Psychological Review
1992, Vol. 99, No. 2, 268-305

0033-295X/92/$3.00

Stimulus Configuration, Classical Conditioning, and Hippocampal Function

Nestor A. Schmajuk and James J. DiCarlo
Northwestern University

Hippocampal participation in classical conditioning is described in terms of a multilayer network that portrays stimulus configuration. The network (a) describes behavior in real time, (b) incorporates a layer of "hidden" units positioned between input and output units, (c) includes inputs that are connected to the output directly as well as indirectly through the hidden-unit layer, and (d) uses a biologically plausible backpropagation procedure to train the hidden-unit layer. Nodes and connections in the neural network are mapped onto regional cerebellar, cortical, and hippocampal circuits, and the effect of lesions of different brain regions is formally studied. Computer simulations of the following classical conditioning paradigms are presented: acquisition of delay and trace conditioning, extinction, acquisition–extinction series of delay conditioning, blocking, overshadowing, discrimination acquisition, discrimination reversal, feature-positive discrimination, conditioned inhibition, negative patterning, positive patterning, and generalization. The model correctly describes the effect of hippocampal and cortical lesions in many of these paradigms, as well as neural activity in hippocampus and medial septum during classical conditioning. Some of these results might be extended to the description of anterograde amnesia in human patients.

In spite of the vast amount of behavioral and physiological data that have been accumulated regarding hippocampal participation in associative learning, the captivating question of hippocampal function remains largely unanswered. In attempting to address this question, numerous theories have been advanced that suggest that the hippocampus is involved in attention, chunking, contextual retrieval of information, internal inhibition, long-term memory selection, recognition memory, response inhibition, spatial mapping, temporal mapping, working memory, or configural memory (see Schmajuk, 1984b, for a review). Typically, theories are tested by contrasting the predicted results of different hippocampal manipulations (e.g., lesions, induction of long-term potentiation, kindling, or neural recording) with empirical results. However, testing is frequently hindered by the fact that most theories do not provide unequivocal predictions of the effects of hippocampal manipulations on learning. This serious deficiency results from the lack of rigorous specifications of the interaction between the alleged hippocampal function and associative learning.

In contrast to most nonrigorous, verbal theories available at the time, Moore and Stickney (1980, 1982) described hippocampal function in the context of a computational model of classical conditioning. They theorized that the hippocampus computes attentional variables as defined in a real-time version of Mackintosh's (1975) model. With adequate changes in the computation of these attentional variables, the model was able to precisely describe the consequences of hippocampal lesions. These precise descriptions were then contrasted with experimental data to test the validity of the theory. In short, the use of a real-time computational model allowed the generation of an explicitly testable theory of hippocampal function.

Also adopting a computational approach, Schmajuk and Moore (1985, 1988, 1989) and Schmajuk and DiCarlo (1991a, 1991b) studied and compared different theories of hippocampal function in classical conditioning. Because these computational models are real-time models, their output can be compared to behavior as it unfolds in real time, and the dynamics of their intervening variables can be contrasted with neural activity. In the context of these computational models, hippocampal theorizing consists of hypothesizing what model variables map onto the hippocampus. Most important, with relevant changes in their computations the hippocampally mapped, real-time models are able to describe the consequences of different hippocampal manipulations on behavior.

Among the aforementioned computational models, Schmajuk and Moore (1985) described a real-time attentional model of hippocampal function based on Pearce and Hall's (P-H; 1980) model of classical conditioning. According to the P-H model, when a conditioned stimulus (CS) is followed by an unconditioned stimulus (US) a CS–US association is formed. This association can be regarded as a prediction of the US by the CS. CS–US associations are changed until the conditioning rate of the CS (CS associability) is zero. On a given trial, CS associability is defined by the absolute difference between the US intensity on the previous trial and the "aggregate prediction" of the US computed upon all CSs present on the previous trial. Schmajuk (1984a) theorized that the hippocampus computes the "aggregate prediction" of environmental events used to determine CS associability. In a real-time model, aggregate predictions triggered by the CSs present at a given time forecast *what* event is going to occur, *when* in time, and *where* in space.

This project was supported in part by BRSG SO7 RR07028-22 awarded by the Biomedical Research Support Grant Program, Division of Research Resources, National Institutes of Health, and by Grant N00014-91-J-1764 from the Office of Naval Research.

We are grateful to Nelson Donegan, John W. Moore, William Revelle, Aryeh Routtenberg, Aaron Thieme, and Philip Wolff for their comments on previous versions of the manuscript.

Correspondence concerning this article should be addressed to Nestor A. Schmajuk, Department of Psychology, Northwestern University, Evanston, Illinois 60208.

According to the aggregate prediction hypothesis, the activity of some hippocampal pyramidal neurons is proportional to the instantaneous value of the aggregate prediction, and the computation of the aggregate prediction is impaired by hippocampal lesions.

Schmajuk (1986, 1989; Schmajuk & Moore, 1988) extended the Schmajuk and Moore (1985) real-time version of the P-H model to include CS–CS associations and designated this rendering the S-P-H (Schmajuk-Pearce-Hall) model. By combining CS–CS and CS–US associations, the S-P-H model describes second-order conditioning and sensory preconditioning. Under the aggregate prediction hypothesis, the S-P-H model correctly describes the effect of hippocampal lesions on acquisition of delay conditioning, extinction, latent inhibition, generalization, blocking, overshadowing, discrimination reversal, and sensory preconditioning. However, the S-P-H model has difficulty describing the effect of hippocampal lesions on conditioned inhibition and mutual overshadowing. In addition to its applications to classical conditioning, Schmajuk (1990) showed that under the aggregate prediction hypothesis the S-P-H model correctly describes the effects of hippocampal lesions in spatial learning: Place learning, but not cue learning, is impaired by hippocampal lesions.

Schmajuk and Moore (1988; Schmajuk, 1986) assumed that CS–CS associations were stored in the form of long-term potentiation (LTP) of hippocampal synapses and that, in consequence, the induction of hippocampal LTP increased the value of CS–CS associations. Unfortunately, under this hypothesis the S-P-H model has difficulty describing Berger's (1984) results showing that LTP induction facilitates discrimination acquisition.[1] Consequently, Schmajuk (1989) later assumed that CS–CS associations are stored in association areas of the neocortex.

Although the S-P-H model can be applied to a wide variety of complex classical conditioning paradigms, including sensory preconditioning and cognitive mapping (Schmajuk, 1987), it cannot address the problems of negative and positive patterning. In negative patterning, two CSs are introduced separately in the presence of the US and presented together in the absence of the US. After training, combined presentation of both CSs generates a response smaller than the sum of the responses to each CS. In positive patterning, two CSs are introduced separately in the absence of the US and together in the presence of the US. After training, combined presentation of both CSs generates a response larger than the sum of the responses to each CS. In order to extend the application of the S-P-H model to positive and negative patterning, we introduce an expanded version of the model referred to as the S-D (Schmajuk-DiCarlo) model. For simplicity, the present rendering of the S-D model does not include CS–CS associations.

After completely specifying the S-D model, we mapped nodes and connections in the network onto regional cerebellar, cortical, and hippocampal circuits. With the network mapped onto the brain, we used computer simulations to formally study neural activity in different brain regions and the effect of lesions of these areas. The present study compares empirical and simulated results regarding (a) the effect of hippocampal and cortical lesions in numerous classical conditioning paradigms and (b) neural activity in hippocampus and medial septum during classical conditioning. Although most of the data addressed in this study refer to the eyeblink or nictitating membrane (NM) response preparation, we also apply the model to studies that involve other classical conditioning techniques.

Experimental Data

In this section we review neurophysiological, anatomical, and behavioral data that constrain both the formulation of the model and its mapping onto the brain circuit. Those readers not concerned with neurophysiological details are referred to the *Summary of Neurophysiological Data* presented later in this section.

Hippocampal Neuronal Activity During Classical Conditioning

In the rabbit NM preparation, hippocampal activity during acquisition of classical conditioning is positively correlated with the topography of the conditioned response (CR), and increments in neural activity precede behavioral acquisition (Berger & Thompson, 1978a). Berger, Rinaldi, Weisz, and Thompson (1983) found that CA1 and CA3 (CA = cornus ammonis) pyramidal cells increase their frequency of firing over acquisition trials and display a within-trial pattern of activity that models the NM response. Berger, Clark, and Thompson (1980) reported that neural activity correlated with the CR is present also in the entorhinal cortex and is amplified over trials in CA1 and CA3 hippocampal regions. During extinction following delay conditioning, Berger and Thompson (1982) found that pyramidal cells in dorsal hippocampus decreased their frequency of firing correlated with behavioral extinction during the period when the US was presented in the course of acquisition, but in advance of behavioral extinction during the CS period.

In addition to CR-related neural activity, other types of activity have also been recorded from hippocampal regions. For instance, Vinogradova (1975) found that neural activity in CA3 and CA1 pyramidal neurons, dentate gyrus, and entorhinal cortex of the rabbit was correlated with the presentation of sensory stimuli (tones and light). Activity in CA3 and CA1 showed habituation after repeated presentation of the stimuli. Berger et al. (1983) found that hippocampal neurons coding CS information do not respond to the US and are also segregated from those coding CR information. Wible, Findling, Shapiro, Lang, Crane, and Olton (1986) observed that, in a variety of learning tasks, hippocampal cells respond not only to individual stimuli but to combinations of stimulus dimensions such as color, shape, and spatial location.

Berger et al. (1983) also reported that theta cells, presumably basket cells that provide recurrent inhibition to pyramidal neurons, respond during paired conditioning trials with a rhythmic 8-Hz bursting pattern. Weisz, Clark, and Thompson (1984) found that granule cells in the dentate gyrus exhibited a CS-

[1] As demonstrated by Schmajuk and DiCarlo (1991a, 1991b), this experimental result is well characterized by assuming that, rather than CS–CS associations, information related to the CS relevance for the learned task is stored in the form of hippocampal LTP.

evoked theta firing when rabbits were trained with a CS followed by a US, but not when they were trained with CS and US unpaired presentations.

In summary, pyramidal neural activity in CA3 and CA1 is correlated either with the behavioral CR or with the presentation of simple and compound CSs. Granule cell activity in the dentate gyrus exhibits theta firing evoked by CS presentation.

Lateral Septal Neural Activity During Classical Conditioning

As mentioned, Berger and Thompson (1978a) and Berger et al. (1983) found that the activity of CA1 and CA3 pyramidal cells is positively correlated with the topography of the CR. Because one important hippocampal output is mediated through CA3 axons that reach the lateral septum, it is not surprising that Berger and Thompson (1978b) and Salvatierra and Berry (1989) reported that neural activity in the lateral septum during acquisition of classical conditioning in the rabbit NM preparation is similar to that recorded in hippocampal pyramidal cells. These results suggest an excitatory CA3–lateral-septum pathway. However, in contrast to this view, Vinogradova, Brazhnik, Karanov, and Zhadina (1980) reported that interruption of hippocampal connections to the lateral septum at the level of the septo-fimbrial nucleus increased spontaneous activity in the lateral septum.

In summary, hippocampal activity reaches the lateral septum through CA3 axons, and neural activity in the lateral septum during acquisition of classical conditioning reflects the activity of hippocampal pyramidal cells.

Cerebellar Involvement in Classical Conditioning

Experimental evidence using the rabbit NM preparation and rat eyeblink conditioning suggests that cerebellar areas are essential for the acquisition and maintenance of classical conditioning (McCormick, Clark, Lavond, & Thompson, 1982; Skelton, 1988). The association of a CS and the US might be mediated by plastic changes at the interpositus nucleus of the cerebellum or at the Purkinje cells of the hemispheric portion of cerebellar lobule VI or at both. Sensory representations of the CS may reach the interpositus nucleus and the cerebellar cortex via mossy fibers from the pontine nuclei, and US representations may reach the interpositus nucleus and the cerebellar cortex via climbing fibers from the dorsal accessory olive (McCormick, Steinmetz, & Thompson, 1985). CR-related activity originates in the cerebellar lobule VI or the interpositus nuclei or both, is relayed to the contralateral red nucleus, and reaches the contralateral accessory abducens nuclei where the NM response is controlled.

Andersson and Armstrong (1987) reported that the inferior olive is selectively active during unexpected somatic events. Consistent with this view, olivary responses to periorbital stimuli are absent after the acquisition of classical conditioning (Berthier & Moore, 1986; Foy & Thompson, 1986). This decreased activity may result from an inhibitory action of the red nucleus on the dorsal accessory olive (C. Weiss, Houk, & Gibson, 1990), mediated by inhibitory projections from the red nucleus to the trigeminal spinal nuclei (Davis & Dostrovsky, 1986), and inhibitory projections from the interpositus nucleus to the dorsal accessory olive (Andersson, Gorwicz, & Hesslow, 1987).

Lesions of the dentate and interpositus cerebellar nuclei ipsilateral to the trained eye cause abolition of both the behavioral CR and CR-related neural activity in CA1 (Clark, McCormick, Lavond, & Thompson, 1984; Sears & Steinmetz, 1990). Information about the behavioral response might be conveyed to the hippocampus via cerebellar-thalamic-cortical pathways (see Ito, 1984).

To summarize, cerebellar areas seem critical for the acquisition and maintenance of eyeblink and NM conditioning. Whereas the pontine nuclei seem to convey CS information, the dorsal accessory olive seems to carry US information to the cerebellar loci of learning. CR-related activity originates in the cerebellar lobule VI or the interpositus nuclei or both, is relayed to the contralateral red nucleus, and reaches the contralateral accessory abducens nuclei where the NM response is controlled. Lesions of cerebellar nuclei cause abolition of both the behavioral CR and the CR-related neural activity in pyramidal cells.

Hippocampal–Cerebellar Interactions: Modulation of Classical Conditioning

Berger, Swanson, Milner, Lynch, and Thompson (1980) proposed that hippocampal projections to the pontine nucleus modulate the activity of mossy fiber projections to cerebellar cortex and interpositus nucleus (Berger, Weikart, Bassett, & Orr, 1986; Steinmetz, Logan, & Thompson, 1988). Two hippocampal–pontine pathways have been described: a hippocampal–retrosplenial-cortex projection via the subiculum that reaches the ventral pons (Berger, Bassett, & Weikart, 1985; Berger et al., 1986; Semple-Rowland, Bassett, & Berger, 1981) and a cingulo–pontine projection (Weisendanger & Weisendanger, 1982; Wyss & Sripanidkulchai, 1984).

In addition to its modulation of the pontine nucleus, we suggest that the hippocampus also acts on the dorsal accessory olive, thereby modulating the activity of climbing fiber projections to cerebellar cortex and interpositus nucleus. The hippocampus might act on the dorsal accessory olive through a lateral-septum–raphe-nucleus–dorsal-accessory-olive pathway. This suggestion is supported by Vinogradova's (1975) data showing that CA3 stimulation has an excitatory influence on the raphe nuclei and by M. Weiss and Pellet's (1982a, 1982b) results demonstrating that the raphe inhibits the inferior olive. More specifically, raphe fibers reaching the inferior olive innervate the dorsal accessory nucleus (Wilklund, Björklund, & Sjölund, 1977), the area proposed to convey US information to cerebellar areas where CS–US associations are stored (McCormick et al., 1985).

To recapitulate, hippocampal output might modulate postulated CS (pontine nucleus) and US (dorsal accessory olive) inputs to cerebellar areas.

Hippocampal–Reticular-Formation Interactions: Control of Theta Rhythm

Vinogradova (1975, p. 39) reported that stimulation of reticular formation evokes theta rhythm in the medial septum,

whereas hippocampal stimulation inhibits activity in the medial septum and leads to a decrease in theta rhythm. Vinogradova suggested that the hippocampus excites the raphe nuclei, which in turn inhibits theta. Consistent with this view, Vertes (1982) suggested that the medial raphe nucleus is responsible for inhibiting and the pontis oralis is responsible for exciting hippocampal theta. Inhibitory and excitatory inputs from raphe and pontis oralis control the generation of theta rhythm at the medial septum.

Berger and Thompson (1978b) found that neuronal activity in the medial septum during acquisition of classically conditioned NM response decreased over trials during both the CS and US periods. Berger and Thompson (1978b) suggested that the medial septum might provide the hippocampus with an "arousal" signal that decreases across trials. In agreement with Vinogradova's (1975) view that the hippocampus inhibits the medial septum and thereby decreases hippocampal theta, Berger and Thompson (1978b) noted that medial septal and hippocampal pyramidal activity are negatively correlated. Also compatible with Vinogradova's view is the negative correlation between hippocampal theta-cell activity and pyramidal cell activity reported in an odor-discrimination paradigm in rats (Eichenbaum, Kuperstein, Fagan, & Nagode, 1987).

In summary, hippocampal output excites the raphe, which in turn inhibits the medial septum and thereby decreases hippocampal theta rhythm. Consistent with this, during acquisition of classical conditioning, as hippocampal activity increases, medial septum neuronal activity decreases.

Medial Septal Modulation of Hippocampal Activity

Axons of the medial septum project to the dentate gyrus, CA3, and to a lesser extent the CA1 region, via the dorsal fornix and fimbria. Medial septal input has a major role in the generation of hippocampal theta rhythm and in controlling the responsiveness of pyramidal cells in CA1 and CA3 regions and of granule cells in the dentate gyrus.

Permanent as well as temporary inactivation of the medial septal input to the hippocampus results in decreased activity of pyramidal cells. Vinogradova (1975) reported that lesions of septal–hippocampal pathways decreased the number of neurons responsive to sensory input cells in CA3, but not in CA1, particularly those activated by multimodal stimuli. Recently, Mizumori, McNaughton, Barnes, and Fox, (1989) examined the effects of reversibly inactivating the medial septum (with lidocaine) on spontaneous hippocampal single-unit activity. They reported that septal inactivation reduced spontaneous firing rates in granule cells, in hilar/CA3 complex spike cells, and in CA1 theta cells, but not in CA1 complex spike cells.

In agreement with the facilitatory view of medial septal function suggested by the inactivation studies, Krnjevic, Ropert, and Casullo (1988) determined that medial septum stimulation can strongly depress tonic inhibition and inhibitory postsynaptic potentials evoked in CA1 and CA2/3 pyramidal cells by fimbrial stimulation. Bilkey and Goddard (1985) reported that stimulation of the medial septum, although unable to elicit a field potential of its own, facilitated the granule cell population spike evoked by medial perforant path stimulation.

In addition to inactivation and stimulation studies, data obtained from freely moving rats show that the excitability of CA1 pyramidal cells and dentate gyrus granule cells reaches a maximum during the positive phase of theta rhythm (Rudell, Fox, & Ranck, 1980).

It has been suggested that medial septal "modulation" of dentate granule cells, CA1, and CA3 pyramidal cells is mediated through either (a) a GABAergic (GABA = gamma-aminobutyric acid) inhibition of inhibitory interneurons or (b) a cholinergic excitation of pyramidal and granule cells (Bilkey & Goddard, 1985; Krnjevic et al., 1988; Rudell et al., 1980; Stewart & Fox, 1990).

