soma or dendrites.

Hawkins and Kandel (1984) suggest that the existence of unifying cell-biological principles underlying nonassociative and associative learning may suggest the beginnings of a "cell-biological alphabet" for learning, in which the basic units may be combined to form progressively more complex learning processes. In particular, they hypothesized that several higher order features of classical conditioning may be derivable from our understanding of the cellular mechanisms for associative learning in *Aplysia*. Our primary goal in this article is to provide a quantitative analysis of this alphabet hypothesis. We develop here a computational model of *Aplysia* circuitry and use it to test Hawkins and Kandel's specific hypotheses regarding possible mechanisms for differential conditioning, second-order conditioning, blocking, and contingency learning.

A long-term goal of our work is to develop quantitative computational models of the more complex learning and memory circuits in the mammalian brain (see Thompson, Berger, & Madden, 1983; Thompson, 1986; Thompson et al., in press). We use the *Aplysia* circuit in our initial work, in part as a heuristic to select an appropriate level of computational analysis, and because it is simpler and more fully characterized at a neurobiological level.

Parallel–Associative Network Models

In developing a computational model of the *Aplysia* circuit, we draw heavily on previous work in cognitive psychology and artificial intelligence on models of parallel-associative networks, often referred to as "connectionist models" (cf. Feldman & Ballard, 1982). Despite differences in terminology, goals, and methodology, these models all have in common the assumption that complex information processes can be generated by networks of simple nodes that pass, in parallel, a form of excitation from node to node. These models have gained increasing usage in recent years as a framework for modeling complex cognitive behaviors including visual recognition (Anderson, 1977; Anderson & Hinton, 1981; Anderson, Silverstein, Ritz, & Jones, 1977; Hinton & Anderson, 1981), the effects of context on letter perception (McClelland, 1979; McClelland & Rumelhart, 1981), and the representation of concepts by patterns of distributed activation (Rumelhart & McClelland, 1986). The aspect of our network modeling approach that is perhaps new and distinctive is the use of actual neuronal circuits, to the extent they are known, to provide the structure of the associative network. The network is constrained by the actual connections of an identifiable neuronal circuit.

Of particular relevance to our own efforts is the work of Sutton and Barto (1981) who describe a neural-like adaptive ele-

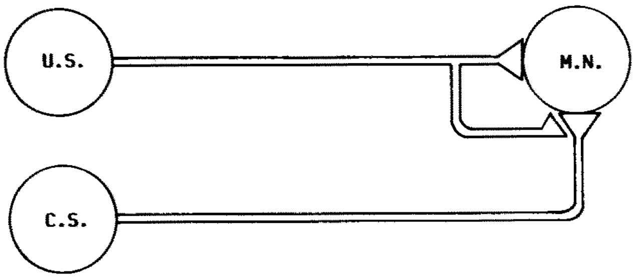

Figure 1. The basic circuit (Stage 1), composed of two sensory neurons (a conditioned stimulus, CS, and an unconditioned stimulus, US) and one motor neuron (MN).

ment more closely in accord with the animal learning literature than previous network models. As an extension of Rescorla and Wagner's (1972) trial-level model for associative learning, the Sutton–Barto model encodes the temporal dynamics of classical conditioning that are not captured by the Rescorla–Wagner model. In addition to providing a more detailed model of the real-time aspects of classical conditioning in animals, their element overcomes many of the stability and saturation problems encountered by network models. Though their model was designed to explain the same behavioral data as ours and used a similar computational framework, Sutton and Barto (1981) made no attempt to develop networks based on specific neural circuits nor was their adaptive element designed to behave in the manner of any identifiable neuron. To the extent that there are significant differences between our models, we might expect these differences to derive from the additional constraints of the relevant biological data.

Computational Model of *Aplysia* Circuit

We begin by describing a simple model of the circuit, including only the critical sensory and motor neurons for associative learning. After implementing this, and understanding what behavioral phenomena it does and does not account for, we add complexity, constrained by the neurobiological data.

The preliminary model consists of three neurons and three synapses, as represented in Figure 1. The neurons include a (to be) conditioned stimulus (the CS neuron), an unconditioned stimulus (the US neuron), and a motor neuron (the MN). One fiber originates at the conditioned stimulus and terminates as a synapse on the motor neuron (the CS→MN synapse). Two fibers originate at the unconditioned stimulus; one terminates as a synapse on the motor neuron (the US→MN synapse) and the other terminates as a synapse on the CS→MN synapse (the US→[CS→MN] synapse). The activations of both the CS and US sensory neurons are specified as input to the model. The primary output from the model is the activation of the MN in response to CS events and the CS→MN synaptic strength.

To implement the model on a digital computer, we adopt the convention of representing time as a series of discrete cycles of arbitrarily short duration. The activation of a neuron during cycle t is represented by $A(t)$ and is interpreted as the probability that the neuron will fire during that cycle. The state of a neuron, $S(t)$, is a binary number indicating whether or not the neuron fired during cycle t:

$$S(t) = \begin{cases} 1 & \text{with probability } A(t) \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Synapses are similarly represented by both a continuous value and a discrete state: Each synapse has a strength, $V(t)$, that is interpreted as the probability that an action potential generated by the presynpatic neuron will be passed postsynaptically. The state of a synapse, $P(t)$, is a binary value indicating whether or not an action potential was passed postsynaptically during cycle t:

$$P(t) = \begin{cases} S(t) & \text{with probability } V(t) \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Combining previous Equations 1 and 2 yields

$$P(t) = \begin{cases} 1 & \text{with probability } A(t)V(t) \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The probability that a neuron transmits an action potential postsynaptically is the product of its activation level and its synaptic strength.

Pairing-Specific Enhancement of Sensitization

Sensitization occurs at the CS→MN synapse according to

$$\Delta V_{CS}(t) = \begin{cases} \beta_1[1 - V_{CS}(t)] & \text{with probability } \Phi(t) \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where β_1 is a parameter governing the rate of sensitization and $\Phi(t)$ encodes the temporal specificity of conditioning as described below.[1] Like the Rescorla–Wagner (1972) model and the Sutton–Barto (1981) model, our model generates a negatively accelerating exponential function for the acquisition of the learned association. Kandel and Hawkins (1984) did not specify the form of the acquisition function in their model.

In our view, the effect of the interstimulus onset interval on conditionability is perhaps the most fundamental property of basic associative learning. The interval over which conditioning occurs varies widely in different preparations and paradigms, being minutes to hours for taste aversion, seconds to minutes for autonomic responses, and milliseconds to seconds for most skeletal muscle responses (see, e.g., Black & Prokasy, 1972; Hilgard & Bower, 1975). But regardless of the duration of the interval, there appears to be a relatively rapid rise and a slower decay of the conditionability function, as Clark Hull emphasized many years ago (Hull, 1943). In *Aplysia,* no learning occurs when the US precedes the CS or when the the two are presented simultaneously; optimal conditioning occurs when the CS precedes the US by .5 s and marginally significant conditioning occurs when the interstimulus interval is extended to 1 s (Hawkins, Carew, & Kandel, 1983).

To encode the temporal specificity of classical conditioning in *Aplysia,* it is necessary that the CS synaptic terminal have the potential to be modified in a pairing-specific manner that peaks some time after the synapse receives a pulse. The time course of this potential determines the possible interstimulus

[1] We use capital Roman letters for variables, capital Greek letters for functions, and lowercase Greek letters for fixed parameters.

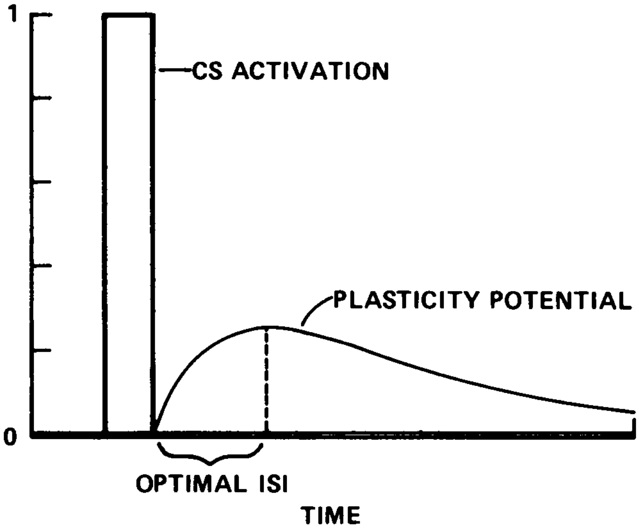

Figure 2. Time course of plasticity potential following a CS pulse with optimal interstimulus interval (ISI) indicated.

intervals (ISI). At this stage of model development we assume only that some mechanism must encode the temporal information and that this information must be present at the site of plasticity. The existence of this information is assumed without specifying the chemical or biological source. The temporal specificity of conditioning is governed in the model by $\Phi(t)$, the pairing-specific sensitization potential, which determines the degree to which US activity sensitizes the CS synapse. If an action potential is passed from the US neuron to the CS→MN synapse via the US→[CS→MN] synapse, then $\Phi(t)$ modulates the amount of sensitization as described above. $\Phi(t)$ is computed as

$$\Phi(t) = T(t)[1 - T(t)], \tag{5}$$

where $T(t)$ is a hidden variable whose default value is 0 but which jumps to 1 when a CS action potential is generated (i.e., $S_{CS}(t) = 1$) and then decays exponentially back to its default value according to

$$\Delta T = -\theta T, \tag{6}$$

where $\Delta T = T(t+1) - T(t)$ and θ determines the rate at which T decays to 0. Like the eligibility traces of Sutton and Barto's (1981) adaptive element model, $\Phi(t)$ acts as a temporal trace at the site of plasticity for encoding the previous occurrence of a CS. This formulation of $\Phi(t)$ as the product of an exponentially decaying function, $T(t)$), and an exponentially rising function, $(1 - T(t))$, produces a temporal specificity that conforms to the behavioral data: Conditionability rises quickly, peaks shortly after the CS event (i.e., when $T(t) = .5$ and $\Phi(t) = .25$), and then slowly decreases (see Figure 2). Presentation of the CS and US simultaneously does not produce effective conditioning because onset of the CS event sets $T(t) = 1$ and therefore the conditionability, $\Phi(t) = 0$.