Besides modulating pyramidal and granule cell activity, the medial septum controls the generation of LTP in perforant-path–dentate-gyrus synapses. Robinson and Racine (1986) demonstrated associative LTP involving septal and entorhinal inputs to the dentate gyrus in the rat. Concurrent medial septal and perforant path activation resulted in stronger LTP of the perforant-path–dentate-gyrus synapses than could be produced by perforant path stimulation alone. Similarly, Robinson (1986) reported that LTP induced in the rat dentate gyrus is enhanced by coactivation of septal and entorhinal inputs. Functionally important is the observation that these perforant-path–dentate-gyrus synapses are endogenously potentiated during classical conditioning of the rabbit NM response (Weisz et al., 1984).

In summary, the medial septum modulates hippocampal CA1 and CA3 pyramidal cells and dentate granule cells. In addition, medial septal activity enhances LTP of perforant-path–dentate-gyrus synapses.

Hippocampal–Cortical Interactions

Hippocampal–cortical interactions have been recently reviewed by Squire, Shimamura, and Amaral (1989). A prominent hippocampal input is the entorhinal cortex. The entorhinal cortex receives inputs from polysensory associational cortices. A major hippocampal output comprises CA1 axons that extend to the subiculum, which in turn excites cells in the entorhinal cortex. Cells in the entorhinal cortex that do not project to the dentate gyrus project to many of the associational cortical fields from which the entorhinal cortex receives input. To summarize, the hippocampus receives information from and sends information back to different polysensory associational cortices through the entorhinal cortex.

Cortical–Cerebellar Interactions

Parietal association cortex and frontal motor cortex project to the cerebellum. Stimulation of the parietal cortex produces responses in mossy fiber that originate in the pontine nuclei, whereas stimulation of the motor cortex produces responses in motor areas of the cerebellar cortex. In turn, cerebellar lateral and interpositus nuclei project to parietal association cortex and motor cortex (Ito, 1984, p. 314). In short, whereas the pontine nuclei receive information from neocortical areas, interpositus nuclei project back to association and motor cortices.

Summary of Neurophysiological Data

The neurophysiological data presented may be summarized as follows:

1. Hippocampal pyramidal activity in CA3 and CA1 has been correlated with the behavioral CR, with simple CSs, and with compound CSs.
2. Hippocampal CA1 and CA3 activity increases during acquisition and decreases during extinction of classical conditioning.
3. Granule cell activity in the dentate gyrus exhibits theta firing evoked by CS presentation.
4. Neural activity in the lateral septum reflects the CR-related neural activity of hippocampal pyramidal cells.
5. Cerebellar areas seem critical for the acquisition and maintenance of classical conditioning. Whereas the pontine nuclei seem to convey CS information, the dorsal accessory olive seems to carry US information to cerebellar loci of learning.
6. Lesions of dentate and interpositus cerebellar nuclei cause abolition of both the behavioral CR and the CR-related neural activity in hippocampal pyramidal cells.
7. Hippocampal output might modulate postulated CS (pontine nucleus) and US (dorsal accessory olive) inputs to cerebellar areas.
8. Hippocampal stimulation excites raphe nuclei, which in turn inhibit the medial septum, thereby decreasing hippocampal theta rhythm.
9. Medial septum activation has a facilitatory effect on CA1 pyramidal activity, presumably by inhibiting inhibitory interneurons.
10. Medial septal activation enhances LTP of the perforant-path–dentate-gyrus synapses.
11. Medial septum neuronal activity decreases during acquisition of classical conditioning.
12. The hippocampus receives cortical information through the entorhinal cortex and sends back information to many neocortical association areas through entorhinal pathways.
13. Pontine nuclei receive information from cortical associational areas, and interpositus nuclei project back to association and motor cortices.

Effects of Hippocampal Manipulations on Classical Conditioning

In this subsection we review the effects of hippocampal lesions (HL) and induction of hippocampal LTP in different classical conditioning paradigms. In general, HL refers to the extensive bilateral removal of the hippocampus. Although most results refer to classical conditioning obtained in the rabbit eyeblink or NM response preparation, some results refer to classical conditioning in the rat.

Acquisition. The effects of HL on acquisition of classical conditioning have been studied under a wide variety of experimental parameters, including different combinations of CS and US durations, interstimulus intervals (ISIs), and types of US.

Acquisition of delay conditioning of the rabbit NM response has been reported to be unaffected or facilitated by HL. Schmaltz and Theios (1972) found that HL rabbits showed faster acquisition using a 250-ms CS, a 50-ms shock US, and a 250-ms ISI. Solomon and Moore (1975) found that HL rabbits displayed normal acquisition using a 450-ms CS, a 50-ms shock US, and a 450-ms ISI. Berger and Orr (1983) found normal acquisition using an 850-ms CS, a 100-ms airpuff US, and a 750-ms ISI. Port and Patterson (1984) found that HL rabbits showed shorter CR onset latency and normal acquisition rate using a 500-ms CS, a 50-ms shock US, and a 450-ms ISI. Port, Mikhail, and Patterson (1985) found shorter CR onset latency and faster acquisition using an 800-ms CS, a 50-ms shock US, and a 150-ms ISI. When the ISI was extended to 300 ms, they found that HL rabbits showed normal CR onset latency and normal acquisition rate. Finally, with a 600-ms ISI, they found that HL rabbits showed normal CR onset latency and acquisition.

Acquisition of trace conditioning of the rabbit NM response has been reported to be impaired, unaffected, or facilitated by HL. Port, Romano, Steinmetz, Mikhail, and Patterson (1986) reported that HL rabbits showed longer CR onset latency and normal acquisition rate with a 250-ms CS, a 50-ms shock US, and a 750-ms ISI. When the shock US was replaced by a 100-ms airpuff US, they found that HL rabbits showed shorter CR onset latency and normal acquisition rate. Solomon, Vander Schaaf, Thompson, and Weisz (1986) reported shorter CR onset latencies and slower acquisition in HL rabbits with a 250-ms CS, a 100-ms airpuff US, and a 500-ms ISI. James, Hardiman, and Yeo (1987) reported normal acquisition rate and shorter onset latency using a 250-ms CS, a 50-ms shock US, and a 750-ms ISI. Recently, Moyer, Deyo, and Disterhoft (1990) found normal CR onset latency and a slightly faster acquisition rate using a 100-ms CS, a 150-ms airpuff US, and a 300-ms ISI. However, they found shorter CR onset latency and a deficit in acquisition when the ISI was 500 ms.

Extinction. Three studies describe the effect of HL on extinction of the rabbit NM response. Berger and Orr (1983) found normal extinction using an 850-ms CS, a 100-ms airpuff US, and a 750-ms ISI in an explicitly unpaired procedure with trials in which the CS was presented alone alternating with trials in which the US was introduced alone. Moyer et al. (1990) found impaired extinction using a 100-ms CS, a 150-ms airpuff US, and a 300-ms ISI in a simple extinction procedure with CS alone presentations. In an acquisition-extinction series of the NM response in rabbits, Schmaltz and Theios (1972) found that the first extinction (with CS-alone presentations) appeared to be unaffected by HL.

Savings. When acquisition is followed by extinction in an NM classical conditioning paradigm, normal rabbits show faster reacquisition, that is, savings effects (Frey & Ross, 1968; Smith & Gormezano, 1965). Schmaltz and Theios (1972) studied the effect of exposing rabbits to successive acquisition and extinction series using a 250-ms CS, a 50-ms shock US, and a 250-ms ISI. Schmaltz and Theios (1972) found that HL rabbits showed faster acquisition than normals in the first acquisition series. HL rabbits did not differ from normal rabbits in the first extinction series, but in the following extinction series normal rabbits decreased the number of trials to reach criterion (showed savings), whereas HL rabbits increased the number of trials to extinction criterion.

Blocking. In blocking, an animal is first conditioned to CS_1, and this training is followed by conditioning to a compound consisting of CS_1 and a second stimulus, CS_2. This procedure results in a weaker conditioning to CS_2 than it would attain if paired separately with the US. Solomon (1977) found that HL disrupted blocking of the rabbit NM response and Rickert et al. (1978) found impairment of blocking after HL using a conditioned suppression paradigm in rats. In contrast to these findings, Garrud et al. (1984) found that blocking was not affected by HL using a conditioned suppression paradigm in rats.

Overshadowing. In overshadowing, an animal is conditioned to a compound consisting of CS_1 and CS_2. This proce-

dure results in a weaker conditioning to the less salient or less reinforced CS than it would achieve if it was independently trained. If both CSs are of similar salience and equally reinforced, both show weaker conditioning than each would achieve separately. Rickert, Lorden, Dawson, Smyly, and Callahan (1979), using a conditioned suppression paradigm, and Schmajuk, Spear, and Isaacson (1983), using a simultaneous discrimination procedure, found that overshadowing is disrupted in HL rats. In contrast to these findings, Garrud et al. (1984) and Solomon (1977) found that overshadowing of the NM response was not affected by HL.

Discrimination acquisition and reversal. In a discrimination paradigm, reinforced trials with CS_1 are alternated with nonreinforced trials with a second CS_2. During reversal, the original nonreinforced CS_2 is reinforced, whereas the CS_1 reinforced in the first phase is presented without the US.

Buchanan and Powell (1982) examined the effect of HL on acquisition and reversal of eyeblink discrimination in rabbits. They found that HL slightly impaired acquisition of discrimination and severely disrupted its reversal by increasing the responding to CS−. Berger and Orr (1983; Orr & Berger, 1985) contrasted HL and control rabbits in a two-tone differential conditioning and reversal of the NM response. Although HL did not affect initial differential conditioning, HL rabbits were incapable of suppressing responding to CS−. Because HL rabbits showed normal explicitly unpaired extinction, Berger and Orr (1983) suggested that the increased responding to the initially reinforced CS_1 was not simply reflecting a deficiency in extinction of the CS_1–US association.

Similar results were reported by Weikart and Berger (1986) in a tone–light discrimination reversal of the rabbit NM response, suggesting that deficits in the two-tone reversal learning after HL are not due to an increased within-modality generalization to the tone CS serving as CS+ and CS−. Port, Romano, and Patterson (1986) found that HL impaired the reversal learning of a stimulus duration discrimination paradigm using the rabbit NM preparation. In addition, Berger et al. (1986) found that lesions of the retrosplenial cortex, which connects the hippocampus to the cerebellar region, produced deficits in reversal learning of the rabbit NM response.

Berger (1984) found that entorhinal cortex stimulation that produced LTP in the perforant-path–granule-cell synapses increased the rate of acquisition of a two-tone classical discrimination of the rabbit NM response. Robinson, Port, and Berger (1989) showed that kindling of the hippocampal perforant-path–dentate projection (which induces LTP) facilitates discrimination acquisition but impairs discrimination reversal of the rabbit NM response.

Feature-positive discrimination. Ross, Orr, Holland, and Berger (1984) studied the effect of HL in a serial feature-positive discrimination using rats. In serial feature-positive discrimination, animals receive reinforced serial compound presentations (CS_1 followed by CS_2) alternated with nonreinforced presentations of CS_2. Ross et al. (1984) found that when HL preceded training, the lesions prevented the acquisition of the conditional discrimination (responding when CS_2 was preceded by CS_1 but not when CS_2 was presented alone). HL lesions also impaired the retention of the paradigm; HL animals responded both to CS_1 followed by CS_2 and to CS_2 presented alone. Before the surgical procedure, both HL and control groups acquired the conditional discrimination. After the surgical procedure, both control and HL groups increased their responding to CS_1 followed by CS_2, but only HL animals wrongly increased their responding to CS_2 presented alone.

Loechner and Weisz (1987) studied simultaneous feature-positive discrimination in the rabbit NM response preparation. In simultaneous feature-positive discrimination, animals receive reinforced compound presentations alternated with nonreinforced one-component presentations. Loechner and Weisz (1987) did not find a significant difference between HL rabbits and controls when a light was the nonreinforced component, because neither HL rabbits nor controls showed conditioning to the light. However, when a tone was the nonreinforced component, HL rabbits exhibited a high level of responding to both the compound and the component, therefore failing to show simultaneous feature-positive discrimination.[2]

Inhibitory conditioning. Inhibitory conditioning can be obtained through different methods, such as conditioned inhibition and differential conditioning. In conditioned inhibition, animals are presented with CS_1 reinforced trials interspersed with CS_1–CS_2 nonreinforced trials. Solomon (1977) found that HL rabbits yield normal conditioned inhibition of the NM response. It is interesting that although HL did not impair conditioned inhibition, Micco and Schwartz (1972) found that HL impaired inhibitory conditioning to CS_2 obtained in a differential conditioning (discrimination acquisition) paradigm using rats.

Negative patterning. In negative patterning, reinforced component presentations (CS_1 or CS_2) are intermixed with nonreinforced compound (CS_1–CS_2) presentations. Negative patterning is attained if the response to the compound is smaller than the sum of the responses to the components. Bellingham, Gillette-Bellingham, and Kehoe (1985) described the differentiation process between the compound and its components in negative patterning of the NM response in rabbits. At the beginning of training animals responded to both compound and components, and this was followed by a gradual decline of the response to the compound.

So far, the effect of HL on the acquisition and retention of negative patterning has not been studied in classical conditioning. However, because of the relevance of this paradigm for the present article, we refer to Rudy and Sutherland's (1989) study on HL effects on negative patterning with rats in an operant discrimination task. Reference to this work is further justified by the view that discriminative stimuli become classically conditioned to the reinforcer during instrumental conditioning (see Mackintosh, 1983, p. 100).

Rudy and Sutherland (1989) found that when HL preceded training, the lesions prevented the acquisition of negative patterning. When HL followed the acquisition of negative patterning, the lesions impaired the correct performance of negative

[2] According to Holland (1990), whereas in simultaneous feature-positive discrimination CS_1 becomes associated with the US, in serial feature-positive discrimination CS_1 acquires the ability to modulate the action of the association formed between CS_2 and the US. Holland says that CS_1 "sets the occasion" (p. 97) for responding to CS_2.

patterning. Before the surgical procedure, both experimental and control groups acquired negative patterning and responded to the components in about 90% of the trials and to the compound in about 40% of the trials. After the surgical procedure, both control and HL groups dramatically increased their responding to the compound: By the end of the first postlesion session (120 trials), HL and control groups responded to the compound in 90% and 70% of the trials, respectively. After four sessions, controls responded to the compound only in 30% of the trials, whereas HL rats responded to the compound in 90% of the trials.

Positive patterning. In positive patterning, reinforced compound (CS_1–CS_2) presentations are intermixed with nonreinforced component (CS_1 or CS_2) presentations. Positive patterning is attained if the response to the compound is larger than the sum of the responses to the components. Bellingham et al. (1985) described the differentiation process between the compound and its components in positive patterning of the rabbit NM response. At the beginning rabbits responded to both compound and components, and this was followed by a gradual decline of the response to the components.

Generalization. Using the rabbit NM response preparation, Solomon and Moore (1975) reported that HL rabbits showed increased stimulus generalization to tones of different frequencies. Generalization testing consisted of nonreinforced presentations of tones of different frequencies (400, 800, 1200, 1600, and 2000 Hz) following conditioning to a 1200-Hz tone.

Effects of Cortical Lesions on Classical Conditioning

In this subsection we review the effects of cortical lesions (CL) in several classical conditioning paradigms. It is important to note that CL refers to the extensive removal (around 80%) of the neocortex and not simply to the ablation of the region of the neocortex overlying the hippocampus, a procedure sometimes used to control for the effect of hippocampal lesions. All results refer to classical conditioning obtained in the rabbit NM response preparation.

Acquisition. Acquisition of delay conditioning of the rabbit NM response has been reported to be slightly retarded or unaffected by bilateral CL (Oakley & Russell, 1972). Acquisition of trace conditioning of the rabbit NM response has been reported to be unaffected by CL (Yeo, Hardiman, Moore, & Russell, 1984).

Extinction. Extinction of the rabbit NM response has been reported to be normal after CL (Oakley & Russell, 1972).

Blocking. J. W. Moore (personal communication, December 1990) found that CL rabbits showed normal blocking of the NM response.

Discrimination acquisition and reversal. Oakley and Russell (1973, 1975) found that CL does not affect the rate of discrimination acquisition and facilitates the rate of reversal of the rabbit NM response.

Inhibitory conditioning. Conditioned inhibition of the rabbit NM response has not been found to be affected by CL (Moore, Yeo, Oakley, & Russell, 1980).

Effects of Cerebellar Lesions on Classical Conditioning

A considerable amount of data shows that lesions of several cerebellar areas permanently abolish the classically conditioned NM and eyeblink response in the rabbit (see Thompson, 1986, for a review). Although, according to Thompson, cerebellar lesions do not affect the generation of the UR, Welsh and Harvey (1989) reported deficits in both CR and UR. Recently, Skelton (1988) found that bilateral lesions of the dentate-interpositus region of the cerebellum disrupted eyeblink conditioning in the rat.

The S-D Model

In this section we introduce a version of the S-P-H model after first describing some issues that especially constrain the design of the network. This version is designated the S-D model. The S-D model is able to describe behavior and brain activity as they unfold in real time, is capable of stimulus configuration using plausible neurobiological mechanisms, and includes input units connected to output units through independent parallel pathways in a brainlike manner.

Stimulus Configuration

As previously described, in negative patterning two CSs are introduced separately in the presence of the US and presented together in the absence of the US. Combined presentation of both CSs generates a response smaller than the sum of the responses to each CS. In order to account for negative patterning, Spence (1936; see also Kehoe, 1988; Mishkin & Petri, 1984; Rudy & Sutherland, 1989) suggested that when both CSs are presented together they generate a "patterned" or "configural" stimulus. As a result of this configuration, the strength of a CR is proportional to the summation of the excitatory effects of the CSs and the inhibitory force of the configural stimulus. Therefore, when presented separately each CS elicits a CR larger than that elicited by both CSs presented together.

In positive patterning two CSs are introduced separately in the absence of the US and together in the presence of the US. Combined presentation of both CSs generates a response larger than the sum of the responses to each CS. In order to account for positive patterning, it is assumed that when both CSs are presented together they generate a "configural" stimulus. During training, the configural stimulus is always reinforced, whereas the individual CSs are reinforced on only half of the trials. Assuming that the individual CSs and the configural stimulus compete to form associations with the US, the association of the configural stimulus with the US becomes increasingly stronger and the associations of the individual CSs with the US become increasingly weaker. After a certain number of trials, combined presentation of the CSs elicits a CR stronger than the sum of the CRs elicited by each CS when presented in isolation. In logical terms, whereas negative patterning is equivalent to learning an "exclusive-or" response rule, positive patterning is similar to learning an "and" response rule.

Neural Networks With Internal Representations

Some neural network theories assume that in classical conditioning CS–US associations can be represented by the efficacy of synapses that connect one neural population excited by the CS with a second neural population that is excited by the US.

This second population controls the generation of the CR. At the beginning of training, synaptic strengths are small, and therefore the CS is incapable of exciting the second neural population and generating a CR. As training progresses, synaptic strengths gradually increment and the CS comes to generate a CR. Different rules have been proposed to describe changes in CS–US associations and, presumably, in the underlying synaptic strengths (e.g., Grossberg, 1975). A popular rule, proposed independently in psychological (Rescorla & Wagner, 1972) and neural network (Sutton & Barto, 1981; Widrow & Hoff, 1960) domains, has been termed the *delta rule.* The delta rule describes changes in the synaptic connections between the two neural populations by minimizing the squared value of the difference between the output of the population controlling the CR generation and the US.

The S-P-H model is an example of a network that utilizes a variation of a delta rule to change the connection strengths between CS inputs and the CR output. In the S-P-H model, CS–US associations are changed until the absolute difference between the US intensity and the CR (or the "aggregate prediction" of the US computed on all CSs present at a given moment) is zero. The S-P-H model is a two-layer network that involves only input and output units. As all other two-layer networks, the S-P-H model is unable to solve exclusive-or problems and hence to explain negative patterning.

Rumelhart, Hinton, and Williams (1986) showed that exclusive-or problems can be solved by networks that incorporate a layer of "hidden" units positioned between input and output units. In such multilayer networks, the information coming from the input units is recoded by hidden units into an "internal representation." The exclusive-or problem can be solved if this internal representation, active only with the simultaneous presentation of both inputs, acquires an inhibitory association with the output. Because the output is proportional to the sum of the excitatory effects of the individual inputs and the inhibitory force of the internal representation, the output will be large when each input is presented separately and small when both inputs are presented together. Rumelhart et al. (1986) suggested that, rather than fixed internal representations, the system might be able to learn internal representations adequate for performing the task at hand.

It is easy to notice the similarity between Spence's (1936) account of negative patterning, suggesting that when both inputs are presented together they generate a "configural" stimulus, and Rumelhart et al.'s (1986) solution for exclusive-or problems in multilayer networks, proposing the learning of an internal representation active with the simultaneous presentation of both inputs. On the basis of this correspondence, we introduce in *The Network* section a multilayer neural network that describes stimulus configuration. In this network, the information coming from the input units is recoded by hidden units into *internal representations* assumed to be equivalent to *configural stimuli* (CNs). The recoding of input-unit activity into hidden-unit activity represents the process of stimulus configuration.

Biologically Plausible Versions of Backpropagation

Rumelhart et al. (1986) described a procedure that instructs the hidden-unit layer in a multilayer network to learn the internal representations needed to solve the problem in question. Rumelhart et al. (1986) called the procedure the "generalized delta rule." Although the generalized delta rule can be applied to networks with multiple outputs, in classical conditioning a single output unit suffices to describe the CR. In this case, the generalized delta rule describes changes in the synaptic strength between a hidden unit j and the output unit as proportional to (a) the error of the output unit, EO, given by the difference between the actual output activity and the desired output value, and (b) the derivative of the output function of the hidden unit, CN_j. In simple mathematical terms, the error signal that regulates the association of hidden unit j with the output is proportional to $d(CN_j)$ EO, where $d(CN_j)$ is the derivative of the output of hidden unit j and EO is the output error.

For the case of classical conditioning, the generalized delta rule describes changes in the synaptic strength between input units and a given hidden unit as proportional to (a) the value of the association of the hidden unit with the single output unit, (b) the output error, and (c) the derivative of the output of the hidden unit. In simple mathematical terms, the error signal that regulates the associations of hidden unit j with the input CSs is proportional to $d(CN_j)$ VN_j EO, where $d(CN_j)$ is the derivative of the output of hidden unit j, VN_j is the association of hidden unit j with the output, and EO is the output error. Because the error of the output unit is propagated back to the hidden units in proportion to the influence of each hidden unit on the output error, hidden units strongly connected to the output unit are modified proportionally more than those weakly connected to the output unit.

Several implementations of backpropagation rules in biologically plausible circuits and consistent with accepted neurobiological principles have been proposed. Parker (1985) suggested that backpropagation of the error signals may be accomplished in a biologically plausible way by using an "error network" connected to the original network. Zipser and Rumelhart (1990) further developed this idea and proposed a neural architecture composed of two networks, "one for the forward propagation of activity and the other for the backward propagation of error" (p. 198). Briefly, errors in the output units are backpropagated to the hidden units by associating the output of the units computing output errors to the output of the hidden units. Because the changes in the weights of the error signals are similar to the changes in the weights of the hidden units, the output of a backpropagation unit to a given hidden unit is proportional to the sum of the output errors multiplied by weights connecting the corresponding hidden unit with the output unit. Recently, Tesauro (1990) proposed a similar scheme to describe backpropagation of errors. In both Zipser and Rumelhart's (1990) and Tesauro's (1990) cases, the approach requires the weights in both forward and backward networks to be homologous and that this homology be maintained during the learning process. Because exactly homologous weights are difficult to guarantee, in the following subsection we suggest an alternative procedure for backpropagating errors in a neurophysiologically possible manner.

The Network

In this section we introduce the S-D network, a multilayer network that accounts for stimulus configuration in classical

conditioning. The S-D network is a real-time model that depicts behavior as a moment-to-moment phenomenon, in contrast to other models of classical conditioning that portray behavior on a trial-to-trial basis (e.g., Mackintosh, 1975; Pearce & Hall, 1980). A formal description of the S-D model, consisting of set of differential equations that depict changes in the values of neural activities and connectivities as a function of time, is presented in Appendix A.

The S-D network (a) describes behavior in real time, (b) incorporates a layer of "hidden" units positioned between input and output units that internally codes *configural stimuli,* (c) includes inputs that are connected to the output directly and indirectly through the hidden-unit layer, and (d) uses a biologically plausible backpropagation procedure to train its hidden-unit layer. In addition, the network can be physiologically implemented using neurons whose adaptive weights assume only positive values (see Appendix B). As shown in the following section (Mapping of the S-D Model Onto the Brain Circuitry), all of these features contribute to the biological relevance of the network: (a) behavior and brain activity are portrayed as they unfold in real time, (b) direct and indirect input-output connections are mapped over parallel brain circuits, and (c) error signals are propagated back to hidden units through an "error network" that extends over different regions of the brain, including the hippocampus.

Figure 1 shows a multilayer network with one input layer, one hidden-unit layer, and two output layers. Input units are activated by conditioned stimuli, CS_1 and CS_2, and the context, CX. Input units form direct associations, VS_1, VS_2, and VS_3, with the first output layer. In addition, input units form associations, VH_j, with the hidden-unit layer. In turn, hidden units form associations, VN_1, VN_2, and VN_3 with the first output layer. The output activities of the hidden-unit layer are assumed to code configural stimuli denoted by CN_1, CN_2, and CN_3.

As explained in Appendix B, we suppose that VS_1, VS_2, VS_3, VN_1, VN_2, and VN_3, whether excitatory or inhibitory, assume only positive values. Therefore, in order to obtain inhibitory effects, activities of the first output layer are summed with their corresponding signs, in a single-unit second output layer that determine the output of the system, the CR. As explained later, the assumption that associations of simple CS_is with the output units, VS_i, and of configural CN_js with the output units, VN_j, are stored in a separate layer of neural elements, instead of in a single output unit as in a traditional backpropagation architecture, has important implications for the present implementation of backpropagation.

Figure 1 shows that a CS accrues a *direct* association with the output units and becomes *configured* with other CSs in a hidden unit that in turn acquires associations with the output units. Stimulus configuration is achieved by adjusting CS–hidden-unit associations, VH_{ij}. The S-D model weights more heavily the output of the hidden units than the output of the direct inputs, a difference that gives an advantage to the hidden units to establish associations with the US and to control the CR output. The model also assumes that the rate of change in the CS–hidden-unit associations, VH_{ij}, is faster than the rate of change in the associations of simple and configural CSs with the US, VS_i and VN_j. In short, configural learning is faster than CS_i–US and CN_j–US learning, and configural stimuli may establish faster and stronger associations with the US.

Figure 2 results from adding an "error network" to Figure 1. This error network regulates the associations formed by simple and configural stimuli with the US, VS_i and VN_j, and the associations of simple stimuli with the hidden units, VH_{ij}. VS_i and VN_j associations are controlled by a delta rule that reduces the output error, EO, between the aggregate prediction of the US and the actual value of the US. This aggregate prediction, B, equals the sum of the activation by simple and configural stimuli of their association with the US ($B = \Sigma_i CS_i VS_i +$

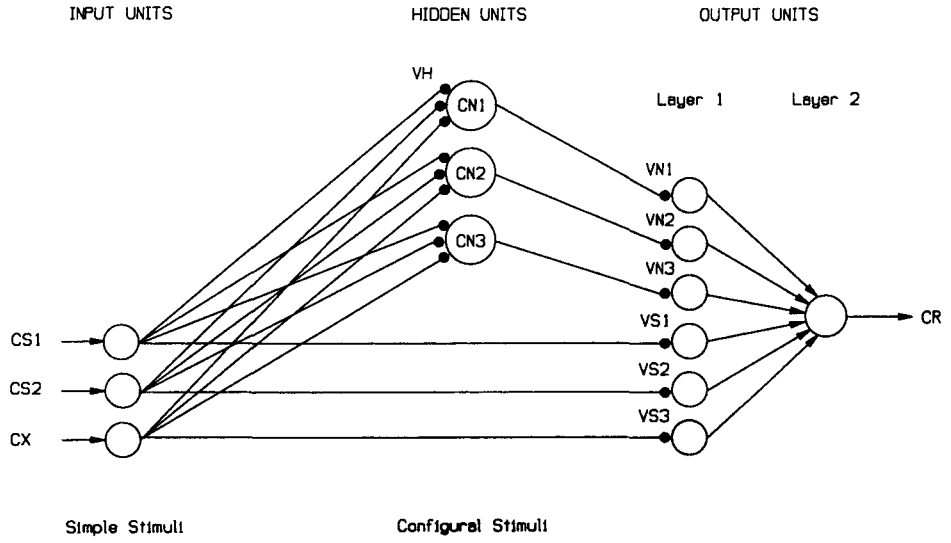

Figure 1. The S-D (Schmajuk-DiCarlo) model: Diagram of a network that incorporates a layer of hidden units capable of describing stimulus configuration in classical conditioning. (CS = conditioned stimulus; CN = configural stimulus; US = unconditioned stimulus, VS = CS–US association; VN = CN–US associations; VH = CS–CN association; CR = conditioned response; CX = context. Arrows represent fixed synapses. Solid circles represent variable synapses.)

$\Sigma_j CN_j VN_j$). Therefore, the output error is given by EO = US − $\bar{B}$. Because the CR is also given by CR = $\Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$, the output error is also expressed as EO = US − CR.

Simple-stimulus–hidden-unit associations, VH_{ij} are regulated by a biologically plausible backpropagation procedure. In the network shown in Figure 2, the associations of hidden units with the US, VN_j, are stored in a separate layer of neural elements, and therefore, the output activity of each of these neural elements is proportional to $CN_j VN_j$. Consequently, the error signal that regulates VH_{ij} associations in hidden units can be calculated by multiplying $CN_j VN_j$ activities by the output error EO. Accordingly, a given hidden unit j changes its associations with the input CSs, VH_{ij}, in proportion to the signal error $EH_j = CN_j VN_j EO$. As in backpropagation, the synaptic strengths between input CSs and a given hidden unit are changed proportionally to (a) the value of the association of the hidden unit with the output unit and (b) the output error. However, contrasting with backpropagation, the derivative of the output of the hidden unit, $d(CN_j)$, is replaced by the output of the hidden unit, CN_j. Because virtually identical results are obtained with both procedures (see Appendix A for a detailed comparison of the present and original versions of backpropagation), the neural architecture shown in Figure 2 implements a biologically plausible version of backpropagation.

The S-D network can be considered an extension of the original S-P-H model (Schmajuk & Moore, 1988), which utilizes a "simple" delta rule to regulate CS_i–US associations, to a model that uses a "generalized" delta rule to train a layer of hidden units that "configure" simple CSs.

Compound Conditioning

In order to illustrate stimulus configuration in the S-D model, we describe compound conditioning. In compound conditioning, two or more stimuli are presented together in the presence of the US. Kehoe (1986) explored the relation between configuration and summation by comparing the responding to a CS_1–CS_2 compound and its components, CS_1 and CS_2, under different proportions of reinforced presentations of CS_1–CS_2, CS_2, and CS_2. He found that responding to each component, CS_1 and CS_2, decreased with increasing proportion of CS_1–CS_2 presentations. These results suggest that CS_1–CS_2, CS_1, and CS_2 have separate representations that compete to establish associations with the US.

Kehoe's (1986) results are readily captured by the parallel input–output architecture of the S-D model shown in Figure 2. In the S-D network, a CS (a) accrues a direct association with the output units and (b) becomes configured with other CSs in a hidden unit that in turn acquires associations with the output units. Direct connections, relaying stimulus components, and indirect connections, relaying configured stimuli from the hid-

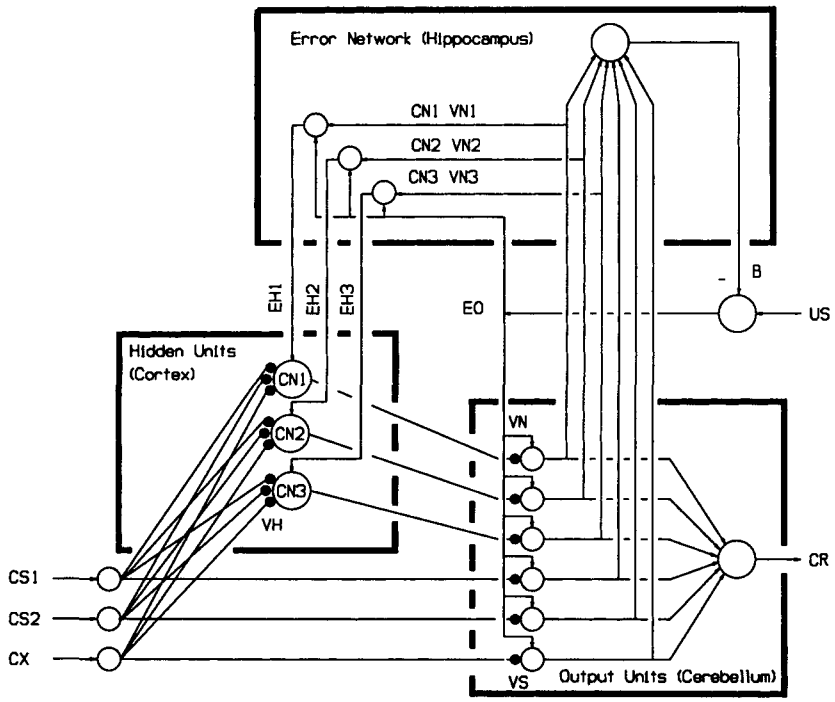

Figure 2. The S-D (Schmajuk-DiCarlo) model: Diagram of the network of Figure 1 showing error signals sent to the hidden units and output units. (CS = conditioned stimulus; CN = configural stimulus; US = unconditioned stimulus; VS = CS–US association; VN = CN–US associations; VH = CS–CN association; B = aggregate prediction; CR = conditioned response; EH = error signal for hidden units; EO = error signal for output units. Arrows represent fixed synapses. Solid circles represent variable synapses. Anatomical areas indicated in parentheses refer to the mapping of different nodes in the network onto various brain regions. CX = context.)

den-unit layer, compete to establish associations with the US at the output units. This competition is regulated by a delta rule.

Figure 3 shows experimental and simulated relationships between simple stimulus and compound stimulus responding as a function of the proportion of simple and compound acquisition trials. Figure 3 displays the average of CR peak amplitude evoked by either component, CS_1 or CS_2, after training with different proportions of CS_1, CS_2, and CS_1–CS_2 reinforced trials. Peak CR amplitude is expressed as a percentage of the peak CR relative to the CR evoked by the compound CS_1–CS_2. The percentage of compound trials is proportional to the number of CS_1–CS_2 trials divided by the total number of CS_1, CS_2, and CS_1–CS_2 trials. During CS_1 and CS_2 reinforced trials, only direct CS_i–US associations are increased. During CS_1–CS_2 reinforced trials, simple (CS_i) and configural (CN_j) stimuli compete to gain association with the US. Small percentages of compound trials result in increased responding to the components, and larger percentages of compound trials result in decreased responding to the components (because of the competition between simple and configural stimuli established by the delta rule). Figure 3 shows that the S-D model describes well Kehoe's (1986) compound conditioning data. As shown in the next section, the parallel input–output architecture of the S-D model not only provides an adequate description of compound conditioning but also contributes to the correct mapping of the model over parallel brain circuits.

Mapping the S-D Model Onto the Brain Circuitry

In this section we describe a brain mapping of the S-D model presented in the previous section under the behavioral and neurophysiological constraints presented earlier. In the case of hippocampal and cortical areas, the mapping refers in general to classical conditioning as obtained in different preparations. In the case of cerebellar areas, however, this mapping alludes specifically to the classically conditioned eyeblink or NM response.

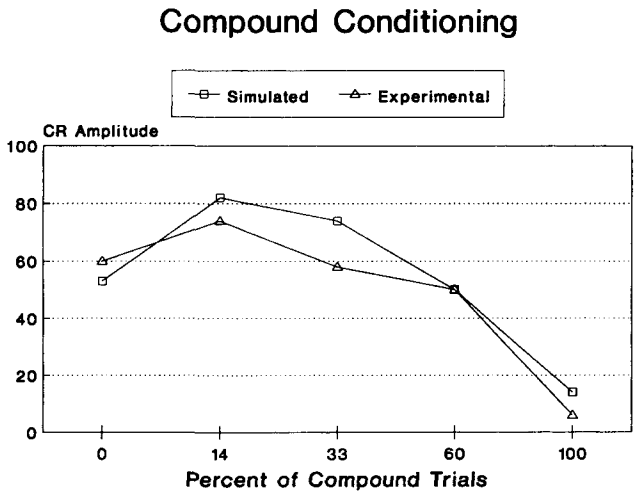

Figure 3. Compound conditioning: Relationship between simple stimulus and compound stimulus responding as a function of the proportion of simple and compound acquisition trials. (Average of CR [conditioned response] peak amplitude evoked by CS_1 and CS_2 [CS = conditioned stimulus] after training with different proportions of CS_1, CS_2, and CS_1–CS_2 reinforced trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1–CS_2. Percentage of compound trials is computed by the number of CS_1–CS_2 trials divided by the total number of CS_1, CS_2, and CS_1–CS_2 trials. Experimental data are from Kehoe, 1986.)

The mapping of different layers in the S-D network onto various brain regions, as advanced in this section, is indicated by the labels in parentheses in Figure 2. Figure 4 displays an equivalent mapping of the network onto a schematic diagram of brain interconnections. Figure 2 and Figure 4 show that output units map onto cerebellar circuits and hidden units map onto cortical circuits. Because circuits controlling output and hidden-unit errors are proposed to map onto the hippocampus, this mapping is referred to as the "error network" hypothesis of hippocampal function.

On the basis of the brain-mapped S-D network, in this section we also portray the effect of lesions of different areas of the brain. In the following section we provide computer simulations of the effects of hippocampal and cortical lesions on classical conditioning.

The Hippocampus and the Computation of Error Signals

Considering that (a) classical conditioning is still possible after HL, (b) induction of hippocampal LTP does not cause classical conditioning but only facilitates discrimination acquisition, and (c) the activity of pyramidal cells in the hippocampus *reflects* the temporal topography of the CR, Schmajuk (1984a, 1989; Schmajuk & Moore, 1985, 1988) suggested that the hippocampus is involved in the computation of the aggregate prediction of ongoing events. Aggregate prediction, $B = \Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$, represents the prediction of the intensity of the US based on all of the stimuli present at a given time. Figure 4 shows that aggregate prediction, B, is used to determine the output error signal, $EO = (US - B)$, that controls the association of different CSs and CNs with the US in the cerebellum. Notice that because the CR is also proportional to $\Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$, B is an efference copy of the CR.