As noted above, this process can be measured in the *Aplysia* circuit and is termed a "pairing-specific enhancement." More generally, such a process, in combination with a process of plasticity, must be postulated to account for the powerful effect of the interstimulus interval on conditionability (see above and Black & Prokasy, 1972). A more complete computational model of the mechanism for classical conditioning in *Aplysia* would need to include the details of the time course of the molecular processes that mediate this temporal specificity. In the final section of this article we discuss some recent progress in this direction by Gingrich and Byrne (1985).

Habituation

To model habituation, we extend the Groves–Thompson model to the neuronal level: Each time the CS synapse passes a pulse to the motor neuron, its strength decreases according to

$$\Delta V_{CS}(t) = \begin{cases} -\beta_2 V_{CS}(t) & \text{if } P_{CS}(t) = 1 \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where β_2 governs the rate of habituation.

Neuronal Firing

The firing rates of the sensory neurons constitute the input to the model. The firing rate of the motor neuron changes according to

$$\Delta A_{MN}(t) = \begin{cases} \delta_1[1 - A_{MN}(t)] & \text{if } (P_{CS}(t) = 1) \text{ or } (P_{US}(t) = 1) \\ -\delta_2 A_{MN}(t) & \text{otherwise,} \end{cases} \tag{8}$$

where δ_1 and δ_2 are the activation growth and decay rate parameters, respectively. The model as presently described has five free parameters: the neuronal activation increment and decrement rates (δ_1 and δ_2) the synaptic sensitization and habituation rates (β_1 and β_2), and a plasticity parameter, θ, that determines the time course of the ISI.

Associative Learning

At this level of detail, the circuit model is capable of producing the most basic associative learning phenomena exhibited by *Aplysia.* When a CS and US are paired with an appropriate interstimulus interval, pairing-specific learning occurs at the CS→MN synapse. The behavior of the model is shown in Figure 3A.[2] Initially the US produces a large amount of activity in the motor neuron compared to the small amount produced by the CS. After repeated presentations of the CS preceding the US, the motor-neuron response produced by the CS increases significantly. This change in the circuit's behavior can also be seen in the increased strength of the CS→MN synapse. Follow-

[2] In all the simulations of classical conditioning paradigms, we varied the full range of parameters in order to test the robustness of the phenomena and their sensitivity to parameter changes. All the basic associative conditioning phenomena were exceedingly robust across manipulations of the parameter values; changes in parameter values affected the rate and strength of conditioning and the duration of the optimal ISI but not the essential conditioning phenomena we sought to model. Parameters for the simulations shown in this article were set so as to generate a sample of conditioning in a sufficiently short number of trials to allow the complete protocol of the simulation to be shown in one figure.

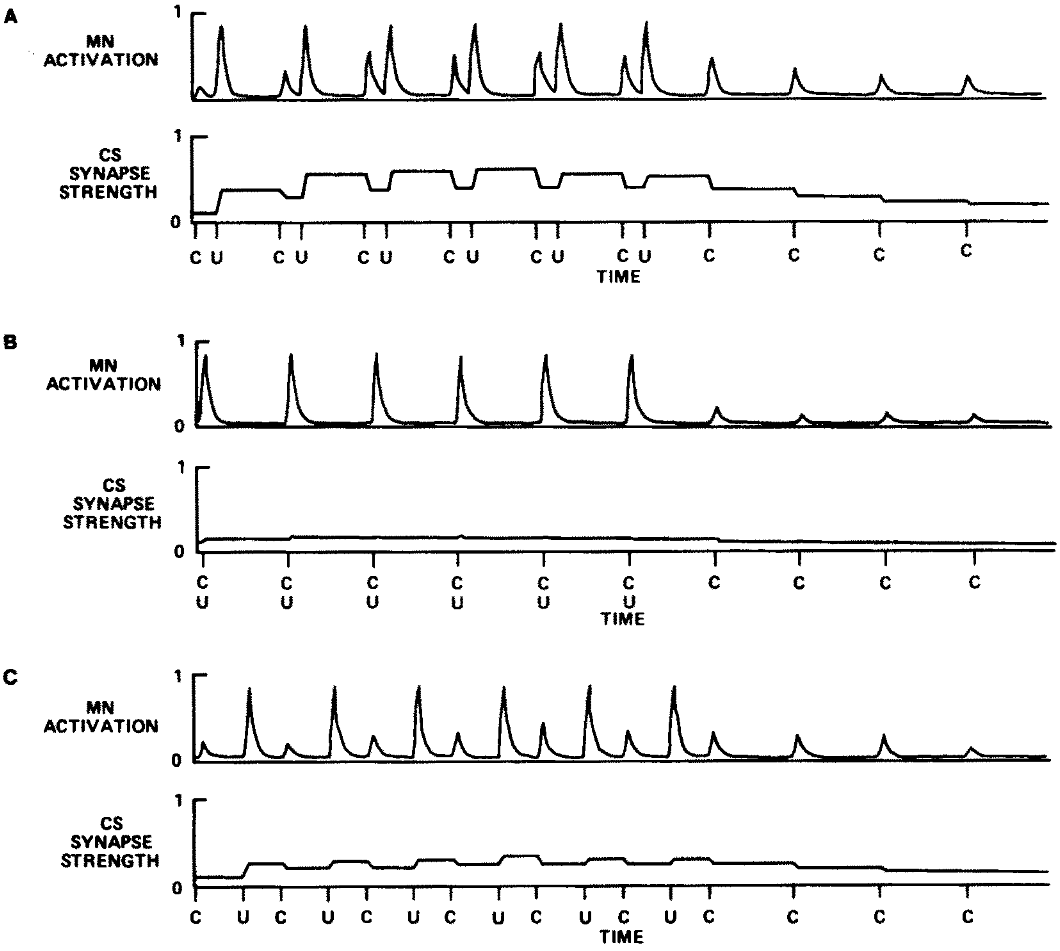

Figure 3. Stage 1 simulation of simple classical conditioning under three conditions: (a) optimal ISI produces maximal conditioning, (b) simultaneous presentation of CS and US (e.g., ISI = 0) produces essentially no conditioning, and (c) ISI much greater than optimal produces relatively small amount of conditioning. (For each of the three conditions, the time course of motor-neuron (MN) activation and the CS synaptic strength are shown. The CS (C) and US (U) pulses are indicated along the bottom of the graphs.)

ing this, repeated presentations of the CS alone habituates the CS→MN synapse strength with the consequence that motorneuron activity during presentation of the CS returns to its initial state. This extinction is mediated by the nonassociative process of habituation that occurs during CS presentations (Carew et al., 1972; Castellucci & Kandel, 1974).

No learning occurs in *Aplysia* when the CS and US are presented simultaneously (Hawkins et al., 1983). The model's behavior under this training paradigm is shown in Figure 3B. No learning occurs because the sensitization potential, $\Phi(t)$, is at 0 when the US fires, inhibiting pairing-specific sensitization. Figure 3C shows simulated conditioning with an ISI longer than optimal; some learning occurs, but less than with an optimal ISI.

This preliminary model omits many components of the full *Aplysia* circuitry, including interneurons and many of the fine-grained details of the molecular processes of nonassociative and associative sensitization as identified and characterized by Kandel and colleagues. This model is presented only as a first approximation of the *Aplysia* circuitry, detailed at a level of description comparable with the connectionist models used in cognitive psychology and artificial intelligence.

By beginning with this simple network model and evolving it as necessary to account for the relevant behavioral data, we hope to come to a greater understanding of the computational roles played by the different circuit components and neurobiological processes in mediating higher order features of conditioning. In the remainder of this article we describe our attempt to instantiate Hawkins and Kandel's (1984) hypothetical elaborations of this basic circuitry to see if they will, in fact, account for (a) differential conditioning, (b) second-order conditioning, and (c) blocking and contingency learning.

Differential Conditioning

In differential conditioning an animal learns to respond specifically to one reinforced stimulus and not to another nonreinforced stimulus. Only those sensory neurons that are active

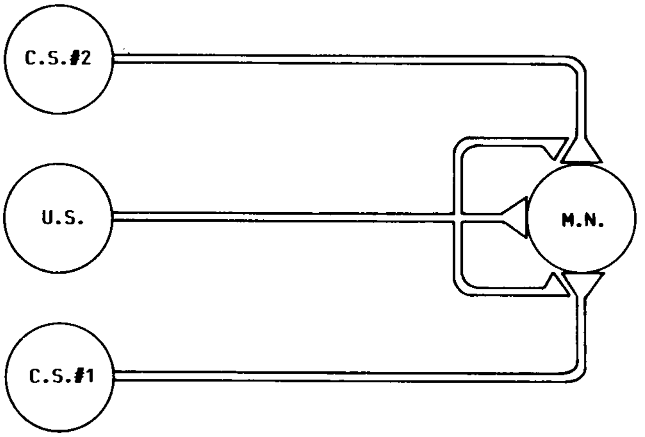

Figure 4. Stage 2 circuit with second CS added. (Both CS1 and CS2 are identically connected to the MN.)

prior to presentation of the US receive enhanced presynaptic facilitation. In *Aplysia,* a CS$^+$ is presented to the siphon paired with a US whereas an unpaired CS$-$ is presented to the mantle, or vice versa (Carew et al., 1983; Hawkins et al., 1983). To model this phenomena, it is necessary to include a second sensory neuron, CS2, connected to the motor neuron in a fashion similar to the other sensory neuron, CS1 (see Figure 4).

Successful differential conditioning using this model circuit is shown in Figure 5A. Activity in the CS1 neuron is paired consistently with the US, and activity in the CS2 neuron occurs randomly. This training procedure results in strong enhancement of the CS1→MN synapse (and thus in the ability of the CS1 neuron to generate activity in the motor neuron). Only mild nonassociative sensitization is produced in the CS2 synapse (Castellucci & Kandel, 1976; Pinsker et al., 1970).