Because classical conditioning paradigms that require stimulus configuration seem to be impaired by HL, it is conjectured that the computation of the error signals used by hidden units, EH_j, also takes place in the hippocampus. As shown in Figure 4, information about the $CN_j VN_j$ activity and information about the output error are combined in the hippocampus to generate hidden-unit error signals, EH_j.

Figure 4 shows that neural activity proportional to $CS_i VS_i$ and $CN_j VN_j$ reaches the hippocampus through an interpositus-nucleus–thalamic–entorhinal-cortex pathway. This information is (a) summed to generate the aggregate prediction, B, and (b) combined with the medial septal input to compute the error signal for hidden units, $EH_j = \theta CN_j VN_j$ (see Appendix C). Because each hidden unit receives its own error signal, a measure of hippocampal neural activity reflecting the computation of all error signals is given by $\Sigma_j \theta CN_j VN_j$. Therefore, according to the model, total hippocampal neural activity is proportional to $B + \Sigma_j \theta CN_j VN_j$.

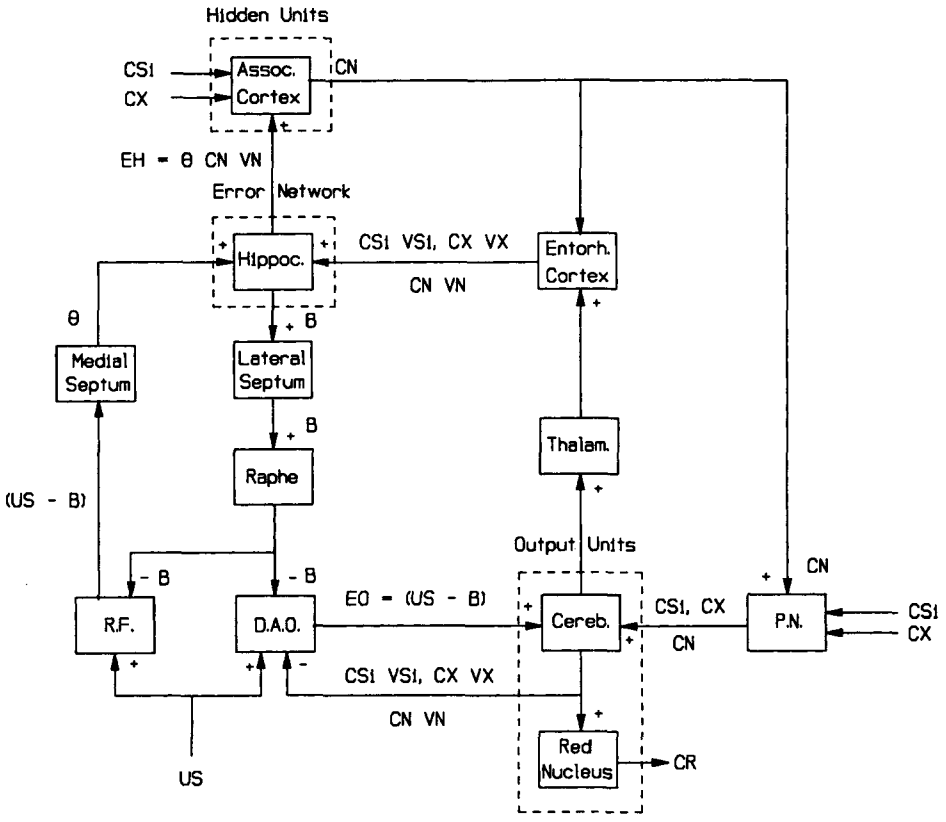

Figure 4. Mapping of the network onto a schematic diagram of cortical, hippocampal, and cerebellar interconnections. (CS_i = conditioned stimulus; CX = context; US = unconditioned stimulus; CN = configural stimulus; VS_i = CS_i–US associations; VX = context–US associations; VN = CN–US associations; CR = conditioned response; B = aggregate prediction; θ = theta rhythm; EO = [US − B] = output error; EH = θ CN VN = hidden-unit error. Assoc. Cortex = association cortex; Entorh. Cortex = entorhinal cortex; Hippoc. = hippocampus; R.F. = reticular formation; D.A.O. = dorsal accessory olive; Cereb. = cerebellum; P.N. = pontine nucleus; Thalam. = thalamus.)

Lateral Septal Neural Activity During Classical Conditioning

Because neural activity in the lateral septum during acquisition of classical conditioning is similar to that recorded in hippocampal pyramidal cells, Figure 4 shows an excitatory hippocampal output (from CA3) to the lateral septum, proportional to $B = \Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$.

Cerebellar Involvement in Classical Conditioning

Considering that classical conditioning of the eyeblink or NM response is impaired after cerebellar lesions, Figure 4 assumes that CS_i–US and CN_j–US associations are stored in cerebellar areas, thereby controlling the generation of CRs. Simple stimuli, CS_i, and configural stimuli, CN_i, reach the pontine nuclei. In the cerebellum, these sensory representations are associated with the US representation conveyed by the dorsal accessory olive. Cerebellar output through the interpositus nucleus reflects the magnitude of $CS_i VS_i$ and $CN_j VN_j$ and generates a CR by acting on the red nucleus.

Both the red and the interpositus nuclei are responsible for an inhibitory gating of olivary responsiveness to the US. We assume that each CS and each CN control their own associations with the US by inhibiting the dorsal accessory olive in proportion to their individual associations with the US. This *local* cerebellar control of CS_i–US and CN_j–US associations is assumed to be independent of the associations accrued by all other CSs and CNs with the US.

To the extent that (a) the activity of pyramidal cells in the hippocampus reflects the temporal topography of the CR, and (b) this neural activity depends on the integrity of cerebellar circuits, Figure 4 shows that copies of $CS_i VS_i$ and $CN_j VN_j$ are relayed from the interpositus nucleus to the thalamus, entorhinal cortex, and the hippocampus.

Hippocampal–Cerebellar Interactions: Modulation of Classical Conditioning

According to the circuit shown in Figure 4, the hippocampus inhibits (through the lateral septum and raphe nuclei) the dorsal accessory olive, which conveys the US input to cerebellar areas. By inhibiting the dorsal accessory olive, the hippocampus controls all CS_i–US and CN_j–US associations. Because hippocam-

pal inhibition is proportional to B, and $B = \Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$, this *global* control signal reflects the associations accrued by all other CSs and CNs with the US. As $CS_i VS_i$, $CN_j VN_j$, and B increase during acquisition, the US becomes less unexpected and the dorsal accessory olive decreases its activity, thereby preventing further increments in CS_i–US and CN_j–US cerebellar associations. Details of the interaction between local and global inhibitory mechanisms are presented in Appendixes A and B.

Reticular–Hippocampal Interactions: Control of Theta Rhythm

In Figure 4, medial septal theta represents the mismatch between the predicted and actual magnitude of the US, $\theta = |US - B|$. Appendix B shows that two theta populations may exist, one active when the US is underpredicted ($\theta_p = US - B$) and another one active when the US is overpredicted ($\theta_n = B - US$). Schmajuk and Moore (1988; Schmajuk, 1989) showed that medial septal activity can be adequately described by the sum of all mismatches between predicted and actual environmental events, including CSs and the US. Because the present version of the S-D model does not include CS–CS associations, and therefore CS predictions, the only mismatch considered here is |US − B|.

Figure 4 suggests that whereas stimulation of the reticular formation increases the US representation in the medial septum, hippocampal stimulation increases the representation of the aggregate prediction, B. According to the brain-mapped model, the hippocampus excites the raphe, which in turn inhibits the reticular formation, in proportion to the magnitude of the aggregate prediction, B. Therefore, whereas reticular formation stimulation elicits hippocampal theta, medial raphe stimulation decreases hippocampal theta.

Medial Septal Modulation of Hippocampal Activity

As mentioned earlier, medial septal activity has a "modulatory" effect on dentate granule cells and on CA1 and CA3 pyramidal cells. This modulation might be mediated by a GABAergic inhibition of inhibitory basket cell interneurons. Figure 4 shows that the hippocampus receives information related to $CN_j VN_j$ through the entorhinal cortex and information regarding theta, $\theta = |US - B|$, from the medial septum. Appendix C shows that under these circumstances the output of some dentate granule cells and CA1 and CA3 pyramidal cells is proportional to $\theta CN_j VN_j$.

Hippocampal–Cortical Interactions

According to Figure 4, stimulus configuration takes place in association cortex. The output of CA1 pyramidal neurons, proportional to $\theta CN_j VN_j$, is conveyed via the entorhinal cortex to association cortical areas, where they regulate learning in cortical hidden units, thereby modulating stimulus configuration.

Figure 4 also shows cortical areas projecting to the hippocampus through the entorhinal cortex. In a version of the S-D model that included CS–CS associations, these projections would provide information about CS–CS associations to the hippocampus in order to compute the aggregate prediction of a given CS.

Cortical–Cerebellar Interactions

In Figure 4, association cortex projections to the pontine nuclei are responsible for conveying CN_j information to the cerebellum, where they become associated with the US.

Effects of Hippocampal, Cortical, Cerebellar, and Septal Lesions

On the basis of the mapping of nodes and connections in the S-D model onto the brain circuitry presented in Figure 4, the effect of hippocampal, cortical, cerebellar, and septal lesions can be described by removing or disconnecting the appropriate blocks in the diagram.

HL produce important changes. One effect of HL is that the aggregate prediction of the US, $B = \Sigma_i CS_i VS_i + \Sigma_j CN_j VN_j$, is no longer computed. As pointed out, B is responsible for a *global* inhibition of the US representation that regulates cerebellar learning. In the absence of B, cerebellar CS_i–US and CN_j–US associations are controlled only by local inhibitions. As noted, *local* cerebellar control of CS_i–US and CN_j–US associations is independent of the associations accrued by the rest of the CSs and CNs with the US. Therefore, after HL, associations accrued by each CS or CN become independent of the associations accrued by other CSs and CNs, thereby impairing paradigms such as blocking. Because HL removes inhibition of the dorsal accessory olive, the S-D model predicts that HL animals will show increased neural activity in the dorsal accessory olive.

A second effect of HL is that hidden-unit error signals are no longer computed, and therefore no new associations are formed in association cortex, thereby impairing paradigms that require stimulus configuration, such as positive and negative patterning. Cortical associations formed before HL remain unchanged.

Notwithstanding the lack of (a) aggregate prediction signals, and (b) hidden-unit error signals, HL animals are still capable of classical conditioning by storing VS_i and VN_j associations in cerebellar areas. A formal description of the effects of hippocampal lesions is presented in Appendix D. Because the present version of the S-D model does not incorporate plastic neural elements in the hippocampus it cannot describe the effects of LTP induction in classical conditioning.

According to the circuit shown in Figure 4, CL destroy old information stored in association cortex and prevent the formation of new configural stimuli, CN_i. However, the hippocampus is still available for the computation of the aggregate prediction, B, and CL animals are still capable of classical conditioning by storing VS_i and VN_j associations in cerebellar areas. A formal description of the effects of CL is presented in Appendix D.

According to the circuit shown in Figure 4, cerebellar lesions destroy previously stored (and preclude the formation of new) CS_i–US and CN_j–US associations. A formal description of the effects of cerebellar lesions is presented in Appendix D.

Because the medial septum provides the output error value

included in the error signals used to train cortical units, medial septal lesions are somewhat equivalent to CL, which decrease the rate of acquisition of classical conditioning. This idea is compatible with Berry and Thompson's (1979) data showing that lesions of the medial septum produce retardation of acquisition of classical conditioning.

According to Figure 4, lateral septal lesions are equivalent to the elimination of hippocampal inhibition (proportional to the aggregate prediction, B) on the dorsal accessory olive. In general, this prediction is supported by the finding that simultaneous medial and lateral septum lesions and HL have similar effects in many learning paradigms (for a review see Gray & McNaughton, 1983). However, the model predicts different effects of medial and lateral septum lesions. Whereas lateral septum lesions suppress the broadcast of aggregate prediction signals to the dorsal accessory olive, consequently altering the normal formation of CS_i–US and CN_j–US associations, medial septum lesions impair the transmission of output error signals to the hippocampus, thereby impairing the formation of configural cortical associations.

The next section shows detailed computer simulations of the effect of hippocampal and cortical lesions in many classical conditioning paradigms.

Computer Simulations

By applying the brain-mapped S-D neural network described in Figure 4, in this section we contrast the experimental results regarding HL and CL in classical conditioning with computer simulations of the following paradigms: (a) acquisition of delay and trace conditioning, (b) extinction, (c) acquisition-extinction series, (d) overshadowing, (e) blocking, (f) discrimination acquisition, (g) discrimination reversal, (h) simultaneous feature-positive discrimination, (i) conditioned inhibition, (j) acquisition and retention of negative patterning, (k) acquisition and retention of positive patterning, and (l) generalization.

In order to show how similar paradigms might be learned with different involvement of diverse brain regions, we present simulations of blocking and negative patterning under various initial conditions, defined by assigning initial random values to simple-stimulus–hidden-unit associations, VH_{ij}s. Although simulation results are robust for a large range of parameter values, all simulations were carried out with identical parameter values. In Appendix E we analyze the effects that varying the number of hidden units or adding noise to the error used to train hidden units have on the acquisition of negative patterning. Parameter values used in the simulations are presented in Appendix F.

Acquisition of Delay and Trace Conditioning

Figure 5 shows real-time simulations on Trials 1, 4, 8, 12, 16, and 20 in a delay conditioning paradigm with a 200-ms CS, a 50-ms US, and a 150-ms ISI for normal, HL, and CL cases. In the normal case, as CR amplitude increases over trials, output weights VS_i and VN_j and hidden weights VH_{ij} may increase or decrease over trials.

In the HL case, CR amplitude and VS_i change at a faster rate than in the normal case. Hidden weights VH_{ij} do not change in the HL case. In the HL case, some CN_j–US learning occurs because of the initial random weights on the hidden units. This learning, however, is limited because the absence of the hippocampus precludes the formation of strong configural stimuli, CN_j. In spite of the lack of strong CN_j–US associations, the absence of the aggregate prediction of the US in the HL case allows more CS_is and CN_js to establish associations with the US, thereby attaining faster learning than in the normal case. This result is in agreement with some experimental data (Port et al., 1985; Schmaltz & Theios, 1972), but in conflict with other empirical results (Berger & Orr, 1983; Port & Patterson, 1984; Solomon & Moore, 1975).

No hidden weights VH_{ij} are present in the CL case. Because CN_j–US associations are lacking in the CL case, CL animals display a slightly slower rate of acquisition than the normal case. This result is in agreement with Oakley and Russell's (1972) data.

Simulated results for trace conditioning with a 200-ms CS, a 50-ms US, and 0-, 200-, 400-, 600-, and 800-ms ISIs show that trace conditioning is not impaired (and may be facilitated) in HL animals. These results are in agreement with some empirical data (James et al., 1987; Moyer et al., 1990; Port, Romano, & Steinmetz et al., 1986) but in conflict with other experimental results (Moyer et al., 1990). In agreement with the findings of Yeo et al. (1984), trace conditioning is not impaired in the CL case. As explained later in the Discussion section, a more elaborated description of the interaction between CS duration and ISI in determining HL effects on conditioning acquisition is offered by an alternative model (Schmajuk & DiCarlo, 1991b).

Extinction

The right panel of Figure 6 shows the number of trials to extinction for normal, HL, and CL cases. Conditioning was considered extinguished when the maximum CR amplitude was less than 0.02. Normal and HL groups exhibit comparable extinction rates, a result in agreement with part of the experimental data (Berger & Orr, 1983; Schmaltz & Theios, 1972) but in conflict with other empirical findings (Moyer et al., 1990). In agreement with findings by Oakley and Russell (1972), CL animals show normal extinction. The right panel of Figure 9 shows that, in contrast to Berger and Orr's (1983) results, the HL group is impaired in an explicitly unpaired extinction procedure (see later subsection entitled *Discrimination and Acquisition Reversal*).

In the S-D model, the rate of extinction is proportional to the output error given by Equation A6 in Appendix A. In the absence of the US, the output error in normal animals is proportional to the aggregate prediction, B, which is absent in the HL case. In simple extinction, normal animals show just a slight advantage over HL animals because the aggregate prediction, proportional to the value of the context–US association, is relatively small. In the explicitly unpaired procedure, animals are exposed to alternating trials with the US presented alone (in a given context) and trials with the CS presented alone (in the same context). Therefore, in the explicitly unpaired extinction procedure, normal animals show an important advantage over HL animals because the aggregate prediction is relatively large. As explained in the Discussion section, a more detailed charac-

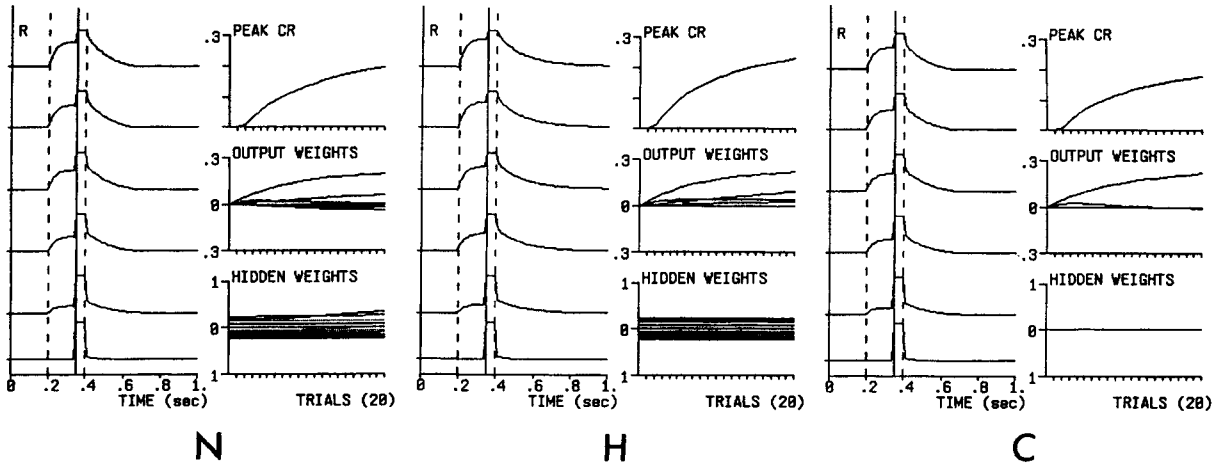

Figure 5. Effect of cortical and hippocampal lesions on acquisition of classical conditioning. (N = normal case; H = hippocampal lesion case; C = cortical lesion case. Left panels: Real-time simulated conditioned and unconditioned responses [R] on Trials 1, 4, 8, 12, 16, and 20. Vertical dashed lines indicate conditioned stimulus [CS] onset and offset. Vertical solid line indicates unconditioned stimulus [US] onset. Trial 1 is represented at the bottom of the panel. Right panels: Peak CR = peak CR as a function of trials. Output weights are average VSs [CS–US associations] and VNs [CN–US associations] as a function of trials. Hidden weights are average VHs [CS–CN associations] as a function of trials. CN = configural stimulus.)

terization of the interaction between CS duration and ISI in determining HL effects on extinction is offered by an alternative model (Schmajuk & DiCarlo, 1991b).