Second-Order Conditioning

In second-order conditioning, a CS1 association is first trained via pairing with the US. After training, the CS1 can serve as a reinforcing stimulus to condition a new stimulus, CS2. For second-order conditioning to occur, the source of associative training could not logically be the US synapse or the CS1 could never gain the ability to sensitize the CS2. The results of Carew et al. (1984) implicate a presynaptic process. Hawkins and Kandel (1984) suggest that a facilitator interneuron plays the role of a local arousal system in second-order conditioning and serves as an intermediary between the sensory neuron (both CS and US) and the motor neuron. The facilitator interneuron produces facilitation not only at the sensory neuron synapses but also at the synapses from the facilitator neuron to itself. As shown in Figure 6 we include this interneuron in the model as an intermediary stage between the sensory and motor neurons. The same equations that were previously described for governing motor neuron activation and sensory to motor neuron synapses also govern the interneuron activation and the sensory to interneuron and interneuron to motor neuron synapses. The facilitator interneuron acts as the single source of sensitization enhancement for the synapses of the sensory neurons and the interneuron. One implication of this is that, unlike the CS synapses that terminate on the motor neuron, the CS synapses that terminate on the facilitator interneuron act as Hebb (1949) synapses, because firing of the CS neuron prior to firing of the facilitator interneuron enhances the ability of the CS neuron to activate the interneuron (Hawkins & Kandel, 1984, p. 384).

With the addition of the facilitator interneuron, the model successfully produces second-order conditioning (see Figure 7). The CS1 is paired with the US until an asymptotic level of conditioning is reached. Following this, the CS2 is paired with the CS1 and the US is omitted. Because the CS1 has now acquired the ability to act as a source of pairing-specific enhancement of sensitization, conditioning occurs at the CS2 synapses via the interneuron, and the CS1 synapse strength slowly extinguishes as a result of habituation and nonreinforcement.

Blocking

A class of behavioral phenomena exists that indicates animals learn not just the temporal contiguity of stimuli but also their predictive or informational value. Hawkins and Kandel (1984)

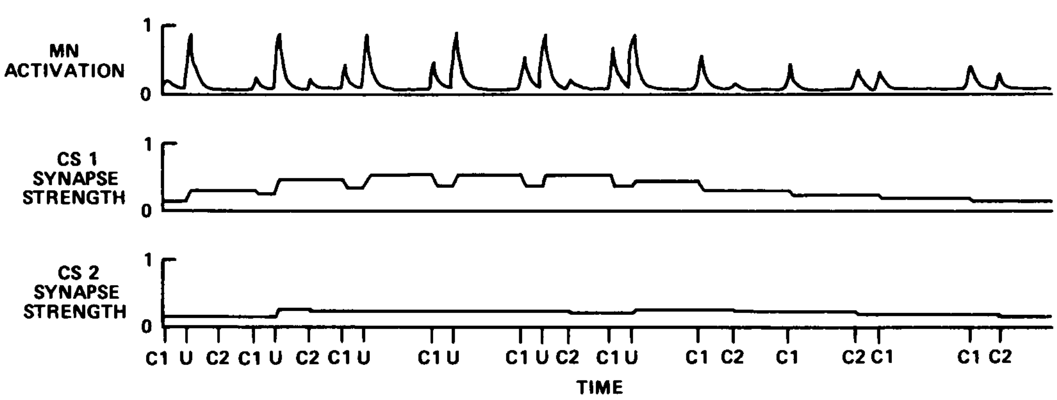

Figure 5. Stage 2 simulation of differential conditioning of the CS1 and CS2. (Three graphs are shown: The time course of motor-neuron (MN) activation and the synaptic strengths of CS1 and CS2. The CS1, CS2, and US pulses are indicated along the bottom by C1, C2, and U, respectively.)

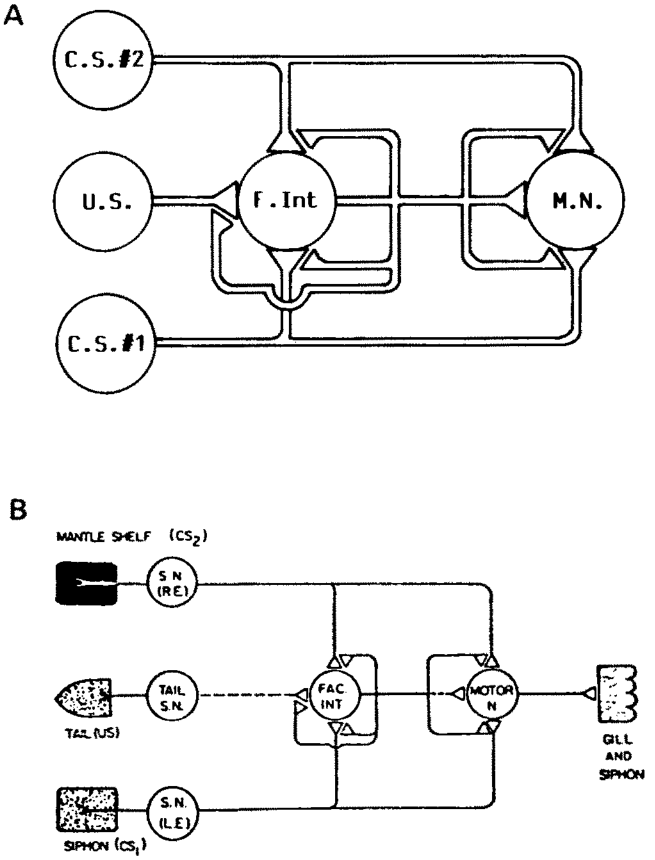

Figure 6. (a) Stage 3 circuit model with facilitator interneuron (F. Int) added; (b) *Aplysia* circuit (from Hawkins & Kandel, 1984). (The sensory neurons (S.N.) from the siphon and mantel shelf are the CS1 and CS2 input lines. The sensory neuron at the tail is the US. The output of the motor neuron excites the gill and siphon withdrawal reflex.)

proposed an elaboration of their basic cellular model that they suggest could account for these higher order features. They illustrate this hypothetical mechanism using blocking, whereby an animal learns not only about the contiguity of stimuli but also about their predictive contingency: If a CS1 is conditioned to predict the US, then the addition of a second stimulus, CS2, presented simultaneously with CS1, results in attenuated conditioning to the CS2 alone. A form of blocking has recently been reported for a behavioral response in intact *Aplysia,* but the magnitude of the effect was not described (Colwill, 1985).

In cognitive terms, this phenomena is described as a lack of association strength accruing to a stimulus that provides no new predictive power (Kamin, 1969). In the Rescorla-Wagner model of classical conditioning, this is formalized as

$$\Delta V_j = \alpha_j \beta_1 (\lambda - \sum V_i) \tag{9}$$

where ΔV_j is the change in the association strength between a stimulus element j and the US, α_j is the salience of the CS element, β_1 is a parameter governing the rate of learning during US presentation, λ is the maximum possible association strength associated with the US, and ΣV_i is the sum of the association strengths between the CS stimulus elements occurring on that trial and the US. In the initial phase of training, CS1 is paired with the US until V_1 approaches λ. In the compound (CS1 + CS2→US) phase of training, CS2 will acquire little associative strength since pretraining on CS1 results in $(\lambda - \Sigma V_i \approx 0)$. To demonstrate blocking, however, it is not necessary that there be an absence of conditioning to CS2; this is only one extreme case that satisfies the definition. More generally, blocking is observed when responding to CS2 is less for subjects who have had pretraining on CS1 than for subjects not pretrained on CS1.

Hawkins and Kandel (1984) suggest that the interneuron may implement the $(\lambda - \Sigma V_i)$ component of the Rescorla-Wagner algorithm in the following way: After being activated by the CS1, the interneuron undergoes a refractory period—caused perhaps by accommodation and recurrent inhibition—that persists throughout the presentation of the US. They speculate that this could mediate the blocking effect if the firing of the interneuron by the CS1, and its resulting inhibition, attenuated the sensitization that accrues to the CS2. Activation of the interneuron during the compound CS1 + CS2 stimulus occurs outside of the window of eligibility for ΔV_{CS2} and the interneuron is refractory during US stimulation, preventing interneuron activation at a time favorable to pairing-specific modification of the synaptic strength of CS2.

In modeling the refractory period of the interneuron we are concerned only with the resultant firing behavior and not the mechanisms for accommodation and recurrent inhibition. We introduce here an additional variable, R_{FI}, which represents the degree to which firing of the facilitator interneuron is inhibited. When the activation level of the facilitator interneuron exceeds a predetermined threshold, R_{FI} is set to 1. R_{FI} then decays slowly back to 0. In the computational model, R_{FI} affects the interneuron by probabilistically governing the growth of interneuron activation in the following manner:

$$\Delta A_{FI}(t) = \begin{cases} \delta_1[1 - A_{FI}(t)] & \text{with probability } (1 - R_{FI}) \text{ if } [P_{CS}(t) = 1] \text{ or } [P_{US}(t) = 1] \\ -\delta_2 A_{FI}(t) & \text{otherwise,} \end{cases} \tag{10}$$

where δ_1 is the rate parameter for activation increase and $\Delta A_{FI}(t)$ is the change in $A_{FI}(t)$, the current activation level of facilitator interneuron. As long as R_{FI} is near 1, the interneuron will be inhibited from firing. To produce the appropriate blocking behavior the decay rate of R_{FI} must be set so that the refractory period of the interneuron is longer than the possible interstimulus interval.

In the Hawkins and Kandel model, associative activation produces a graded refractoriness in the interneuron proportional to the strength of associative activation. A single behavioral trial is characterized in our model by multiple cycles of the simulation. Our model produces a somewhat continuous effect of associative strength on the degree of refractoriness that is proportional to the degree of overlap between interneuron activity and CS eligibility.