Acquisition-Extinction Series: Savings Effects

Five acquisition-extinction series were simulated by alternating (a) delay conditioning until a maximum CR amplitude of 0.19 was reached, with (b) extinction until a maximum CR amplitude of 0.02 was attained. The left panel of Figure 6 shows the number of trials to criterion in acquisition series for normal, CL, and HL cases. Consistent with the findings of Schmaltz and Theios (1972), HL animals display faster acquisition than normals in the first acquisition series.

In normal animals, reacquisition is faster than acquisition because the output of the hidden units increases over trials and causes VN_j associations to increase. The increasing output of the hidden units causes the system to rely more on VN_j associations than on VS_i associations over successive series. Because the learning rate in hidden units is faster than the learning rate of VN_j associations, the system becomes increasingly faster (savings effect). In HL animals, the output of the hidden units (which cannot be changed but keeps its original random value) is small, and therefore hidden units acquire and extinguish their associations with the US at a slow rate. Because VN_j and VS_i associations are not totally extinguished when the extinction criterion is reached, VN_j and VS_i associations start at a higher level, one closer to the behavioral threshold, providing some savings during successive series. In the CL case, only the second series shows savings because VS_i associations start at a higher level closer to the behavioral threshold.

The right panel of Figure 6 shows number of trials to criterion in extinction series for normal, CL, and HL cases. In the normal case, the model predicts faster extinction over series. These results are in agreement with the acquisition-extinction series data of the Schmaltz and Theios (1972) study showing that normal animals decreased the number of trials to extinction criterion. Also in agreement with Schmaltz and Theios's (1972) results, simulations show an increasing impairment in extinction in HL animals.

Stronger reliance on hidden units in normal animals implies faster extinction over successive series. As mentioned before, in HL animals the output of the hidden units is small, and therefore hidden units acquire and extinguish their association with the US at a slow rate. Because these associations extinguish at a slower rate, they require the direct associations to decrease to a lower level to attain extinction, therefore necessitating more trials to reach criterion in successive extinctions. Because the CL case lacks configural stimuli, the model predicts that CL keeps the rate of extinction constant over all series.

Overshadowing and Blocking

Figure 7 shows peak CR amplitude evoked by CS_2 in acquisition, overshadowing, and blocking for normal, HL, and CL groups. Acquisition consisted of CS_2 presentations paired together with the US during 10 trials. Overshadowing consisted of CS_1 and CS_2 presentations paired together with the US during 10 trials. Blocking consisted of 10 CS_1–CS_2 acquisition trials following 10 CS_1 acquisition trials. Figure 7 shows that the model exhibits overshadowing in normal and CL cases because the CR for CS_2 is smaller in the overshadowing condition than in simple acquisition. In addition, Figure 7 shows that the model exhibits blocking in normal and CL cases because the

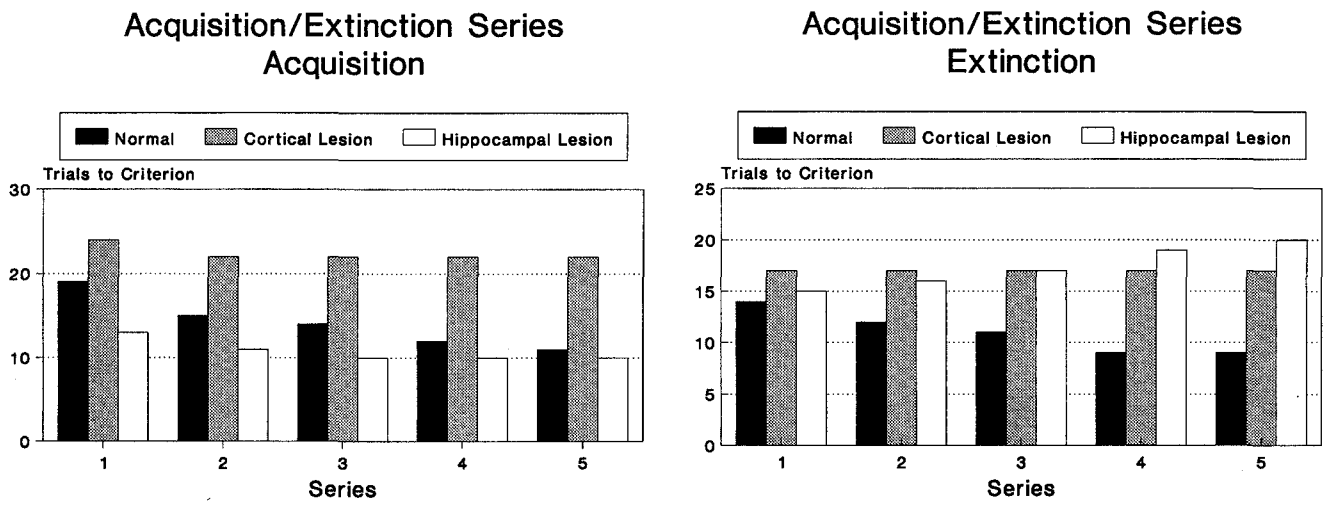

Figure 6. Effect of cortical and hippocampal lesions on acquisition-extinction series. (Acquisition [left panel]: Simulated trials to acquisition criterion [0.19] for normal and lesioned cases on five series of acquisition trials. Extinction [right panel]: Simulated trials to extinction criterion [0.02] following acquisition for normal and lesioned cases on five series of extinction trials.)

CR for CS_2 is smaller in the blocking than in the overshadowing case. The CL result for blocking is in agreement with experimental data (J. W. Moore, personal communication, December 1990).

According to the brain-mapped S-D model, HL preclude the computation of the aggregate prediction, *B*, thereby making the associations accrued by CS_2 independent of the associations accrued by CS_1 and producing deficits in overshadowing and blocking. Consistent with the results of Rickert et al. (1979) and Schmajuk et al. (1983), but not with those of Garrud et al. (1984), Figure 7 shows that overshadowing is disrupted by HL. In addition, in agreement with the findings of Rickert, Bent, Lane, and French (1978) and Solomon (1977), but not with those of Garrud et al. (1984), Figure 7 shows that HL eliminates blocking.

Figure 8 shows that even when virtually identical blocking is generated in two simulations using different initial conditions, various areas of the brain may be differentially engaged in learning. Initial conditions are defined by assigning initial random values to simple-stimulus–hidden-unit associations, VH_{ij}s. Figure 8 shows the average of the absolute change in input–hidden-unit (cortical) weights, VH_{ij}, hidden-unit–output (cerebellar) weights, VN_j, and input–output (cerebellar) weights, VS_i, after training in a blocking paradigm under two different initial conditions. Under Initial Condition A (characterized by an average VH_{ij} value of 0.028), animals establish few configural stimuli in association cortex and many CS_i–US associations in the cerebellum. Under Initial Condition B (characterized by an average VH_{ij} value of −0.023), animals show more cortical configural learning, and less CS_i–US cerebellar associations, than under Initial Condition A. The number of CN_j–US cerebellar associations is equivalent for both conditions. It is important to notice that according to the model, although blocking can be obtained without cortical involvement, cortical learning may be present during this paradigm.

Discrimination Acquisition and Reversal

The left panel of Figure 9 shows simulations of discrimination acquisition for normal, HL, and CL groups. During discrimination acquisition, 20 reinforced CS_1–US trials alternated with 20 nonreinforced CS_2 trials. CS duration was 300 ms, US duration was 50 ms, and the ISI was 250 ms. Discrimination acquisition is similar for normal, HL, and CL cases. These results agree with experimental results for HL (Berger & Orr,

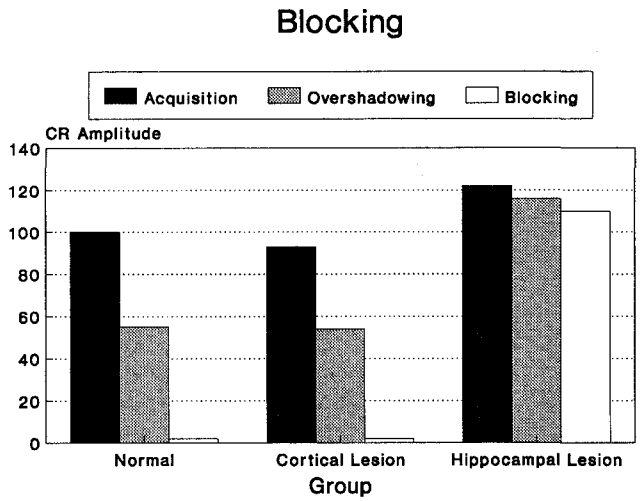

Figure 7. Effect of cortical and hippocampal lesions on blocking and overshadowing. (Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by [conditioned stimulus] CS_2 after 10 reinforced CS_1–CS_2 trials following 10 reinforced CS_1 trials in the case of blocking and 10 CS_1–CS_2 reinforced trials in the case of overshadowing. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of normal animals after 10 reinforced CS_2 trials.)

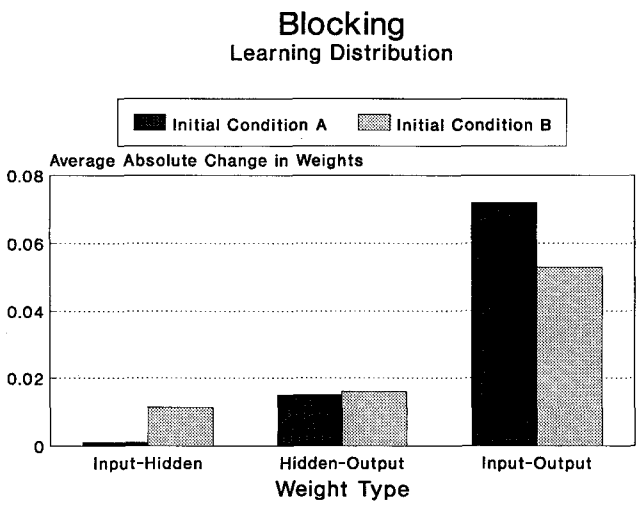

Figure 8. Learning distribution during blocking under two different initial conditions. (Average of the absolute change in input–hidden-units [VH], hidden-unit–output [VN], and input–output [VS] weights after training in a blocking paradigm. In both cases virtually identical behavioral responses were generated. Initial conditions are defined by assigning initial random values to VHs.)

1983; Berger et al., 1986; Port, Romano, Patterson, 1986; Weikart & Berger, 1986) and CL (Oakley and Russell, 1975). Also in agreement with the results of Micco and Schwartz (1972), the simulations show that CS_2 might acquire inhibitory association with the US during discrimination acquisition in normal but not in HL animals.

The right panel of Figure 9 shows simulations of the CR amplitude elicited by CS_1 in a discrimination reversal paradigm. During reversal, the original nonreinforced CS_2 is reinforced for seven trials and these trials alternate with seven nonreinforced CS_1 trials. There is a slightly impaired discrimination reversal in the CL case and a strongly impaired discrimination reversal in the HL case. The HL result is in agreement with results obtained by Berger and Orr (1983), Berger et al. (1986), Buchanan and Powell (1982), Port, Romano, and Patterson (1986), and Weikart and Berger (1986), which show a large impairment due to an increased responding to CS_1. However, contrary to experimental data obtained by Oakley and Russell (1973), the figure shows that discrimination reversal might be somewhat retarded in the CL group.

According to the model, discrimination acquisition is facilitated by HL because of the absence of the aggregate prediction of the US that allows more CSs and CNs to establish association with the US. During discrimination reversal in normal animals, because the context is alternately reinforced in the presence of CS_2 but not reinforced in the presence of CS_1, the association of the context with the US becomes inhibitory, thereby contributing to a decreased responding to CS_1 and facilitating reversal. Discrimination reversal is impaired in HL animals because when the context is alternately reinforced in the presence of CS_2 but not reinforced in the presence of CS_1, the association of the context with the US becomes excitatory (instead of inhibitory as in the normal case), thereby contributing to an increased responding to CS_1 and hindering reversal.

As mentioned before, because HL rabbits show normal explicitly unpaired extinction, Berger and Orr (1983) suggested that the increased responding to the initially reinforced CS_1 during reversal, did not reflect a deficiency in extinction of the CS_1–US association. In order to address this issue the right panel of Figure 9 shows the effect of HL and CL on explicitly unpaired and simple extinction. Explicitly unpaired extinction consisted of seven nonreinforced CS_1 trials alternated with seven reinforced context-only trials. Simple extinction consisted of seven nonreinforced CS_1 trials. The right panel of Figure 9 shows that the HL group exhibits almost normal simple extinction (see the previous subsection *Extinction*). However, in contrast to Berger and Orr's (1983) data, the HL case is impaired in explicitly unpaired extinction. Therefore, for the brain-mapped S-D model, deficits in discrimination reversal and deficits in explicitly unpaired extinction are related phenomena.

Feature-Positive Discrimination

Figure 10 shows the effect of CL and HL on the acquisition of feature-positive discrimination. During feature-positive discrimination, 50 reinforced CS_1–CS_2 trials, in which CS_1 and CS_2 were simultaneously presented, alternated with 50 nonreinforced CS_2 trials.

Figure 10 shows peak CR amplitude for normal, CL, and HL cases evoked by CS_1–CS_2 and CS_2. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1–CS_2 of normal animals. Figure 10 shows that normal and CL cases exhibit feature-positive patterning because the response to CS_1–CS_2 is large and the response to CS_2 almost negligible. The HL case does not show feature-positive discrimination because it strongly responds to CS_2, and the response to the compound CS_1–CS_2 is simply the sum of the responses to the components CS_1 and CS_2. These results are in agreement with Loechner and Weisz's (1987) data showing that acquisition of simultaneous feature-positive discrimination is impaired in HL animals. No data are available for CL effects on feature-positive discrimination.

According to the brain-mapped S-D model, feature-positive discrimination is present in normal and CL animals because CS_2 is reinforced only on half of its presentations whereas CS_1 is always reinforced, thereby overshadowing CS_2–US associations. To the extent that HL animals lack aggregate predictions, CS_2 is not overshadowed and therefore generates a strong CR.

Conditioned Inhibition

Figure 11 shows peak CR amplitude for normal, CL, and HL cases evoked by CS_1 and CS_1–CS_2 after 20 reinforced CS_1 trials alternated with 20 nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1 of normal animals. The model describes conditioned inhibition for normal animals because CS_2 becomes inhibitory and, therefore, responses to CS_1 are larger than responses to CS_1–CS_2. In contrast to experimental results showing that conditioned inhibition of the rabbit NM response is not affected by HL (Solomon, 1977), the simulations do not display conditioned inhibition in the HL case. The absence of inhibitory conditioning

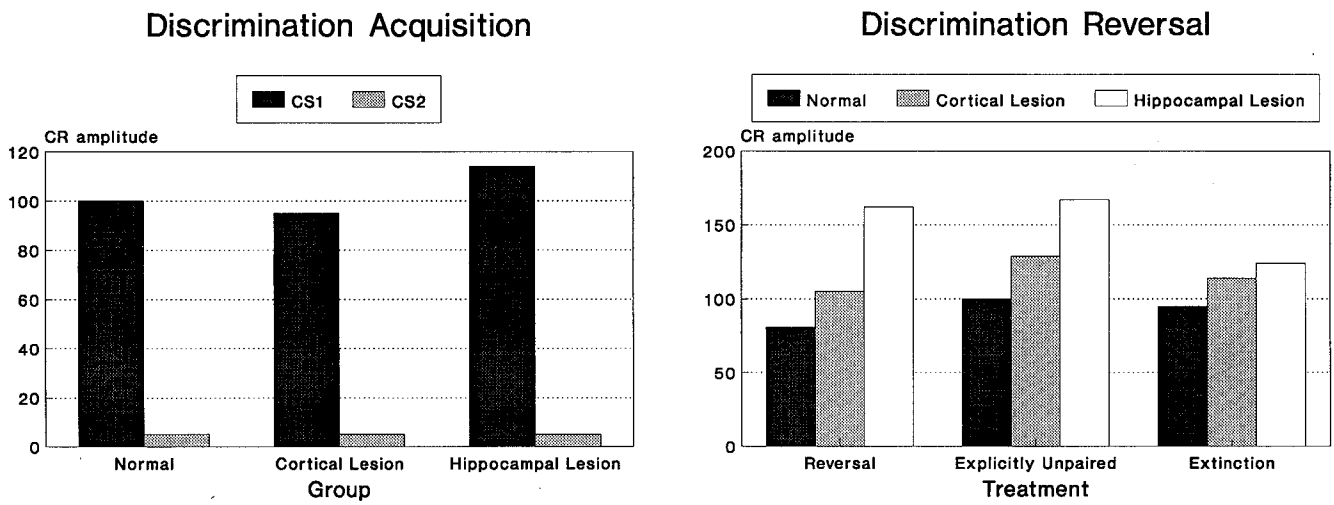

Figure 9. Effect of cortical and hippocampal lesions on discrimination. (Discrimination acquisition [left panel]: Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by CS_1 and CS_2 [CS = conditioned stimulus] after 20 reinforced CS_1 trials alternated with 20 nonreinforced CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1 of normal animals. Discrimination reversal [right panel]: Peak CR amplitude for normal and lesioned cases evoked by CS_1 after reversal, explicitly unpaired extinction, and simple extinction. Reversal consisted of 7 nonreinforced CS_1 trials alternated with 7 reinforced CS_2 trials. Explicitly unpaired extinction consisted of 7 nonreinforced CS_1 trials alternated with 7 reinforced context-only trials. Extinction consisted of 7 nonreinforced CS_1 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1 of normal animals under the explicitly unpaired extinction procedure.)

predicted by the brain-mapped S-D model is in agreement, however, with data showing that differential conditioning is impaired in HL rats (Micco & Schwartz, 1972). According to the S-D model, inhibitory conditioning is impaired in HL animals because the aggregate prediction is needed in order to generate an inhibitory association between CS_2 and the US (see Equation A6 in Appendix A and Equation D2 in Appendix D). In agreement with the findings of Moore et al. (1980), the simulations display normal conditioned inhibition in the CL case.

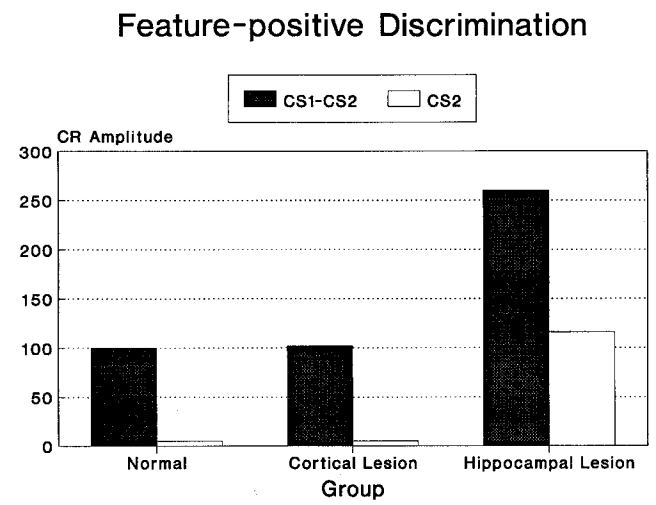

Figure 10. Effect of cortical and hippocampal lesions on feature-positive discrimination. (Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by CS_1–CS_2 and CS_2 [CS = conditioned stimulus] after 50 reinforced CS_1–CS_2 trials alternated with 50 nonreinforced CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1–CS_2 of normal animals.)

Negative Patterning

Figure 12 displays simulated data regarding the temporal course of negative patterning acquisition. Figure 12 shows the CR amplitude evoked by CS_1, CS_2, and CS_1–CS_2 over 60 simulated blocks of trials. Each block consisted of one reinforced CS_1 trial, one reinforced CS_2 trial, and one nonreinforced CS_1–CS_2 trial. CR amplitude is expressed as a proportion of the peak CR amplitude of normal animals after 20 acquisition trials. As in the Bellingham et al. (1985) experiment, at the beginning animals respond to both compound and components, and this is followed by a gradual decline of the response to the compound. However, the initial difference between responding to compound and components is larger in the simulations than in the experimental data.

Although it is usually accepted that negative patterning can be solved by creating a configural stimulus that is active when both CSs are active together and that this configural stimulus acquires inhibitory association with the US (see the previous subsection entitled *Neural Networks With Internal Representations*), computer simulations show that a variety of configural stimuli are built during negative patterning and that they acquire excitatory and inhibitory associations with the output unit.

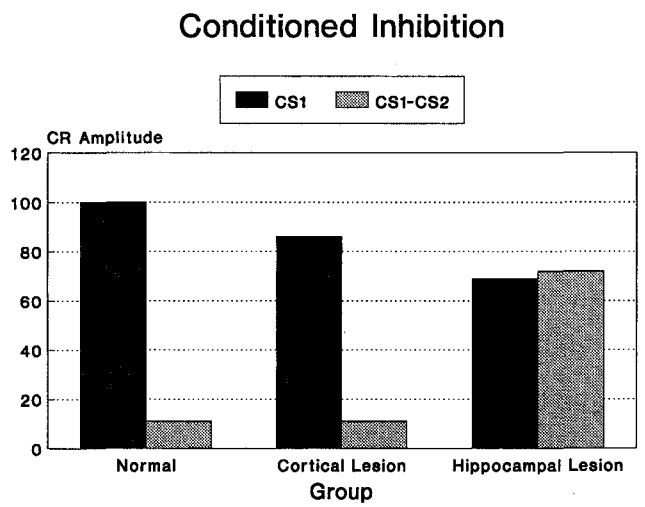

Figure 11. Effect of cortical and hippocampal lesions on conditioned inhibition. (Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by CS_1 [CS = conditioned stimulus] and CS_1–CS_2 after 20 reinforced CS_1 trials alternated with 20 nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1 of normal animals.)

Figure 13 shows that even when virtually identical negative patterning is obtained in two simulations using different initial conditions, cortex and cerebellum may be differently engaged in learning. As in blocking, initial conditions are defined by assigning initial random values to simple-stimulus–hidden-unit associations, VH_{ij}s. Figure 13 shows the average of the absolute change in input–hidden-unit (cortical) weights, VH_{ij}, hidden-unit–output (cerebellar) weights, VN_j, and input–output (cerebellar) weights, VS_i, after training normal animals in a negative patterning paradigm under two different initial conditions. In both cases, simulations proceeded until the CR to CS_1 and CS_2 was 80% of, and the CR to CS_1–CS_2 was less than 30% of, the asymptotic CR in simple acquisition. Under Initial Conditions A and B, animals establish a similar number of configural stimuli in association cortex. However, whereas under Condition A animals establish more CN_j–US cerebellar associations than under Condition B, this relation is reversed for the case of CS_i–US cerebellar associations. It is interesting to notice that negative patterning always requires cortical learning (Figure 13), whereas blocking can be solved with or without cortical associations (Figure 8).

Figure 14 shows the effect of CL and HL on negative patterning acquisition. Figure 14 shows peak CR amplitude for normal, CL, and HL cases evoked by CS_1, CS_2, and CS_1–CS_2 after 60 reinforced CS_1 trials, alternated with 60 reinforced CS_2 trials, and 60 nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of normal animals. Figure 14 shows that although normal animals exhibit negative patterning (the response to CS_1–CS_2 is smaller than the response to the sum of its components CS_1 and CS_2), CL and HL animals do not show negative patterning. This result is in agreement with Rudy and Sutherland's (1989) results in operant conditioning.

Figure 15 displays the effect of CL and HL on negative patterning retention; that is, CL and HL effects were simulated following the acquisition of negative patterning. Prelesion values in Figure 15 are those for the normal case presented in Figure 14. Figure 15 shows postlesion peak CR amplitude evoked by CS_1, CS_2, and CS_1–CS_2 after five reinforced CS_1

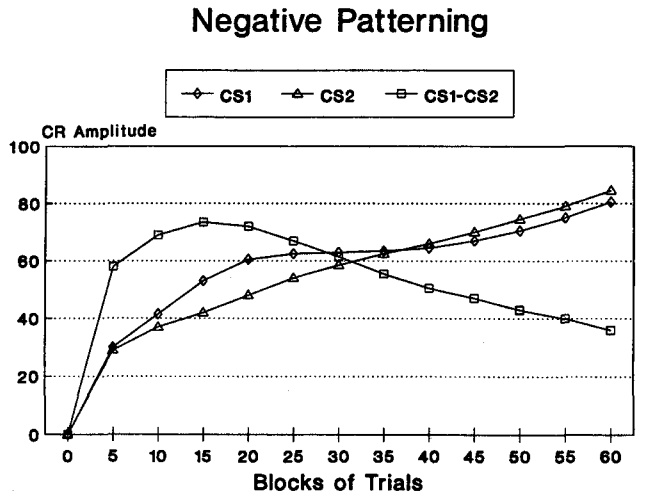

Figure 12. Temporal course of negative patterning acquisition. (CR [conditioned response] amplitude evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 over 60 blocks of trials. Each block consisted of one reinforced CS_1 trial, one reinforced CS_2 trial, and one nonreinforced CS_1–CS_2 trial. CR amplitude is expressed as a proportion of the peak CR amplitude of normal animals after 20 acquisition trials.)

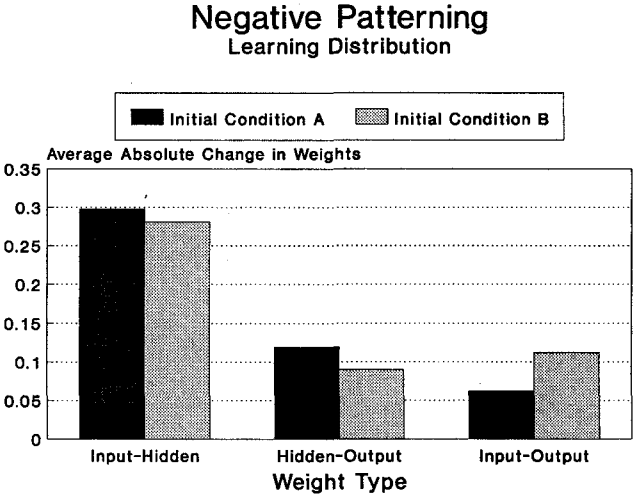

Figure 13. Learning distribution during negative patterning under two different initial conditions. (Average of the absolute change in input–hidden-unit [VH], hidden-unit–output, [VN], and input–output [VS] weights after training in a negative patterning paradigm. In both cases virtually identical behavioral responses were generated. Initial conditions are defined by assigning initial random values to VHs.)

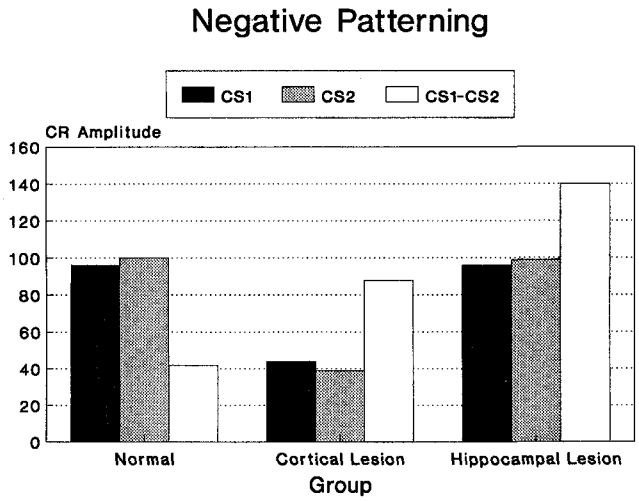

Figure 14. Effect of cortical and hippocampal lesions on negative patterning acquisition. (Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 after 60 reinforced CS_1 trials, alternated with 60 reinforced CS_2 trials, and 60 nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of normal animals.)

trials, alternated with five reinforced CS_2 trials, and five nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the prelesion peak CR to CS_2. According to Figure 15, CL and HL eliminate the previously learned negative patterning. These results are in agreement with those of Rudy and Sutherland (1989), who found that both control and HL animals lose some negative patterning after the surgical procedure. Control animals recover to prelesion levels, and HL worsen their responding, within 120 trials.

According to the brain-mapped S-D model, acquisition of negative patterning is impaired in CL and HL animals because stimulus configuration is absent in both cases. Retention of negative patterning, however, is explained in different terms. The CL case does not show retention of negative patterning because previously formed CN_js are obliterated by the lesion. The HL case does not show retention of negative patterning because, although the previously formed CN_js are not changed after HL, the system cannot sustain CN_j–US or CS_j–US inhibitory connections that reduce responding to the compound.

Appendix E analyzes the effects that (a) varying the number of hidden units or (b) adding noise to the error used to train hidden units, have on the acquisition of negative patterning. Appendix E shows that at least 4 hidden units are required in the S-D model to solve negative patterning 75% of the time. Appendix E also shows that the performance of the S-D model degrades rapidly with error noise levels greater than 5%.

Positive Patterning

Figure 16 exhibits simulated data regarding the temporal course of positive patterning acquisition. Figure 16 shows the percentage of CRs evoked by CS_1, CS_2, and CS_1–CS_2 over 60 blocks of trials. Each block consisted of one nonreinforced CS_1 trial, one nonreinforced CS_2 trial, and one reinforced CS_1–CS_2 trial. CR percentage is expressed as a proportion of the peak CR of normal animals after 20 acquisition trials. As in the Bellingham et al. (1985) experiment, at the beginning animals respond to both compound and components, and this is followed by a gradual decline of the response to the components. The temporal course of acquisition of positive patterning is very similar in the simulations and the experimental data.

Figure 17 shows the effect of CL and HL on positive patterning acquisition. Figure 17 shows peak CR amplitude for normal, CL, and HL cases evoked by CS_1, CS_2, and CS_1–CS_2 after 60 reinforced CS_1 trials, alternated with 60 reinforced CS_2 trials, and 60 nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of normal animals. Figure 17 shows that normal animals exhibit positive patterning because the response to CS_1–CS_2 is larger than the sum of the responses to its components, CS_1 and CS_2. CL and HL lesion cases do not show positive patterning because the response to the compound is simply the sum of the responses to the components. No data are available for the effects of HL and CL on acquisition of positive patterning.

Figure 18 displays the effect of CL and HL on the retention of positive patterning. Normal, CL, and HL cases were simulated following positive patterning acquisition. Prelesion values in Figure 18 are those for the normal case presented in Figure 17. Figure 18 shows postlesion peak CR amplitude evoked by CS_1, CS_2, and CS_1–CS_2 after five nonreinforced CS_1 trials, alternated with five nonreinforced CS_2 trials, and five reinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of prelesion animals. According to Figure 18, CL and HL eliminate the previously learned positive

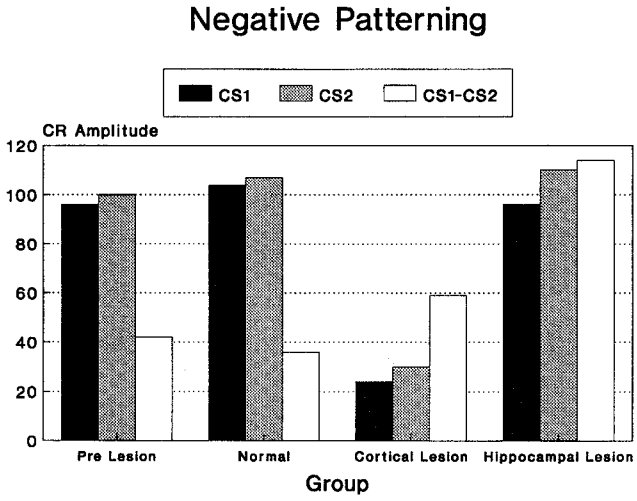

Figure 15. Effect of cortical and hippocampal lesions on negative patterning retention. (Normal and lesioned cases were simulated following negative patterning acquisition. Peak conditioned response [CR] amplitude evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 after five reinforced CS_1 trials, alternated with five reinforced CS_2 trials, and five nonreinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_2 of prelesion animals.)

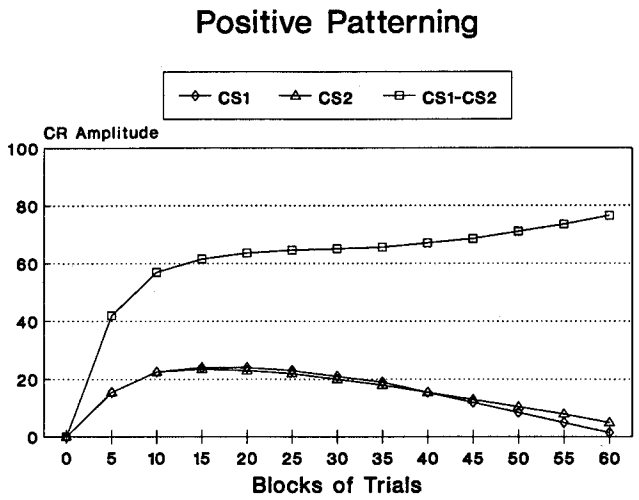

Figure 16. Temporal course of positive patterning acquisition. (Percentage of conditioned responses [CRs] evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 over 60 blocks of trials. Each block consisted of one nonreinforced CS_1 trial, one nonreinforced CS_2 trial, and one reinforced CS_1–CS_2 trial. CR percentage is expressed as a proportion of the peak CR of normal animals after 20 acquisition trials.)

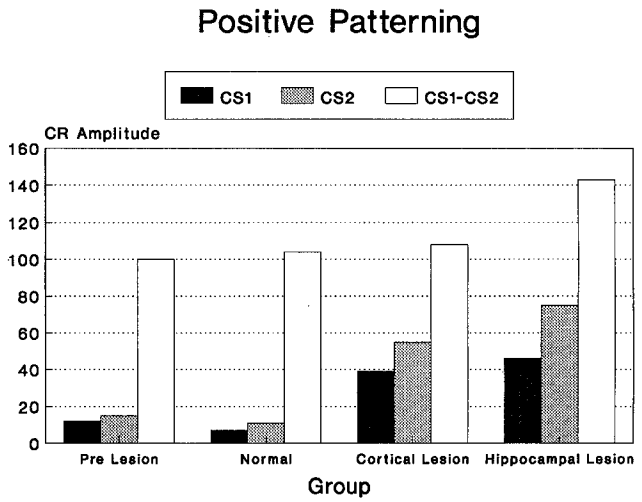

Figure 18. Effect of cortical and hippocampal lesions on positive patterning retention. (Normal and lesioned cases were simulated following positive patterning acquisition. Peak conditioned response [CR] amplitude evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 after five nonreinforced CS_1 trials, alternated with five nonreinforced CS_2 trials, and five reinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1–CS_2 of prelesion animals.)

patterning. No data are available for the effects of HL and CL on retention of positive patterning.

According to the brain-mapped S-D model, acquisition of positive patterning is impaired in HL and CL animals because stimulus configuration is absent in both cases. The CL case shows simple overshadowing, that is, the response to the compound is similar to the normal response and approximately equal to the sum of the responses to the components. Because the HL case does not show overshadowing (see previous subsection *Overshadowing and Blocking*), the response to the compound is increased over normal levels. Retention of positive patterning, however, is explained in different terms. The CL case does not show retention of positive patterning because the previously formed CN_js are obliterated by the lesion. The HL case does not show retention of positive patterning because, although the previously formed CN_js are not changed after HL, the system cannot limit the formation of CS_i–US simple associations, thereby increasing responding to the components.

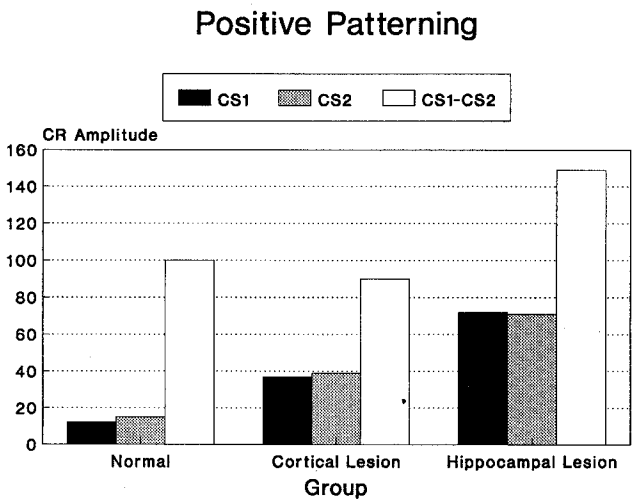

Figure 17. Effect of cortical and hippocampal lesions on positive patterning acquisition. (Peak conditioned response [CR] amplitude for normal and lesioned cases evoked by CS_1 [CS = conditioned stimulus], CS_2, and CS_1–CS_2 after 50 nonreinforced CS_1 trials, alternated with 50 nonreinforced CS_2 trials, and 50 reinforced CS_1–CS_2 trials. Peak CR amplitude is expressed as a percentage of the peak CR to CS_1–CS_2 of normal animals.)

Generalization

Generalization is measured by the responses elicited by tones of different frequencies following conditioning to a tone of a given frequency. By incorporating Blough's (1975) ideas into the S-D model, we were able to describe generalization. Blough proposed that training to a tone of given frequency generalizes to tones of other frequencies according to a generalization factor, γ, $0 \leq \gamma \leq 1$. Blough assumed that γ increases with stimulus similarity and, therefore, that γ is proportional to the ordinate of a Gaussian density distribution centered on the stimulus being presented. Because Blough proposed that associations between the different tone components and the US are regulated by a Rescorla-Wagner (delta) rule, his model is conceptually similar to, and compatible with, the S-D model.

We simulated generalization by first pairing a CS (representing a tone of a given training frequency) together with 10 CSs (representing tones of frequencies separated $\pm1, \pm2, \pm3, \pm4$, and ±5 arbitrary steps from the training frequency) with the US for

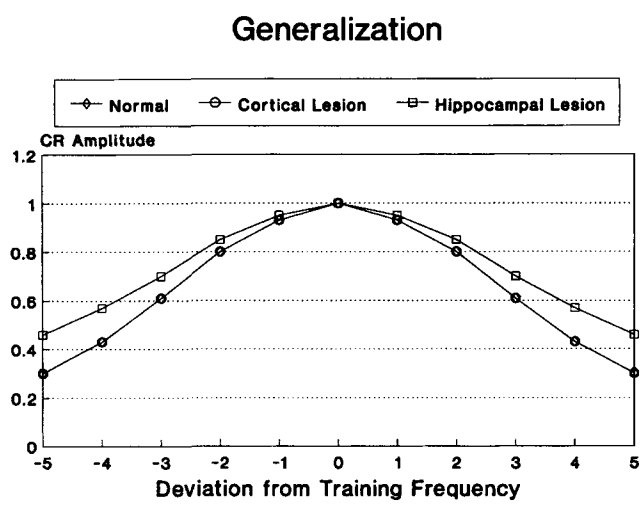

Figure 19. Effect of cortical and hippocampal lesions on generalization. (Simulated generalization curves for normal and lesioned cases generated by 10 simulated tones with frequencies that deviate from the training frequency by −66%, −33%, 0%, 33%, and 66%. In each case, CR [conditioned response] amplitude is expressed as a percentage of the peak CR to the peak CR of the original tone.)