The implementation of a refractory period for the interneuron longer than the acceptable ISI necessitates the explicit inclusion of a direct US→MN connection (see Figure 8) in order to get an appropriate unconditioned response to the US. Although these pathways exist, they are often not represented in less ex-

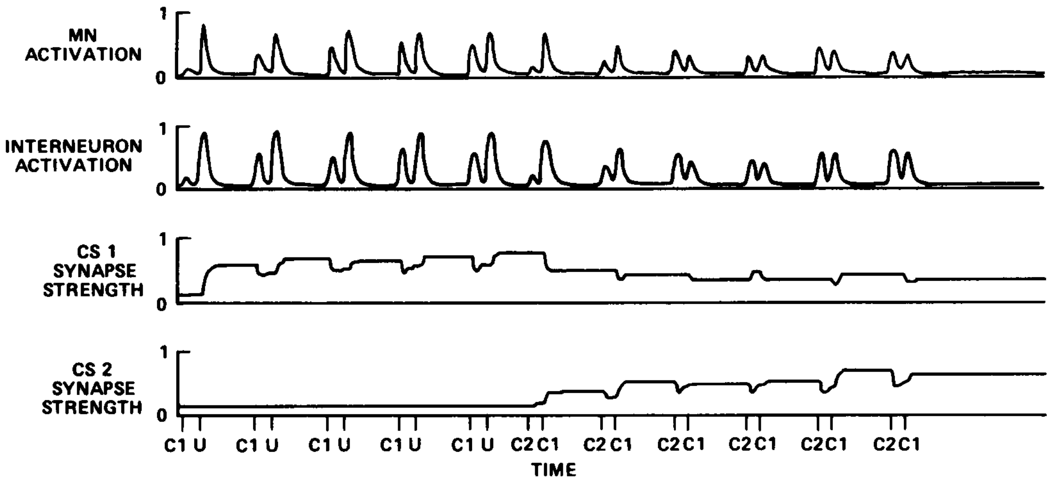

Figure 7. Stage 3 simulation of successful second-order conditioning of CS2 to CS1, showing the time course of the motor-neuron (MN) activation, the interneuron activation, and the CS1 and CS2 synaptic strengths. (The CS1, CS2, and US pulses are indicated along the bottom by, C1, C2, and U, respectively.)

plicit models of the circuitry for learning and memory. As shown in Figure 9, the addition of the implementation of this refractory behavior still allows for normal conditioning and second-order conditioning.

However, contrary to expectations, this circuit model failed to produce asymptotic blocking. Across variations of all the relevant parameters the asymptotic levels of conditioning after extended compound training were identical for CS1 and CS2 and indistinguishable from the levels attained without pretraining to the CS1.

However, as suggested by Nelson Donegan (personal communication, March 19, 1986) the model is capable of a short-term preasymptotic form of blocking. Figure 10A shows a trial-by-trial analysis of the changes in activation levels and synaptic strengths for a simulated blocking experiment using the same parameter setting that produced the successful second-order conditioning shown in Figure 9. In this simulation both CS and US events were 5 cycles long and the interstimulus interval was 30 cycles, an interval optimally favorable to associative learning given the time course of eligibility (as determined by θ). The rate of synaptic sensitization is $\beta_1 = .4$ and the rate of habituation is $\beta_2 = .05$, parameter settings that in previous simulations were sufficient to produce second-order conditioning.

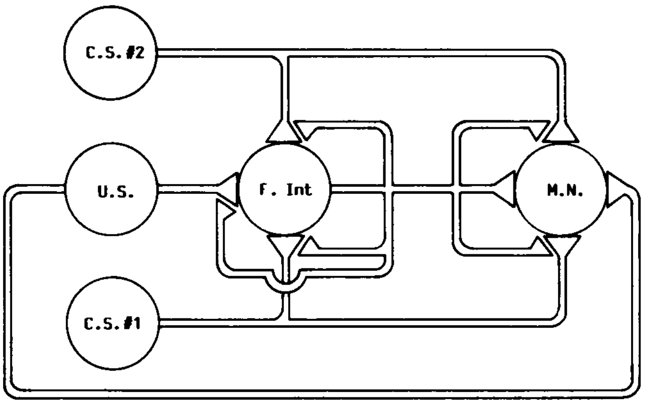

Figure 8. Stage 4 circuit with direct US→MN pathway.

After eight CS1→US trials, the CS1 synaptic strength reached an asymptotic level. During this period both the interneuron and motor neuron responses to CS1 increased considerably. Hawkins and Kandel suggest that as training to the CS1 reaches asymptote, activity in the interneuron during US presentation will disappear. Although the interneuron response during US presentation does decrease significantly during training, it cannot be entirely eliminated. If the associative strength of the CS stimulus were strong enough to entirely eliminate the interneuron response during US presentation, there would be no source of pairing-specific sensitization for the CS terminals. This would be an inherently unstable state because the nonassociative process of habituation would drive down the CS associative strength until enough interneuron activity occurred during the eligibility period for CS synapse modification (i.e., during US presentation) to offset the habituation. Thus, total refractoriness of the interneuron is not consistent with the basic mechanisms for strength revision. Interneuron activity during the US presentation does not entirely disappear, but rather decreases to a level where it is just sufficient to offset the effects of habituation. This can be seen in both the simulation of second-order conditioning shown in Figure 9 and in the simulation shown in Figure 10.

Following pretraining to the CS1 stimulus, 12 compound CS1 + CS2→US trials were presented. As is clear from the time course of CS2 synaptic strength, the CS2 synapse gains considerable associative strength; the important comparison, however, is with the time course of CS2 synaptic strength during compound training without pretraining to CS1 (shown in Figure 10 as a dashed line). Although these two curves reach the same asymptotic levels after extended compound training, there is an initial preasymptotic period in which the CS2 strength is below that found without pretraining to the CS1.

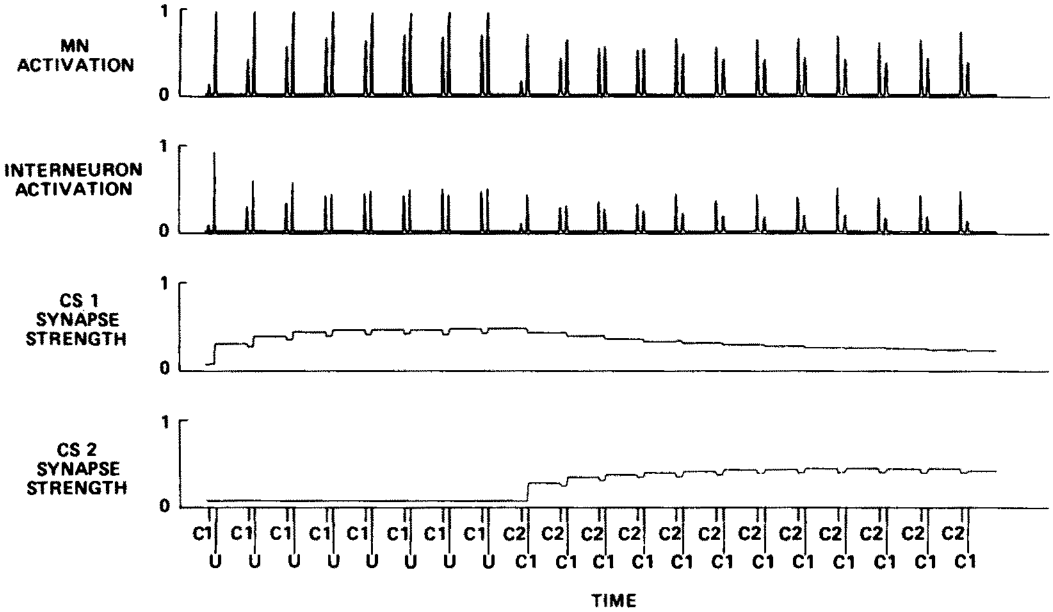

Figure 9. Successful second-order conditioning with refractory interneuron. (Eight CS1–US pairings are followed by 12 CS2–CS1 pairings. Note that at asymptotic CS1 conditioning (Trial 8), interneuron activity is attenuated but not eliminated during US presentation. This graph shows only the first half of the paradigm in which the decreasing strength of the CS1 association is still strong enough to maintain the CS2 association. Repeated pairings of CS2–CS1 would eventually habituate the CS1 association and, in turn, the CS2 association. Parameter settings: sensitization rate (β_1) = .4, habituation rate (β_2) = .05, ISI factor (θ) = .15, neuron activation increment and decrement (δ_1, δ_2) = .8 and .6. Initial synaptic strength settings: $V_{US} = 1$, $V_{CS_1} = V_{CS_2} = .05$. Figure shows average values of variables for 100 repetitions of paradigm.)

Thus, initial compound trials do produce a limited form of blocking. Because pretraining to CS1 attenuates interneuron activity during the eligibility period, the change in associative strength available to the CS2 (as measured by interneuron activity) is significantly less on the first trial than if there had been no pretraining to CS1, in much the same way that Hawkins and Kandel (1984) suggest. By the second trial, the CS2 has acquired some associative strength because of interneuron activity during the first trial. Because there is now less interneuron activity during the US presentation because of the combined effects of the compound stimuli, the CS1 strength decreases slightly. The CS2 strength, however, rises because (a) the absolute effect of habituation is smaller for a weak association than for a strong association, the rate of habituation being proportional to the absolute level of associative strength (Groves & Thompson, 1970); and (b) the sensitizing effect of the interneuron is greater for a weak association than for a strong association because of the negatively accelerated growth function for strength revision. This rise in the CS2 strength along with the slight decay in the CS1 strength continues until the associative strengths of the two stimulus elements are equal and stable: The combined effect of the stimulus elements decreases the level of interneuron activity until its sensitizing ability just counteracts the effects of habituation. With extended training on the compound stimulus, the effect of pretraining to the CS1 diminishes and disappears asymptotically. As is apparent from the simulation results, there is no difference in the CS2 synapse strengths between those subjects who were pretrained to the CS1 and those who were not.