10 trials. Generalization curves were generated by simulating the CR elicited by a CS (representing a tone of a testing frequency) together with 10 CSs (representing tones of frequencies separated ±1, ±2, ±3, ±4, and ±5 arbitrary steps from the testing frequency).

Figure 19 shows generalization curves for normal, CL, and HL cases. In each case, CR amplitude has been normalized to the value of the CR amplitude elicited by the training frequency. In all cases animals respond maximally to the training frequency and gradually less with increasing differences between the training and testing frequencies. Figure 19 shows that the generalization gradient decreases in HL animals, in agreement with Solomon and Moore's (1975) data, and does not change in CL animals. The generalization gradient decreases in HL animals because the generalized frequencies as well as the context are not overshadowed by the training CS, thereby acquiring stronger associations with the US. Therefore, all testing frequencies generate stronger CRs in HL animals than they do in the normal case.

Hippocampal Pyramidal Cell Activity During Classical Conditioning

As described earlier, the activity of some pyramidal cells is proportional to CS_iVS_i and CN_jVN_j values used to compute the aggregate prediction, $B = \Sigma_i CS_i VS_i + \Sigma_j CN_j$, and the activity of other pyramidal cells is proportional to the product θCN_jVN_j. Therefore, pyramidal cell activity is computed as $B + \theta\Sigma_j CN_jVN_j$. Because both the CR and pyramidal activity are proportional to $\Sigma_i CS_iVS_i + \Sigma_j CN_jVN_j$, in agreement with empirical results (Berger et al., 1983; Berger & Thompson, 1978a), pyramidal activity is positively correlated with the topography of the CR.

Figure 20 shows simulated CR amplitude and hippocampal neural activity during the CS and US periods for normal animals over 20 acquisition and 20 extinction trials. CR amplitude and neural activity are expressed as a percentage of their maximum values. Figure 20 shows that, in agreement with experimental data (Berger et al., 1983; Berger & Thompson, 1978a, 1982), changes in hippocampal activity during the US period precede both behavioral acquisition and extinction. Changes in pyramidal activity precede behavioral acquisition during the US period because a behavioral threshold should be exceeded to generate a CR (see Equation A11 in Appendix A). Changes in pyramidal activity precede extinction during the US period because the sum of error terms for the hidden units, $\theta\Sigma_j CN_jVN_j$, which contributes to determine pyramidal activity, decreases faster than $\Sigma_i CS_iVS_i + \Sigma_j CN_jVN_j$, which sustains both pyramidal activity and the CR.

However, in partial disagreement with empirical data (Berger & Thompson, 1982), simulated changes in pyramidal activity during the CS period precede behavioral changes during acquisition but not during extinction. During extinction, pyramidal activity and behavior decrease at the same time because θ does not decrease during the CS period (see the following subsection).

Medial Septal Neural Activity During Classical Conditioning

In the model, medial septal neural activity is proportional to the absolute difference between actual and predicted US values, $|US - B|$. Because the medial septum controls the generation of theta, hippocampal theta rhythm is also proportional to

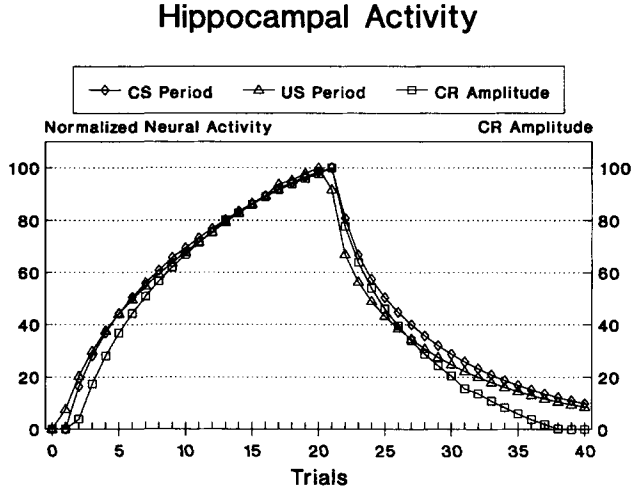

Figure 20. Hippocampal pyramidal activity during acquisition and extinction of classical conditioning. (Simulated peak conditioned response [CR] amplitude and hippocampal pyramidal activity during conditioned stimulus [CS] and unconditioned stimulus [US] periods for normal animals over 20 acquisition and 20 extinction trials. CR amplitude and neural activity are expressed as a percentage of their maximum values. Hippocampal neural activity is assumed proportional to the aggregate prediction of the US, B, plus the sum of error signals for the hidden units, $\theta \Sigma_j CN_j VN_j$.)

this value. In agreement with the data of Berger and Thompson (1978a), because hippocampal activity is proportional to B, medial septal and hippocampal activity are negatively correlated during acquisition of classical conditioning.

Figure 21 shows simulated medial septum neural activity during the US period for normal animals over 20 paired and unpaired trials. Paired and unpaired cases show reductions in $|US - B|$ because the US is predicted by the CS in the paired case and by the sound of the airpuff in the second case. Because the US is better predicted by the CS and the sound of the airpuff than by the airpuff alone, the final value of $|US - B|$ during the US period is smaller in the paired case than in the unpaired case. These simulation results are in agreement with experimental data (Berger & Thompson, 1978b). Because the present version of the S-D model does not include CS–CS associations, it cannot capture the decrease found in medial septal activity (and θ) during the CS period, a phenomenon that is well described by the S-P-H model (Schmajuk & Moore, 1988).

Discussion

The present article delineates hippocampal participation in classical conditioning in terms of a neural network that (a) describes behavior in real time, (b) incorporates a layer of "hidden" units positioned between input and output units that codes configural stimuli, (c) includes inputs that are connected to the output directly and indirectly through the hidden-unit layer, and (d) uses a biologically plausible backpropagation procedure to train the hidden-unit layer. These features contribute to the biological relevance of the network: (a) Behavior and brain activity are portrayed as they unfold in real time; (b) error signals are propagated to hidden units through an "error network" that includes the hippocampus; and (c) direct and indirect input–output connections are mapped over parallel brain circuits.

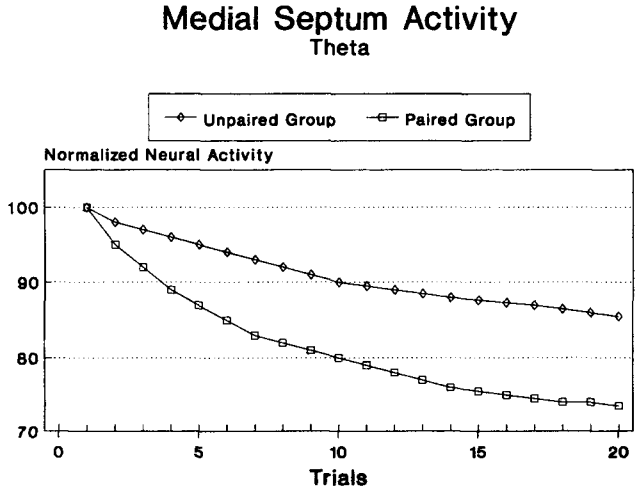

Figure 21. Medial septum activity during acquisition of classical conditioning. (Simulated medial septum neural activity during the unconditioned stimulus [US] period for normal animals over 20 paired and unpaired trials. Neural activity is expressed as a percentage of its maximum value at the beginning of acquisition. Medial septum neural activity is proportional to the absolute difference between actual and predicted US values.)

The present article extends the application of backpropagation to classical conditioning (see also Tesauro, 1990). Backpropagation has been applied to the computation of retinal and eye-position information (Zipser & Andersen, 1988), the computation of object shape from shading (Lehky & Sejnowski, 1988), text-to-speech translation, and disturbances of this phenomenon in surface dyslexia (Seidenberg & McClelland, 1990).

The biologically plausible, real-time rendition of backpropagation offered here differs from the original version in that the error signal used to train hidden units, instead of including the derivative of the activation function of the hidden units, simply contains their activation function. A possible consequence of this difference is that, although negative patterning or the exclusive-or problem can be solved with only one hidden unit in the original architecture (Rumelhart et al., 1986), the S-D model requires at least four units to solve the task more than 75% of the time. However, because of the large number of neurons available in association cortices, the use of more hidden units is not a serious obstacle for a neurophysiologically viable system.

Mapping the Network Onto the Brain Circuitry

Modelling of neural systems has been accomplished by means of both top–down and bottom–up approaches. Top–down analysis consists first of proposing possible neural mechanisms that might subserve a given behavior and then mapping these mechanisms onto brain circuits. Bottom–up analysis consists first of describing the functional properties of given brain circuits and then integrating this function with the behavior under study. In general, top–down approaches are useful as a first approximation of the understanding of the neural basis of behavior. Bottom–up approaches are valuable when the physiological interactions within a given brain circuit are known and the relation between circuit functioning and overt behavior is to be established.

In our previous studies (Schmajuk and DiCarlo, 1991a, 1991b; Schmajuk & Moore, 1988, 1989) we have made use of a top–down approach to hippocampal function: first portraying a neural network that describes associative learning and then proposing a plausible mapping of the network onto the brain circuitry. The advantage of a top–down approach, clearly demonstrated in previous articles, is that it can organize a large amount of otherwise seemingly conflicting data into the framework of a functional theory. In contrast, bottom–up approaches have serious difficulties in their attempt to derive the functional meaning of complex brain circuits from neurophysiological and anatomical information. For instance, so far it has been impossible to understand the functional significance of hippocampal CR-related neural activity in the hippocampus with a bottom-up approach. However, a serious limitation of top–down approaches is that they might disregard important aspects of the brain circuitry that remain unaddressed by the theory.

The present article shows some degree of integration of a top-down strategy (providing a functional interpretation for components and connections) and a bottom-up method (contributing descriptions of these physiological components and

connections). In the network, real-time equations describing neural activity (see Appendix A) and error equations used to train hidden units (see Appendix C) are derived from neurophysiological data.

Mapping of the S-D model onto brain circuitry is constrained by a large amount of neurophysiological information: (a) Interactions between neural elements are constrained by known anatomical connections; (b) learning sites in the model are constrained by known locations of plasticity in the cerebellum and association cortex; and (c) activities of different neural elements are constrained by neural recording data.

The S-D network assumes different types of long-term memory (LTM): (a) CS_i–US and CN_j–US associations, and (b) configural stimuli associations, VH_{ij}. The model assumes that (a) CS_i–US and CN_j–US associations are stored in cerebellar regions, and (b) VH_{ij} configural associations are stored in association cortex. A plausible alternative to the assumption that CN_j–US associations are stored in cerebellar areas is that CN_j–US associations are stored in motor cortex and relayed to the red nucleus to contribute to the generation of the CR. In this case, a copy of the aggregate prediction of the US should be conveyed to the motor cortex in order to regulate CN_j–US associations. LTM is modulated by two circuits: (a) an aggregate prediction inhibitory loop to the cerebellum, and (b) an error excitatory loop to the neocortex. Figure 4 shows that the hippocampus is part of both loops.

Based on its different types of LTM, the S-D model shows how different conditioning paradigms can be learned using alternative strategies (see *Overshadowing and Blocking* and *Conditioned Inhibition* in the previous section). Each strategy uses different proportions of simple and configural associations (VS_i, VN_j, and VH_{ij}) and, therefore, differently engages cortical and cerebellar regions even when a classical conditioning paradigm is learned in a behaviorally undistinguishable fashion. One important conjecture in the present article is that cortical VH_{ij} associations change at a faster rate than cerebellar VN_j associations. This difference in the temporal course of cortical and cerebellar LTM generates savings effects.

Once the S-D network is mapped onto the brain, the effect of lesions of different brain regions is formally studied. HL eliminates the computation and broadcasting of (a) aggregate predictions to cerebellar areas (thereby making independent the associations that CS inputs and hidden units might accrue with the US) and (b) error signals to cortical hidden units (thereby impairing the formation of new configural associations). CL eliminates old established configural representations and precludes the formation of new ones. Appendix E shows that cortical lesions of increasing extension produce increasing behavioral deficits, in a way that is reminiscent of Lashley's (1950) principle of mass action. Cerebellar lesions abolish previously learned CS_i–US and CN_j–US associations and preclude the formation of new ones.

Because the model presented here is a system that describes classical conditioning in terms of the interaction between hippocampus and other brain regions, it establishes constraints on the models for those other brain regions. In the case of the cerebellum, it has been proposed that it implements learning models such as the Rescorla and Wagner (1972) model (e.g., Donegan, Gluck, & Thompson, 1989; Thompson, 1989) or the Sutton and Barto (1981) model (Moore & Blazis, 1989). Whereas according to these two models classical conditioning paradigms such as blocking are independent of hippocampal function, according to our model, blocking is dependent on the integrity of the hippocampus. Therefore, our model requires a cerebellar model that assumes independence among the associations that each CS accrues with the US. Such a model might be attained with small variations in the mentioned models. For instance, Donegan et al. (1989) proposed that the activation of the interpositus nucleus not only generates a CR by exciting the red nucleus but also inhibits the US representation in the dorsal accessory olive. Under this assumption, blocking is a purely cerebellar phenomenon. In the brain-mapped S-D circuit, interpositus and red nuclei provide a *local* inhibition to the dorsal accessory olive proportional to CS_iVS_i and CN_jVN_j, whereas the hippocampus provides a *global* inhibition to the dorsal accessory olive proportional to the aggregate prediction *B*. Consequently, in the brain-mapped S-D model, blocking is a hippocampal-dependent phenomenon.

Functional Organization of the Limbic System

The brain-mapped S-D model shown in Figure 4 has some similarities with a model presented by Vinogradova (1975). Vinogradova suggested that the limbic system can be regarded as two interconnected circuits: an informational circuit (CA1–mammillary-bodies–anterior-thalamic-nucleus–cingulate-cortex–entorhinal-cortex) and a regulatory circuit (lateral-septum–hypothalamus–reticular-formation–medial-septum). The regulatory system controls signal processing in the informational system, through the dentate and CA3 regions. The main difference between the model described in Figure 4 and Vinogradova's (1975) model is the place where the comparison between actual and predicted events takes place. Actual and predicted events are compared in the reticular formation according to the S-D model, and in region CA3 according to the Vinogradova model.

In the brain-mapped S-D model, the comparison between real and anticipated environmental events is reflected by the medial-septal-driven theta rhythm, which controls information processing in hippocampus and cortex. It has been proposed that theta rhythm is related to orienting responses (Grastyan, Lissak, Madarasz, & Donhoffer, 1959), attentional processes (Adey, 1966), or attention to environmental cues (Bennet, 1975). The brain-mapped S-D model is in agreement with Grastyan et al.'s (1959) view that theta is proportional to the orienting response to environmental stimuli during learning. According to the model, as learning progresses during acquisition, theta decreases.

Simulation Results

The evaluation of the brain-mapped model is a critical component in the present investigation. We evaluated the model at anatomical, behavioral, and neural levels. At the anatomical level, the interconnections among neural elements in the model reflect the neuroanatomical data; that is, links in the neural architecture of the model have equivalents in pathways connecting different brain regions and in the intrinsic circuit of the

hippocampus. At the behavioral level, we contrasted computer-simulated results with experimental data for normal animals. By evaluating the model in the large variety of learning paradigms listed in Table 1 (using the same parameter values) we expect it to reflect properties of the real nervous system and to reveal more than a fortuitous selection of parameter values. At the neural level, we compared the performance of neural elements in the model with data about activity of different neural populations. This comparison provides insights into the biological plausibility of the computational processes occurring in the model.

Further insights into the model's performance come from evaluating its success after "lesioning" (removing elements from) the network. Table 1 summarizes the results of the simulation experiments for HL and CL. It shows that the brain-mapped S-D model is able to describe many HL results on delay conditioning, trace conditioning, extinction, acquisition and extinction series, blocking, overshadowing, discrimination acquisition and reversal, differential conditioning, feature-positive discrimination, and generalization. In addition, the model correctly describes the effects of HL in acquisition and retention of negative patterning in an operant discrimination task. The model has difficulty simulating the effect of HL on explicitly unpaired extinction and conditioned inhibition.

The model is also able to describe CL effects on delay conditioning, trace conditioning, extinction, conditioned inhibition, blocking, and discrimination acquisition. Yet the model has difficulties describing the effects of CL on discrimination reversal. Some of the limitations of the model seem to come from the fact that the hippocampus only interacts with the US input to the cerebellum. As explained in a later subsection, some of these difficulties are resolved by assuming that the hippocampus modulates CS inputs to cerebellar areas as well.

Table 1
Simulations of Lesion Effects Obtained With the S-D Model Compared With Experimental Results in Classical Conditioning

	Hippocampal lesion		Cortical lesion	
Paradigm	Data	Model	Data	Model
Delay conditioning	+, 0	+	−, 0	−
Trace conditioning	+, 0, −	+	0	0
Extinction	0, −	0	0	0
Explicitly unpaired extinction	0	−*	?	−
Acquisition series	−	−	?	−
Extinction series	−	−	?	−
Latent inhibition	−	?	?	?
Blocking	−, 0	−	0	0
Overshadowing	−, 0	−	?	0
Discrimination acquisition	0	0	0	0
Discrimination reversal	−	−	+	−*
Feature-positive discrimination	−	−	?	0
Conditioned inhibition	0	−*	0	0
Differential conditioning	−	−	?	0
Negative patterning acquisition	?#	−	?	−
Negative patterning retention	?#	−	?	−
Positive patterning	?	−	?	−
Generalization	+	+	?	−

Note. − = deficit; + = facilitation; 0 = no effect; ? = no available data; * = the model fails to describe accurately the experimental data; # = deficit shown in an operant discrimination task; S-D = Schmajuk-DiCarlo model.

It is interesting that even though the S-D model predicts that both acquisition and retention of negative and positive patterning are impaired by HL, the effects of HL on acquisition and retention are explained in different terms. According to the S-D model, acquisition of negative and positive patterning is impaired in HL animals because the stimulus configuration required to solve these tasks is absent. However, HL animals do not show retention of negative patterning because, although the previously formed CN_js are not changed after HL, the system cannot sustain CS_i–US or CN_j–US inhibitory connections that reduce responding to the compound. Similarly, HL animals do not show retention of positive patterning because, although the previously formed CN_js are not changed after HL, the system cannot limit direct CS_i–US associations, thereby showing increased responding to the elements of the compound.

Although the S-D model is able to describe normal behavior in all of the paradigms listed in Table 1, two important classical conditioning paradigms in which the effect of HL has been investigated are missing—namely, latent inhibition and sensory preconditioning. Latent inhibition refers to the finding that repeatedly presenting the CS alone before pairing it with the US produces retardation in acquisition. In sensory preconditioning, animals are presented with CS_1–CS_2 nonreinforced trials, followed by CS_1 reinforced trials. This procedure yields responding to CS_2. In the S-P-H model, both phenomena are explained in terms of CS–CS associations. Latent inhibition may be described as the result of the acquisition of the prediction of the CS by the context during CS preexposure, which decreases the salience of the CS, thereby retarding acquisition. Impairment of the acquisition of CS–CS associations (including context–CS associations) in HL animals precludes latent inhibition. Sensory preconditioning may be explained in terms of CS–CS associations formed during the CS–CS pairing phase, which activates CS_i–US associations formed during the second phase. Deficits in the acquisition of CS–CS associations in HL animals preclude sensory preconditioning. As is the case of the S-P-H model, an expanded version of the S-D model incorporating CS–CS associations should be able to describe HL effects on latent inhibition and sensory preconditioning. In this version of the S-D model, both configural and CS–CS associations would be stored in cortical areas under hippocampal control. Because HL impair cortical learning, the model predicts hindered latent inhibition and sensory preconditioning.