The magnitude of this preasymptotic blocking is dependent on the relative strengths of the sensitization and habituation parameters: As shown in Figure 10B, when the sensitization rate is decreased (relative to the previous simulation), the absolute difference between the pretrained and nonpretrained conditions is less on any given trial. However, the total number of trials on which there is a significant difference between the two conditions is increased, in other words, it takes longer to reach the equilibrium point.

The failure of our computational model to produce the asymptotic blocking does not make a convincing case that the real circuit is unable to produce this behavior, nor does it make a strong argument that any formal model consistent with the model proposed by Hawkins and Kandel (1984) will be unable to produce blocking. Rather it only supports the weak claim that this particular form of the model does not robustly predict asymptotic blocking in that there exist interpretations of the model that do not produce this behavior.

Circuitry Models and Behavioral Learning Algorithms

In suggesting neuronal mechanisms for higher order features of classical conditioning, Hawkins and Kandel (1984) were guided by an attempt to identify possible neuronal-level correlates of Rescorla and Wagner's (1972) behavioral model of clas-

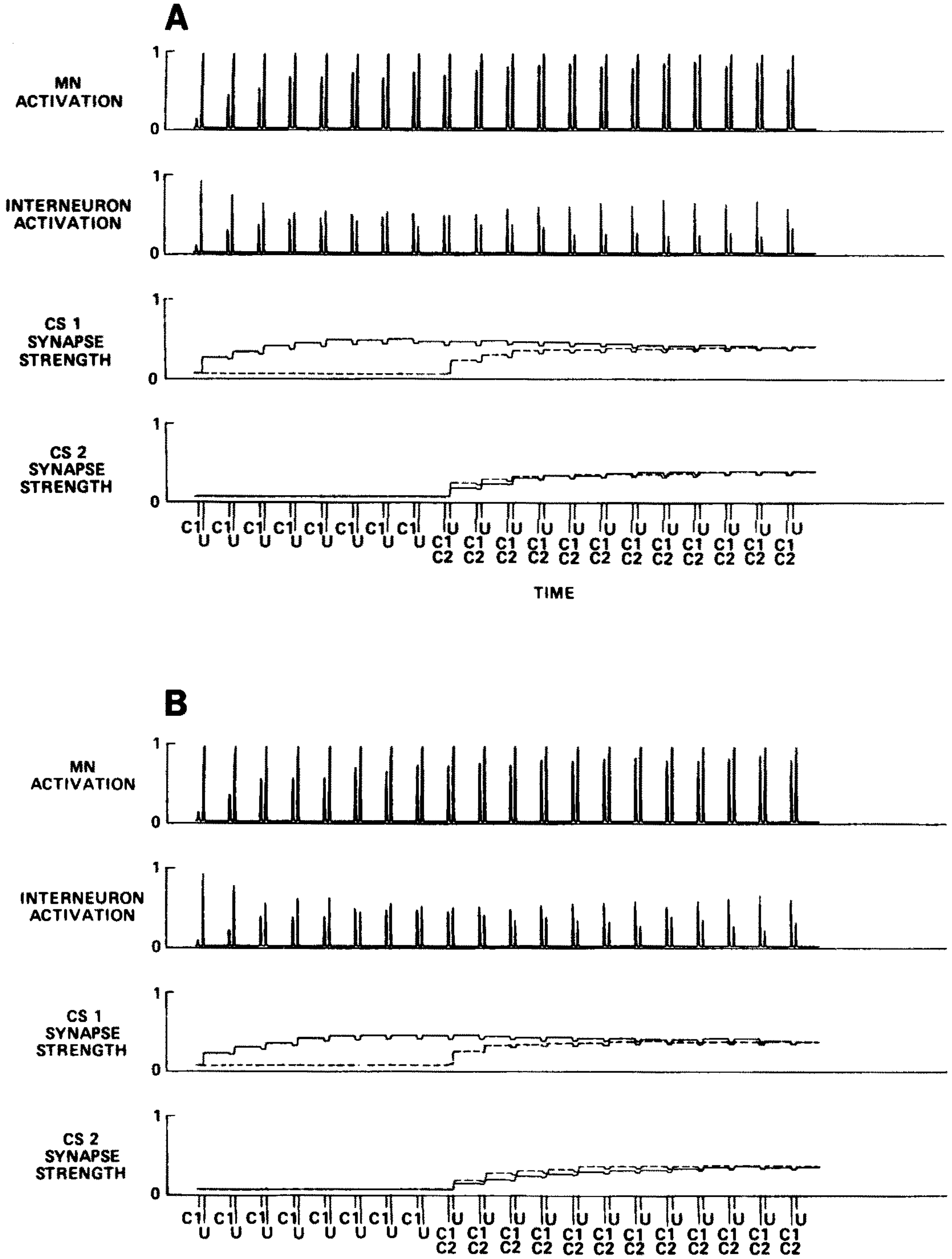

Figure 10. (a) Pre-asymptotic blocking with refractory interneuron. The solid lines graph the MN activation, interneuron activation, and CS1 and CS2 synapse strengths for 8 CS1–US pairings followed by 12 compound CS1/CS2–US pairings. The dashed lines show the CS1 and CS2 synaptic strengths from the simulated control experiment in which the 12 compound pairings were not preceded by CS1 pretraining. With the same parameter settings used in Figure 9, conditioning to the CS2 is attenuated on the first few trials but difference disappears after a few trials with no significant asymptotic difference between the pretrained and nonpretrained conditions. (b) Decreasing the sensitization rate (β_1) to .3 from .4 decreases the magnitude of the blocking effect on the first trial but increases the number of trials before the pretrained and nonpretrained conditions asymptote at the same level.

sical conditioning. The Rescorla–Wagner model does, however, predict blocking; it was proposed because previous simpler models of associative learning could not account for blocking. To the extent that the emergent behavior of the circuit-level model (Hawkins & Kandel, 1984) differs from both the behavior of the learning algorithm that inspired it (Rescorla & Wagner, 1972) and the target animal behavior, it might be fruitful to examine more closely the relationship between these two models.

Hawkins and Kandel's (1984) model of the *Aplysia* circuitry and the Rescorla–Wagner behavioral model differ primarily regarding the mechanisms for negative learning: The Rescorla–Wagner model posits two associative processes by which association strengths can be weakened, whereas the Hawkins–Kandel model posits one nonassociative process for the weakening of association strengths. In the Rescorla–Wagner model, the presentation of a spurious CS (i.e., not followed by a US) changes the association strength of a CS element, *j*, according to

$$\Delta V_j = -\alpha_j\beta_2\sum V_i \qquad (11)$$

where α_j is as before, β_2 is a parameter governing the rate of extinction and ΣV_i is the sum of the association strengths of the CS stimuli. One nonintuitive prediction of the Rescorla–Wagner model is that negative learning can occur during a positive CS–US presentation if the sum of the association strengths between the CS and US elements is greater than the maximum association strength available for the US (i.e., $\Sigma V_i > \lambda$). Kamin and Gaioni (1974) confirmed this by demonstrating that the novel compounding of two independently trained CS stimulus elements produces an overexpectancy effect resulting in a decrement of associative strength for the component stimuli. In summary, the Rescorla–Wagner model posits that decreases in associative strength always occur as the result of an associative process both on positive (i.e., US present) and negative trials. In contrast, the Hawkins–Kandel model of the *Aplysia* circuitry proposes that all decreases in association strength are mediated by the nonassociative process of habituation. As Hawkins and Kandel note (1984, p. 386), their model in this regard is more closely in accord with the Groves–Thompson (1970) model of habituation and sensitization than with the Rescorla–Wagner model. Adapting the Groves–Thompson behavioral model of habituation to the neuronal level, we have modeled habituation as

$$\Delta V_j = -\beta_2 V_j, \qquad (12)$$

where the notation is as before.

The Rescorla–Wagner model predicts that an asymptotic level of conditioning to a CS element will be reached when the sum of the CS elements equals the maximum level of conditioning available for the US (i.e., λ). As Hawkins and Kandel note, their model differs in this regard in that the asymptotic level of conditioning is predicted to occur when there is an equilibrium between the associative process of pairing-specific enhancement of sensitization and the nonassociative process of habituation (Hawkins & Kandel, 1984, p. 386). Actually, the ability of a conditioned CS element to inhibit interneuron firing during US presentation provides an additional source of modulation for the growth of the CS synaptic strength. The asymptotic level of conditioning will occur in the *Aplysia* model when three interacting processes are in equilibrium: (a) habituation, (b) sensitization of the CS synapses via the facilitator interneuron, and (c) attenuation of pairing-specific sensitization of the CS synapses due to inhibition of facilitator interneuron activity by the conditioning of the CS synapses.

What is the implication of this for the blocking paradigm? For blocking to occur, there should be an attenuation of conditioning to the CS2 due to prior training to the CS1. According to the Hawkins and Kandel hypothesis, this should occur because the presentation of the previously conditioned CS1 stimulus will inhibit the interneuron from firing during the US presentation, eliminating the necessary source for pairing-specific learning. As we have seen in the simulations, however, it is necessary that there be some activity in the facilitator interneuron in order to maintain an equilibrium between pairing-specific enhancement of sensitization, which occurs during US presentations, and habituation, which occurs during CS presentations. If the maintenance of a conditioned response in *Aplysia* depends on an equilibrium between pairing-specific sensitization enhancement and habituation, then a process that eliminates or attenuates the source of pairing-specific sensitization will clearly change the equilibrium point. As a tentative hypothesis we suggest that the reason our quantitative simulation of their model circuit does not generate asymptotic blocking is that there is a conflict between the need to maintain interneuron firing during US presentation in order to maintain a learned association and the need to inhibit the interneuron from firing during the US presentation in order to resist the acquisition of a new association.

It appears that the conflict between acquisition and maintenance is attributable to the difference between the associative algorithm for extinction proposed by Rescorla and Wagner (1972) and the nonassociative circuitry mechanism proposed by Hawkins and Kandel (1984); this explains why the Rescorla–Wagner model—and Sutton and Barto's (1981) temporal extension of this model—both avoid the conflict. If the basic mechanisms for associative learning in *Aplysia*—an associative process for sensitization enhancement and a nonassociative process for habituation—are in fact the building blocks for classical conditioning, then we suggest that additional mechanisms must be identified and characterized that allow the circuit to functionally distinguish between acquisition and maintenance, that is, to give the circuit a way of differentially affecting novel and pretrained stimuli in a manner consistent with the behavioral data. In a speculative vein, we consider here two classes of extensions to the Hawkins–Kandel model, consistent with the known *Aplysia* physiology, which might produce blocking.

Single-Interneuron Models

If both retention and acquisition are governed by the same interneuron, as Hawkins and Kandel (1984) suggest, then the activity of this interneuron during US presentation must be sufficient to maintain the CS1 association but insufficient to acquire the CS2 association. The interneuron cannot be totally turned off (or the CS1 association would extinguish) nor can it be left entirely on or there would be no blocking of the CS2 association. This implies that if a single facilitator interneuron mediates both the acquisition of new conditioned pathways and the maintenance of previously learned pathways, then it must settle at an intermediate level of firing (i.e., less than for an unpredicted US). Furthermore, this activity must have a differential effect on the CS1 and the CS2; in other words, it must maintain the learned CS1 associations but resist in the acquisition of the new CS2 association. Because activity in the CS1 and CS2 neurons will be the same, the differential effect of the activity in the interneuron must be attributable to the different strengths of the CS1 and CS2 synapses. More precisely, the activity in the facilitator interneuron during the US presentation must be sufficient to maintain the stronger CS1 association but insufficient to strengthen the weaker CS2 association.

Within this single-interneuron framework, we consider one possible modification to the model of synaptic plasticity that would make it easier to maintain an old association than to acquire a new association. As discussed earlier, our model of the learning mechanisms for positive synaptic weight change follows the models of Rescorla and Wagner (1972) and Sutton and Barto (1981) in using a negatively accelerated acquisition function. Learning in these models is faster and easier for a novel association than for an existing association: The opposite of what we suggest is necessary to mediate blocking with a single source of sensitization. Because any learning curve must eventually be negatively accelerated to reach an asymptote, the critical issue is what happens during the early stages of learning. Negatively accelerated growth in the Rescorla–Wagner model is not behaviorally unrealistic: The model tracks associative strengths, not behavior. It is assumed that additional assumptions are necessary to map associative strengths to behavioral measures such as response probability or response strength. For example, a recent extension to the Rescorla–Wagner model (Frey & Sears, 1978) incorporates a mapping of the negatively accelerated growth of associative strength to a more behaviorally realistic S-shaped ogive learning curve that at first is positively accelerated (cf. Mackintosh, 1974). The initial positively accelerated learning curve could be implemented at the neuronal level by making the rate of sensitization (conditional on the temporal trace)

$$\Delta V_j = \beta_1[V_j(1 - V_j)], \tag{13}$$

rather than using the negatively accelerated rule given in Equation 4. A precondition for using this learning function, however, is that V_j may approach but never equal 0. The motivation for using this learning rule would be to make learning more difficult for a novel stimulus element than for a previously trained stimulus element. If the interneuron is only slightly active during US presentation, this might be sufficient to maintain a learned association but insufficient to acquire a novel association. We incorporated this rule within the model, and the resulting successful simulated blocking behavior is shown in Figure 11. Initial presentation of a single CS1 stimulus causes small bursts of activity in the interneuron and motor neuron relative to the more significant effects on these two neurons from presentation of the US. Repeated pairings of the CS1 and US result in a significantly increased asymptotic level of CS1 synaptic strength along with increased responsiveness of the motor neuron to the CS1 stimulus. The interneuron, which previously fired during US presentations, now fires primarily during CS1 presentations with significantly attentuated firing during the US. Following this, repeated compound trials, CS1 + CS2 → US, produce no significant change in the CS2 synapse strength, a clear case of blocking.

Multiple Interneuron Model

Though Hawkins and Kandel (1984) limited themselves to considering the possible functional significance of a single interneuron, many other interneurons exist that could in principle contribute to higher order features of classical conditioning, including at least four interneurons that receive excitatory input from the sensory neurons and two inhibitory interneurons (Hawkins et al., 1981). It seems reasonable, therefore, to ask what functional properties an additional interneuron might possess that would contribute to resolving the acquisition/maintenance conflict? We may speculate here about one possible mechanism: If a second interneuron does not go into a refractory period and sensitizes the CS synapses proportional to the current learned association, this might counteract the effect of the habituation of an already learned association but have little or no effect on an unlearned association. We implemented this in our model by adding an additional interneuron, FI_2, which is connected just like the original interneuron, FI_1, which provides an additional source of pairing-specific enhancement of presynaptic facilitation according to

$$\Delta V_{CS}(t) = \begin{cases} \beta_1[V_{CS}(t)] & \text{with probability } \Phi(t) \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

where Φ and β_1 are defined as before.

We implemented this additional interneuron and the behavior of the circuit in a blocking paradigm is shown in Figure 12. The circuit clearly generates an extreme case of blocking. We note that the asymptotic level of conditioning accruing to the CS1 is significantly higher in this model than in the previous model (without the additional interneuron); however, without further comparisons of these two models in a variety of additional behavioral paradigms, it is unclear that this is of any theoretical interest.

Contingency Learning

As discussed earlier, blocking is just one example of a class of behavioral phenomena, including overshadowing and the effect of US-alone trials, in which animals learn about the contingency or informational value of stimuli (Prokasy, 1965; Rescorla, 1968) rather than simply their contiguity or co-occurrence (Hull, 1943; Spence, 1956). The importance of the Rescorla–Wagner model is that it posits a single process that accounts for the role of these informational variables in addition to predicting a wide range of additional effects, especially those dealing with the learning of inhibitory associations. Similarly, Hawkins and Kandel (1984) propose that their hypothetical mechanism for blocking might also account for the degradation of learning due to intermittent presentation of US alone. Extending the arguments of Rescorla and Wagner (1972) to the neuronal level, they suggest that if context is viewed as an additional CS, then the presentation of US-alone trials would serve to increase the context→US association and, via the interneuron refractory mechanisms, attenuate the CS association. Thus,

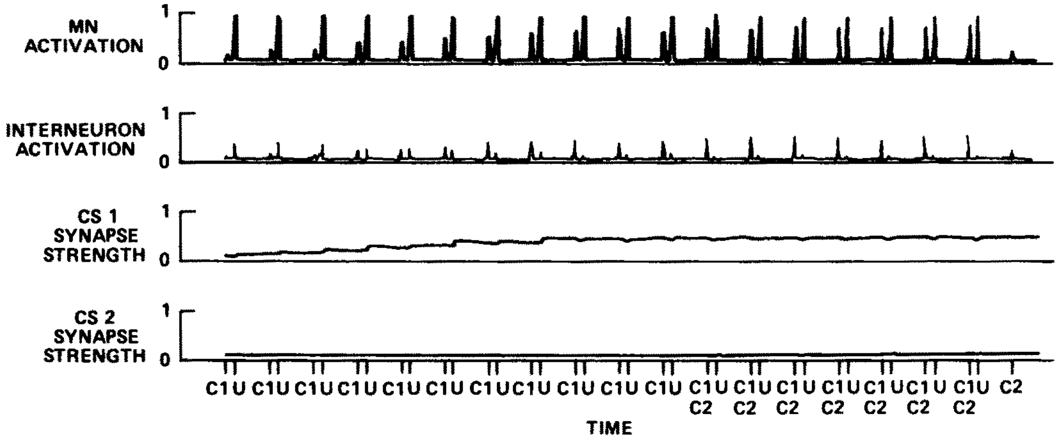

Figure 11. Successful blocking with new S-shaped learning function. (The degree to which a CS synapse can be strengthened is proportional to the strength × (1 − strength). The refractory state is interpreted as 1 minus the probability that the interneuron can fire. Thus, when the refractory state is high, the interneuron cannot fire. As the refractory state decays toward 0, the interneuron is again able to be fired.)

whether or not the circuit model can mediate blocking has important implications for a wider range of behavioral phenomena.

Summary and Conclusions

Our computational model of the *Aplysia* circuit suggests that several of the higher order features of classical conditioning do, as Hawkins and Kandel (1984) suggest, follow as natural elaborations of identified cellular mechanisms for associative and nonassociative learning. In particular we have provided quantitative support for their models of acquisition, extinction, differential conditioning, and second-order conditioning. In doing this, we found that we did not need to concern ourselves with the biophysical properties of neurons (e.g., ionic membrane properties), fine-grained temporal properties and mechanisms of neurotransmitter release, or the kinetics of transmitter–receptor interactions. Rather, the models suggested by Hawkins and Kandel (1984) for differential and second-order conditioning appear to be robust at a cellular level of description, a level comparable to that used by most cognitive-level connectionist models. Quantitative simulations of their models for blocking and contingency learning suggest, however, that it is necessary to assume a particular form of acquisition function (S-shaped) to robustly predict these higher order features of classical conditioning. We have speculated on two possible classes of extensions to the model, consistent with the known *Aplysia* physiology, that could in principle generate blocking behavior. The critical functional feature of these models is that they provide a mechanism for distinguishing between acquisition and maintenance of learned responses. If a single interneuron does mediate blocking and contingency learning, then we suggest that the rate of synaptic sensitization will be a critical factor. To test the plausibility of this model will necessitate modeling the sensitization process with far greater detail than we have attempted here. Recent computational models of the sub-

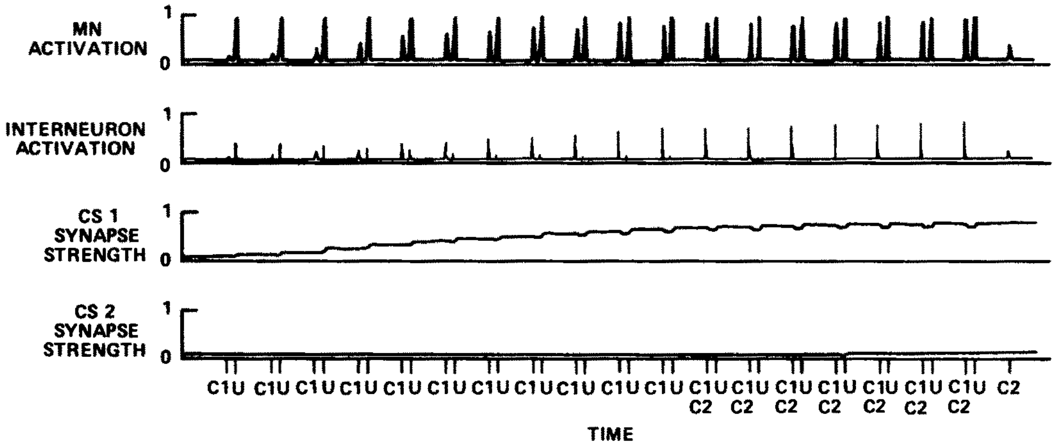

Figure 12. Successful and strong blocking with additional interneuron mediating retention added to previous circuit model.

cellular mollecular processes mediating associative and nonassociative sensitization in *Aplysia* (Gingrich & Byrne, 1985) may provide the necessary constraints to determine if in fact a single-interneuron model for classical conditioning in *Aplysia* is tenable. If other interneurons are indicated as being critical to contingency learning in *Aplysia,* then we speculate that there may exist interneurons whose function is to maintain learned associations but that have little or no effect on the acquisition of new associations.

Similarities to Mammalian Circuitry

The *Aplysia* circuit has certain similarities to the more hypothetical neuronal circuit defined in the mammalian spinal cord that subserves habituation and sensitization of spinal flexion reflexes (Groves & Thompson, 1970; Spencer, Thompson, & Neilson, 1966; Thompson & Spencer, 1966). In both, there is a direct circuit from CS afferents to motor neurons, monosynaptic in *Aplysia* and polysynaptic in the spinal cord, which exhibits habituation with repetitive activation. Strong stimulation of afferents induces sensitization of the response to CS via interneuron actions. In both systems, habituation appears due to a process of synaptic depression, and sensitization is a separate and independent superimposed increase in excitability. In *Aplysia,* sensitization appears due to a process of presynaptic facilitation (Castellucci & Kandel, 1976; Castellucci, Pinsker, Kupfermann, & Kandel, 1970; Kandel, 1976). In the spinal cord, it is not known whether presynaptic facilitation is involved in sensitization; postsynaptic increases in motor-neuron excitability do typically accompany sensitization (Spencer et al., 1966; Thompson & Spencer, 1966).

Both circuits are capable of elementary associative learning (Beggs, Steinmetz, Romano, & Patterson, 1983; Carew et al., 1983; Carew et al., 1981; Durkovic, 1975; Fitzgerald & Thompson, 1967; Patterson, 1975; Patterson, 1976; Patterson, Cegavske, & Thompson, 1973; Patterson, Steinmetz, Beggs, & Romano, 1982). In *Aplysia,* this appears to be a result of a persisting and pairing-specific sensitization-like presynaptic process (Carew et al., 1983; Hawkins & Kandel, 1984; Kandel, Abrams, Bernier, Carew, Hawkins, & Schwartz, 1983); in the spinal cord, the associative process is not yet understood at the synaptic level.

There is also a close correspondence between the properties of classical conditioning in the spinal mammal and those of the classically conditioned gill-withdrawal reflex in *Aplysia:* In both, the initial small response to the CS increases in amplitude as a result of pairing. They are true instances of associative learning in that the increase with pairing is significantly greater than any increase that may occur with unpaired control stimulation. The effect of the interstimulus onset interval on the degree of learning is evident in both preparations and essentially identical to that in classical conditioning of skeletal muscle responses in intact mammals (Gormezano, 1972). No learning occurs with backward pairings (US onset preceding CS onset) or with simultaneous CS–US onset, and the best learning occurs with a CS–US onset interval of about ¼–½ s (Hawkins et al., 1983; Patterson, 1975, 1980). Because *Aplysia* and spinal conditioning exhibit such strikingly parallel phenomena, it is at least possible that the underlying mechanisms of plasticity may be similar, perhaps the most basic or elementary form of associative learning. In any event, as noted above, the circuits and the associative learning they exhibit have many similar properties.

Levels of Analysis in Modeling Learning and Memory

In understanding a complex information-processing system, Marr (1982) described three distinct but interrelated levels of explanation: the level of the computation performed, the level of the algorithm for this computation, and the level of the physical mechanisms that implement this algorithm. In the domain of the neurobiology of learning and memory, the work of Kamin (1969), Rescorla (1968), and Wagner (1969) provided an important constraint on what is being computed in classical conditioning by demonstrating that it is contingency, and not merely contiguity, that determines the association strengths which develop between a CS and a US (see also Granger & Schlimmer, 1986). Various algorithms have since been proposed (e.g., Donegan & Wagner, in press; Rescorla & Wagner, 1972; Wagner, 1981) to describe the iterative trial-by-trial changes in association strengths by which animals learn to respond according to these contingencies. In attempting to identify neuronal correlates of the Rescorla–Wagner learning model, Hawkins and Kandel (1984) have taken a formidable step in attempting to bridge the gap between algorithmic-level models of classical conditioning and implementation-level models of the underlying neurophysiology. The particular advantage of formulating these models within a similar computational framework is that it allows researchers to test more precisely, at a quantitative level, whether the models are both computing the same target behavior.

Our analyses illustrate the complexities that arise in trying to understand a circuit involving only four neurons that generates phenomena of associative learning. If the functioning of this simple circuit is not evident at a qualitative level, then the more complex circuits that code, store, and retrieve memories in the mammalian brain will certainly require quantitative modeling.

References

Alkon, D. L. (1980). Membrane depolarization accumulates during acquisition of an associative behavioral change. *Science, 210,* 1375–1376.

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., & Hinton, G. E. (1981). Models of information processing in the brain. In G. E. Hinton & J. A. Anderson (Eds.), *Parallel models of associative memory* (pp. 9–48). Hillsdale, NJ: Erlbaum.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Beggs, A. L., Steinmetz, J. E., Romano, A. G., & Patterson, M. M. (1983). Extinction and retention of a classically conditioned flexor nerve response in acute spinal cat. *Behavioral Neuroscience, 97,* 530–540.

Bernier, L., Castellucci, V. F., Kandel, E. R., & Schwartz, J. H. (1982). Facilitatory transmitter causes a selective and prolonged increase in adenosine 3:5 monophosphate in sensory neurons mediating the gill and siphon withdrawal reflex in Aplysia. *Journal of Neuroscience, 2,* 1682–1691.

Black, A. H., & Prokasy, W. F. (1972). *Classical conditioning II: Current research and theory.* New York: Appleton-Century-Crofts.

Carew, T. J., Hawkins, R. D., Abrams, T. W., & Kandel, E. R. (1984). A test of Hebb's postulate at identified synapses which mediate classical conditioning in *Aplysia. Journal of Neuroscience, 4,* 1217–1224.

Carew, T. J., Hawkins, R. D., & Kandel, E. R. (1983). Differential classical conditioning of a defensive withdrawal reflex in Aplysia californica. *Science, 219,* 397–400.

Carew, T. J., Pinsker, H. M., & Kandel, E. R. (1972). Longterm habituation of a defensive withdrawal reflex in Aplysia. *Science, 175,* 451–454.

Carew, T. J., Walters, E. T., & Kandel, E. R. (1981). Classical conditioning in a simple withdrawal reflex in Aplysia californica. *Journal of Neuroscience, 1,* 1426–1437.

Castellucci, V. F., & Kandel, E. R. (1974). A quantal analysis of the synaptic depression underlying habituation of the gill-withdrawal reflex in Aplysia. *Proceedings of the National Academy of Sciences, 71,* 5004–5008.

Castellucci, V. F., & Kandel, E. R. (1976). Presynaptic facilitation as a mechanism for behavioral sensitization in Aplysia. *Science, 194,* 1176–1178.

Castellucci, V. F., Nairn, A., Greengard, P., Schwartz, J. H., & Kandel, E. R. (1982). Inhibitor of adenosine 3:5-monophosphate-dependent protein kinase blocks presynaptic facilitation in Aplysia. *Journal of Neuroscience, 2,* 1673–1681.

Castellucci, V. F., Pinsker, H., Kupfermann, I., & Kandel, E. F. (1970). Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in Aplysia. *Science, 167,* 1745–1748.

Chang, J. J., & Gelperin, A. (1980). Rapid taste aversion learning by an isolated molluscan central nervous system. *Proceedings of the National Academy of Sciences, 77,* 6204.

Cohen, D. H. (1980). The functional neuroanatomy of a conditioned response. In R. F. Thompson, L. H. Hicks, & V. B. Shvyrkov (Eds.), *Neural mechanisms of goal-directed behavior and learning* (pp. 283–302). New York: Academic Press.

Colwill, R. W. (1985). Context conditioning in Aplysia Californica. *Society for Neuroscience Abstracts, 11,* 796.

Davis, W. J., & Gillette, R. (1978). Neural correlates of behavioral plasticity in command neurons of Pleurobranchaea. *Science, 199,* 801–804.

Donegan, N. H., & Wagner, A. R. (in press). Conditioned dimunition and facilitation of the UCR: A sometimes-opponent-process interpretation. In I. Gormexano, W. Prokasy, & R. Thompson (Eds.), *Classical conditioning II: Behavioral, neurophysiological, and neurochemical studies in the rabbit.* Hillsdale, NJ: Erlbaum.

Durkovic, R. G. (1975). Classical conditioning, sensitization, and habituation of the flexion reflex of the spinal cat. *Physiology and Behavior, 14,* 297.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science, 6,* 205–254.

Fitzgerald, L. A., & Thompson, R. F. (1967). Classical conditioning of the hindlimb flexion reflex in the acute spinal cat. *Psychonomic Science, 9,* 511–512.

Frey, P. W., & Sears, R. J. (1978). Model of conditioning incorporating the Rescorla–Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review, 85,* 321–340.

Gingrich, K. J., & Byrne, J. H. (1985). Simulation of synaptic depression, posttetanic potentiation, and presynaptic facilitation of synaptic potentials from sensory neurons mediating gill-withdrawal reflex in Aplysia. *Journal of Neuroscience, 53,* 652–669.

Goldman-Rakic, P. (1984). The frontal lobes: Uncharted provinces of the brain. *Trends in Neurosciences, 7,* 425–429.

Gormezano, I. (1972). Investigations of defense and reward conditioning in the rabbit. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 151–181). New York: Appleton-Century-Crofts.

Granger, R. H., & Schlimmer, J. C. (1986). The computation of contingency in classical conditioning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 20, pp. 137–192). New York: Academic Press.

Groves, P. M., & Thompson, R. F. (1970). Habituation: A dual-process theory. *Psychological Review, 77,* 419–450.

Hawkins, R. D. (1981). Interneurons involved in mediation and modulation of gill-withdrawal reflex in Aplysia. III. Identified facilitating neurons increase Ca2+ current in sensory neurons. *Journal of Neurophysiology, 45,* 327–339.

Hawkins, R. D., Abrams, T. W., Carew, T. J., & Kandel, E. R. (1983). A cellular mechanism of classical conditioning in Aplysia: Activity-dependent amplification of presynaptic facilitation. *Science, 219,* 400–404.

Hawkins, R. D., Carew, T. J., & Kandel, E. R. (1983). Effects of interstimulus interval and contingency on classical conditioning in Aplysia. *Society for Neurophysiology Abstracts, 9,* 168.

Hawkins, R. D., Castellucci, V. F., & Kandel, E. R. (1981). Interneurons involved in mediation and modulation of gill-withdrawal reflex in Aplysia. I. Identification and characterization. *Journal of Neurophysiology, 45,* 304–314.

Hawkins, R. D., & Kandel, E. R. (1984). Is there a cell-biological alphabet for simple forms of learning? *Psychological Review, 91,* 376–391.

Hebb, D. (1949). *Organization of behavior.* New York: Wiley.

Hilgard, E. R., & Bower, G. H. (1975). *Theories of learning.* Englewood Cliffs, NJ: Prentice-Hall.

Hinton, G. E., & Anderson, J. A. (1981). *Parallel models of associative memory.* Hillsdale, NJ: Erlbaum.

Hoyle, G. (1980). Learning, using natural reinforcements, in insect preparations that permit cellular neuronal analysis. *Journal of Neurobiology, 11,* 323–354.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Ito, M. (1982). Cerebellar control of the vestibulo-ocular reflex around the flocculus hypothesis. *Annual Review of Neuroscience, 5,* 275–296.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kamin, L. J., & Gaioni, S. J. (1974). Compound conditioned emotional response conditioning with differentially salient elements in rats. *Journal of Comparative Physiological Psychology, 87,* 591–597.

Kandel, E. R. (1976). *Cellular basis of behavior: An introduction to behavioral neurobiology.* San Francisco, CA: Freeman.

Kandel, E. R., Abrams, T., Bernier, L., Carew, T. J., Hawkins, R. D., & Schwartz, J. A. (1983). Classical conditioning and sensitization share aspects of the same molecular cascade in Aplysia. *Cold Spring Harbor Symposium on Quantitative Biology, 48,* 821–830.

Kandel, E. R., & Schwartz, J. H. (1982). Molecular biology of learning: Modulation of transmitter release. *Science, 218,* 433–443.

Kandel, E. R., & Spencer, W. A. (1968). Cellular neurophysiological approaches in the study of learning. *Physiological Reviews, 58,* 65–134.

Kapp, B. S., Gallagher, M., Applegate, C. D., & Frysinger, R. C. (1982). The amygdala central nucleus: Contributions to conditioned cardiovascular responding during aversive Pavlovian conditioning in the rabbit. In C. D. Woody (Ed.), *Conditioning: Representation of involved neural functions* (pp. 581–600). New York: Plenum Press.

Klein, M., Shapiro, E., & Kandel, E. R. (1980). Synaptic plasticity and the modulation of the Ca^{++} current. *Journal of Experimental Biology, 89,* 117–157.

Krasne, F. B. (1969). Excitation and habituation of the crayfish escape reflex: The depolarizating response in lateral giant fibers of the isolated abdomen. *Journal of Experimental Biology, 50,* 29–46.

Lashley, K. S. (1929). *Brain mechanisms and intelligence.* Chicago: University of Chicago Press.

Lashley, K. S. (1950). In search of the engram. *Symposium of the Society for Experimental Biology, 4,* 454–482.

Mackintosh, N. J. (1974). *The psychology of animal learning.* New York: Academic Press.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information.* San Francisco, CA: Freeman.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–328.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88,* 375–407.

Mishkin, M. (1978). Memory in monkeys severely impaired by combined but not separate removal of amygdala and hippocampus. *Nature, 273,* 297–298.

Patterson, M. M. (1975). Effects of forward and backward classical conditioning procedures on a spinal cat hindlimb flexor nerve response. *Physiological Psychology, 3,* 86–91.

Patterson, M. M. (1976). Mechanisms of classical conditioning and fixation in spinal mammals. In A. H. Riesen & R. F. Thompson (Eds.), *Advances in psychobiology* (pp. 381–436). New York: Wiley.

Patterson, M. M., Cegavske, C. F., & Thompson, R. F. (1973). Effects of a classical conditioning paradigm on hindlimb flexor nerve response in immobilized spinal cat. *Journal of Comparative & Physiological Psychology, 84,* 88–97.

Patterson, M. M. (1980). Mechanisms of classical conditioning of spinal reflexes. In R. F. Thompson, L. H. Hicks, & V. B. Shvyrkov (Eds.), *Neural mechanisms of goal-directed behavior and learning* (pp. 263–272). New York: Academic Press.

Patterson, M. M., Steinmetz, J. E., Beggs, A. L., & Romano, A. G. (1982). Associative processes in spinal reflexes. In C. D. Woody (Ed.), *Conditioning: Representation of involved neural functions* (pp. 637–650). New York: Plenum.

Pavlov, I. (1927). *Conditioned reflexes.* London: Oxford University Press.

Pinsker, H. M., Kupfermann, I., Castellucci, V., & Kandel, E. R. (1970). Habituation and dishabituation of the gill-withdrawal reflex in Aplysia. *Science, 167,* 1740–1742.

Prokasy, W. F. (1965). Classical eyelid conditioning: Experimental operations, task demands, and response shaping. In W. F. Prokasy (Ed.), *Classical conditioning* (pp. 208–225). New York: Appleton-Century-Crofts.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of CS in fear conditioning. *Journal of Comparative and Physiological Psychology, 66,* 1–5.

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

Rumelhart, D. E., & McClelland, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition, Vol. 1: Foundations.* Cambridge, MA: Bradford Books/MIT Press.

Sahley, C. L., Rudy, J. W., & Gelperin, A. (1981). An analysis of associative learning in the terrestrial mollusk. I: Higher-order conditioning, blocking, and a US-preexposure effect. *Journal of Comparative Physiology, 144,* 1–8.

Schneiderman, N., McCabe, P. M., Haselton, J. R., & Ellenberger, H. H. (in press). Neurobiological bases of conditioned bradycardia. In I. Gormezzano, W. F. Prokasy, & R. F. Thompson (Eds.), *Classical conditioning III: Behavioral, neurophysiological, and neurochemical studies in the rabbit.* Hillsdale, NJ: Erlbaum.

Spence, K. W. (1956). *Behavior theory and conditioning.* New Haven, CT: Yale University Press.

Spencer, W. A., Thompson, R. F., & Neilson, D. R., Jr. (1966). Decrement of ventral root electronic and intracellularly recorded PSPs produced by iterated cutaneous afferent volleys. *Journal of Neurophysiology, 29,* 253–274.

Squire, L. R. (1982). The neurophysiology of human memory. *Annual Review of Neuroscience, 5,* 241–273.

Squire, L. R., & Zola-Morgan, S. (1983). The neurology of memory: The case for correspondence between the findings for man and nonhuman primate. In J. A. Deutsch (Ed.), *The physiological basis of memory* (2nd ed., pp. 199–268). New York: Academic Press.

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

Thompson, R. F. (1986). The neurobiology of learning and memory. *Science, 233,* 941–947.

Thompson, R. F., Berger, T. W., Cegavske, C. F., Patterson, M. M., Roemer, R. A., Teyler, T. J., & Young, R. A. (1976). The search for the engram. *American Psychologist, 31,* 209–227.

Thompson, R. F., Berger, T. W., & Madden, J. (1983). Cellular processes of learning and memory in the mammalian CNS. *Annual Review of Neuroscience, 6,* 447–491.

Thompson, R. F., Clark, G. A., Donegan, N. H., Lavond, D. G., Madden, J. IV, Mamounas, L. A., Mauk, M. D., & McCormick, D. A. (1984). Neuronal substrates of basic associative learning. In L. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 424–442). New York: Guilford Press.

Thompson, R. F., Donegan, N. H., & Lavond, D. G. (in press). *The psychobiology of learning and memory.* New York: Wiley.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review, 173,* 16–43.

Tsukahara, N. (1981). Synaptic plasticity in the mammalian central nervous system. *Annual Review of Neuroscience, 4,* 351–379.

Wagner, A. R. (1969). Stimulus selection and a modified continuity theory. In G. Bower & J. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3, pp. 1–41). New York: Academic Press.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. Spear & G. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5–47). Hillsdale, NJ: Erlbaum.

Walters, E. T., & Byrne, J. H. (1983). Associative conditioning of single sensory neurons suggests a cellular mechanism for learning. *Science, 219,* 404–407.

Woody, C. D. (1982). *Conditioning: Representation of involved neural function.* New York: Plenum Press.

Received June 17, 1985
Revision received April 2, 1986 ■