Table 2 summarizes the descriptions of neural activity provided by the brain-mapped S-D model in different brain regions. It shows that the model correctly describes neural activity of hippocampal pyramidal cells, medial septum, and lateral septum during acquisition and extinction of classical conditioning. Only pyramidal cell activity during the CS period in the course of extinction was not correctly described.

Although in the present article we have concentrated only on classical conditioning, Schmajuk (1990) showed how the S-P-H model under the aggregate prediction hypothesis can be applied to spatial learning. In order to test the aggregate prediction hypothesis in both spatial and temporal tasks, Schmajuk (1990) presented a real-time neural network capable of describ-

ing temporal discrimination and spatial learning in a unified fashion. The neural network incorporates detectors that can be tuned to a particular value of continuous temporal or spatial variables. In the temporal domain, the model describes temporal discrimination in classical conditioning and instrumental learning, classical conditioning under different ISIs, and classical conditioning and instrumental learning, classical conditioning under different ISIs, and classical conditioning with mixed ISIs. In the spatial domain, the model describes place and cue learning. Schmajuk (1990) showed that under the "aggregate prediction" hypothesis the network correctly describes activity of hippocampal "place fields" and the effect of HL in temporal and spatial learning. The aggregate prediction hypothesis proposes that HL impair paradigms in which several CSs are concurrently reinforced. In agreement with this notion, whereas acquisition of place learning (a paradigm in which animals make use of multiple distal visual stimuli to find reward) is impaired by HL (Jarrard, 1983; Morris, Garrud, Rawlins & O'Keefe, 1982), acquisition of cue learning (a paradigm in which animals make use of only one stimulus to find reward) remains unaffected by HL (Jarrard, 1983; Morris et al., 1982). Because the S-D model is an extension of the S-P-H model, the results for spatial learning obtained with the S-P-H model are also valid for the S-D model.

Novel Predictions

The model provides novel predictions for the behavioral effects of HL and CL. Some of these predictions, indicated in Table 1, are still unexplored.

In addition to HL and CL predictions, the brain-mapped S-D model predicts that medial and lateral septal lesions have different behavioral effects. Because the medial septum provides the output error values included in the error signal used to train the cortex, medial septal lesions are similar to CL, which reduce the rate of conditioning. As mentioned, this conclusion is compatible with Berry and Thompson's (1979) data showing that lesions of the medial septum produce retardation of acquisition of classical conditioning. Because the hippocampal activity that inhibits the dorsal accessory olive travels through the lateral septum, lateral septal lesions are equivalent to HL, which increase rate of acquisition. Therefore, the model predicts that, in contrast to medial septal lesions, lesions of the lateral septum produce facilitation of acquisition of classical conditioning.

The model also offers some novel neurophysiological predictions, listed in Table 2. The model predicts that, because the hippocampus has an inhibitory action over the dorsal accessory olive (see Figure 4), neural activity in the dorsal accessory olive during US presentation is larger in HL animals than in normal animals. In addition, the model suggests that because the hippocampus has an excitatory action on cortical areas (see also Figure 4), cortical activity during acquisition should be larger in normal animals than in HL animals.

Comparison With Other Theories of Hippocampal Function

Unitary psychological theories of hippocampal function. Most theories of hippocampal function are characterized by (a) assigning a unitary psychological function to the hippocampus and (b) proposing that different hippocampal regions collaborate in the same psychological function (Schmajuk, 1984b). In contrast, the brain-mapped S-D model is distinguished by (a) assigning several computational functions to the hippocampus and (b) advancing that different hippocampal regions might process incoming information in different ways.

According to Figure 4, the hippocampus receives information about (a) the associations of simple and configural stimulis with the US and (b) the mismatch between the actual and predicted US intensity. On the basis of these inputs the hippocampus computes and broadcasts (a) a signal proportional to the sum of the associations of simple and configural stimuli with the US (the aggregate prediction of the US) to cerebellar areas

Table 2
Simulations of Neural Activity Obtained With the S-D Model Compared With Experimental Results in Classical Conditioning

Brain region	Paradigm	Data	Model
Hippocampal pyramidal cells	Acquisition CS period	Increases Precedes behavior	Increases Precedes behavior
	Extinction CS period	Decreases Precedes behavior	Decreases Succeeds behavior*
	Acquisition US period	Increases Precedes behavior	Increases Precedes behavior
	Extinction US period	Decreases Precedes behavior	Decreases Precedes behavior
Lateral septum	Acquisition	Increases	Increases
Medial septum	Acquisition	Decreases	Decreases
Dorsal accessory olive	Acquisition	?	Decreases

Note. CS = conditioned stimulus; US = unconditioned stimulus; * = the model fails to describe the experimental data accurately; ? = no available data; S-D = Schmajuk-DiCarlo model.

in order to control the associations formed with the US, and (b) error signals to association cortex in order to modulate stimulus configuration. At a more molecular level, computation of aggregate prediction signals might be carried out in the CA3 region, whereas computation of error signals might be accomplished in the CA1 region. In sum, the brain-mapped S-D network strictly defines hippocampal function as the multiple transformations that occur between the hippocampal inputs and outputs, rather than as a unitary psychological function.

Configural theories of hippocampal function. The brain-mapped S-D model is not the first to incorporate the idea that the hippocampus participates in configural learning. Such an idea was first proposed by Wickelgren (1979) and more recently reelaborated by Sutherland and Rudy (1989). In this section we describe these informal models and compare them with the S-D network.

According to Wickelgren (1979), the organization of various stimuli into an associative group (chunking) is the basis of configuring in conditioning. Wickelgren (1979) proposed that the hippocampus plays a critical role in cortical chunking. The hippocampus partially activates free, as opposed to bound, cortical neurons. Bound neurons reduce their connections to the hippocampus, consolidating memory by protecting the neurons from hippocampal input. HL produce amnesia because of a disruption in chunking. According to the backpropagation procedure used in the S-D model, cortical neurons associate various components into a configural stimulus when (a) the output of that neuron has an effect on the CR output and (b) there is a mismatch between the actual and predicted US. In contrast, in Wickelgren's model cortical neurons associate different components into a new chunk when they are not yet committed to the representation of a compound stimulus. Whereas the S-D model specifies that simple and configural stimuli compete to gain association with the US according to a simple delta rule, Wickelgren's model does not indicate how component and compound stimuli interact with the US.

Recently, Sutherland and Rudy (1989) proposed that the hippocampus participates in the acquisition and storage of configural associations. According to Sutherland and Rudy (1989), although two memory systems (simple and configural) subserve learning, only the configural association system depends critically on the integrity of the hippocampal formation. Sutherland and Rudy assume (a) that memory is simultaneously stored in both the simple and configural systems and (b) that if the configural association has a greater predictive accuracy than a simple association involving one of the relevant elements, then the simple association's output is suppressed. These assumptions are not sufficient to describe blocking. As mentioned, in blocking an animal is first conditioned to CS_1, and this training is followed by conditioning to a compound consisting of CS_1 and a second stimulus, CS_2. This procedure results in a weaker conditioning to CS_2 than it would attain if paired separately with the US. There are several aspects of blocking that the theory does not seem able to address. First, the predictive value of the configural stimulus CS_1–CS_2 is identical to—not greater than—that of CS_2, and therefore it cannot inhibit the CS_2–US association. Second, even in the case that the configural stimulus could inhibit CS_2–US associations, the model cannot describe the difference between overshadowing (CS_1–CS_2–US presentations) and blocking (CS_1–US presentation preceding CS_1–CS_2–US presentations). Sutherland and Rudy claimed that the processing afforded by the hippocampal configural association system may be the basis for declarative memory as described by Cohen and Squire (1980). It should be noted, however, that because Sutherland and Rudy's theory assumes that configural associations are stored in the hippocampus, it wrongly predicts anterograde *and* retrograde impairment in memory (see later subsection *Human Amnesia*).

Although the S-D model and Sutherland and Rudy's theory share several basic assumptions, such as the inclusion of simple and configural associations, some differences should be noted. For example, whereas Sutherland and Rudy's theory asserts that HL impair only configural learning and simple learning rules remain intact, the S-D model contends that HL impair configural learning *and* change the rules for simple associations. The S-D model also diverges from the Sutherland and Rudy theory in that it assumes that configural associations are stored in association cortex and not in the hippocampus.

Integration with other computational models of hippocampal function. In the past, we have studied computational attentional models of hippocampal function that are alternatives to the S-P-H model and the aggregate prediction hypothesis. For instance, Schmajuk and Moore (1985, 1989) analyzed the effects of various hippocampal manipulations on the classically conditioned NM response in an elaborated rendering of the Moore and Stickney (1980, 1982) model called the M-S-S (Moore-Stickney-Schmajuk) model. Under the "tuning out" hypothesis suggested by Moore and Stickney (1980), the M-S-S model correctly describes the experimental effects of HL on delay conditioning, conditioning under optimal ISI, conditioned inhibition, extinction, latent inhibition, blocking, and mutual overshadowing. The model, however, is inconsistent with experimental findings describing the effects of HL on (a) trace conditioning with shock as the US under short and long ISIs, (b) trace conditioning with airpuff as the US under long ISIs, (c) discrimination reversal, and (d) sensory preconditioning. Schmajuk (1986) suggested that LTP facilitates the "tuning in" of good predictors. Under the "tuning in" hypothesis, the M-S-S model is unable to describe the effects of hippocampal induction of LTP in the acquisition of classical discrimination. Under the assumption that hippocampal neuronal activity is proportional to the magnitude of CS–US associations, the M-S-S model correctly describes hippocampal neuronal activity during acquisition and extinction of classical conditioning.

A second alternative to the S-P-H model was recently introduced by Schmajuk and DiCarlo (1991a, 1991b), who described hippocampal participation in classical conditioning in terms of Grossberg's (1975) attentional theory. According to Grossberg's attentional theory, pairing of a CS with a US causes both an association of the sensory representation of the CS with the US (conditioned reinforcement learning) and an association of the drive representation of the US with the sensory representation of the CS (incentive motivation learning). Sensory representations compete among themselves for a limited-capacity short-term memory (STM) activation that is reflected in an LTM storage. Schmajuk and DiCarlo (1991a) proposed that, in the context of Grossberg's (1975) model, the hippocampus controls self-excitation and competition among sensory representations

and stores incentive motivation associations, thereby regulating the contents of a limited-capacity STM.

Based on this hypothesis, termed the *STM regulation hypothesis,* the model predicts that HL impair phenomena that depend on competition for STM, such as blocking and overshadowing, and phenomena that depend on the duration of STM, such as trace conditioning with long ISIs. In the case of a CS presented in isolation, the model specifies the conditions under which acquisition or extinction paradigms are impaired, unaffected, or even facilitated. Therefore, the model replaces the notion that trace conditioning and extinction are or are not affected by HL with the concept that STM is more or less changed under different experimental parameters. The model also predicts that hippocampal LTP or kindling facilitate the acquisition of classical discrimination by increasing the stored values of incentive motivation associations. In addition, the model describes hippocampal neural population activity as proportional to the strength of CS–US associations. In sum, under the STM regulation hypothesis, the model provides (a) a correct description of the effect of LTP induction on discrimination acquisition (but not in reversal), (b) novel descriptions of LTP induction and blockade effects in many classical conditioning paradigms, and (c) a description of the interaction between ISI and CS durations during acquisition of classical conditioning.

Whereas the brain-mapped S-D model describes how the hippocampus regulates cerebellar learning by interacting with the US representation at the dorsal accessory olive, the M-S-S and the STM-regulation models describe how the hippocampus regulates cerebellar learning by interacting with the CS representation at the pontine nucleus. Rather than being mutually exclusive, both approaches might be complementary. A neural network incorporating both features would be able to (a) describe a wider range of paradigms and (b) yield improved descriptions of hippocampal, cortical, and cerebellar manipulations. Whereas hippocampal control of the US representation in the dorsal accessory olive is needed to explain paradigms such as conditioned inhibition, negative patterning, and positive patterning, hippocampal regulation of the CS representation in the pontine nucleus is necessary to describe paradigms such as latent inhibition, sensory preconditioning, ISI effects, and the effects of LTP induction in discrimination acquisition.

A diagram of an architecture in which the hippocampus controls both CS and US inputs to the cerebellum is shown in Figure 22. Figure 22 suggests that the hippocampus inhibits the US input to cerebellar areas in proportion to the magnitude of the aggregate prediction of the US (as defined in the S-P-H and S-D models) and excites the CS input to cerebellar areas in proportion to the magnitude of the association of the CS with the US (as defined in the M-S-S and Grossberg [1975] models).

The S-D model and the STM regulation hypothesis may also be combined to yield an improved functional description of the intrinsic hippocampal circuits. Schmajuk and DiCarlo (1991b) mapped Grossberg's neural network onto the intrinsic circuit of the hippocampus according to the STM regulation hypothesis. This mapping is constrained by the three-dimensional organization of the hippocampus, the distribution and properties of hippocampal LTP, and the firing characteristics of hippocampal CA3, CA1, and granule neurons. According to this mapping, one class of CA3 pyramidal cell population receives CS representations from the entorhinal cortex and helps to sustain these sensory representation activities through a positive feedback loop. A second class of CA3 pyramidal cell population associates (in the form of LTP) CR-related inputs with CS representations from the entorhinal cortex and constitutes part of an incentive motivation loop. A third class of CA1 pyramidal cell population is assumed to add CS information from different rostrocaudal levels of the hippocampus and constitutes a competition loop. Hippocampal outputs regulate the CS representation at the pontine nuclei.

The intrinsic mapping suggested by the STM regulation hypothesis (Schmajuk & DiCarlo, 1991b) ignores important hippocampal components and connections, such as septal inputs and basket cells, and overlooks LTP in other than perforant-path–CA3 synapses. The S-D network contains elements that might complete these aspects of the intrinsic mapping. First, the S-D model describes the interaction between medial septal activation of basket cells and entorhinal activation of pyramidal cells (see Figure C1 and Equation C3 in Appendix C). Second, the S-D model provides a functional interpretation for the information stored as LTP of perforant-path–dentate-gyrus synapses. LTP of perforant-path–dentate-gyrus synapses increases with the simultaneous activation of the perforant path and medial septal inputs to the dentate (see previous subsection *Medial Septal Modulation of Hippocampal Activity*). Whereas the perforant path input conveys CS sensory representations, the me-

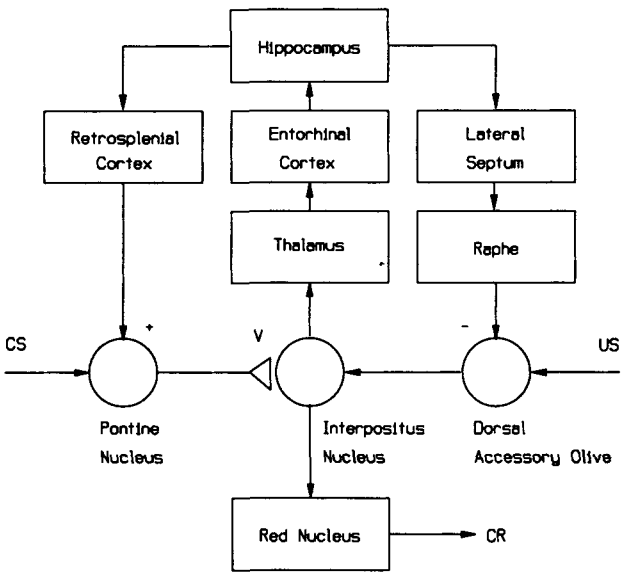

Figure 22. Hippocampal regulation of CS [conditioned stimulus] and US [unconditioned stimulus] inputs to cerebellar locus of learning. (The hippocampus inhibits the US input to cerebellar areas in proportion to the magnitude of the aggregate prediction of the US [as defined in the Schmajuk-Pearce-Hall and Schmajuk-DiCarlo models] and excites the CS input to cerebellar areas in proportion to the magnitude of the association of the CS with the US [as defined in the Moore-Stickney-Schmajuk and Grossberg (1975) models]. The open triangle represents variable synapses [V] in cerebellar areas. CR = conditioned response.)

dial septum input provides a signal proportional to the novelty of environmental events (theta rhythm). Therefore, LTP of the perforant-path–dentate-gyrus synapses might code the novelty value of a given CS. This association, stored in the form of LTP, enhances the STM of those CSs active at the time when novel events occur, thereby modulating their rate of conditioning.

Alternatives to Backpropagation

The current version of the S-D model accomplishes stimulus configuration by incorporating a hidden-unit layer trained through a backpropagation procedure. An alternative rendering of the model might incorporate a different mechanism to provide the internal representations needed to solve configural tasks.

One such mechanism is characterized in Kehoe's (1988) model. Kehoe (1988) presented a three-layer network model that describes configuration, learning to learn, and savings effects. As in Kehoe's (1988) model, the S-D model adopts a unique-stimulus hypothesis that assumes that during learning organisms "maintain a fully active representation of the separate components as well as synthesizing a configural representation of the compound" (Kehoe & Graham, 1988, p. 332). Also as in Kehoe's (1988) model, the connections of the hidden layer to the output units in the S-D model change at a slower rate than the connections from the inputs to the hidden units, thereby describing savings.

In contrast to the backpropagation procedure employed in the S-D model, in Kehoe's model all configural units are trained with the same US signal and, therefore, they all learn the same configuration, a procedure that might subtract power from the system. Unlike the S-D model, Kehoe's model does not offer an adequate description of stimulus compounding (p. 422), the increased responding to the components in positive patterning (p. 420), or a good description of savings in successive extinctions (p. 416). The major distinction, however, between the S-D model and the Kehoe model resides in the way they configure stimuli. Whereas in the Kehoe model configuring is equivalent to creating a compound stimulus active when the component stimuli are presented together (e.g., CS_1 and CS_2), in the S-D model configuring is accomplished by training hidden units that may respond to various combinations of stimuli (e.g., respond to CS_1 in the absence of CS_2).

Another alternative to backpropagation is Carpenter and Grossberg's (1987) adaptive resonance theory (ART) model. The ART model learns stimulus patterns in real time. In ART, interactions between an attentional system and an orienting subsystem enable the network to stabilize its learning, without an external teacher. The attentional system learns bottom–up codes and top–down expectancies. Mismatches between the learned top–down expectancies against input patterns activate an orienting subsystem, which resets incorrect codes and searches for new codes. According to Carpenter and Grossberg (1988), the hippocampus plays a role in detecting novelty and memory consolidation. Breakdown of the hippocampal orienting system in the network produces an amnesic syndrome. Anterograde amnesia is explained because in the absence of the orienting system, the system cannot build new categories. On the other hand, retrograde memory is intact because codes established before HL can be directly accessed without intervention of the orienting system. Therefore, the S-D model is similar to ART in that (a) cortical learning is controlled through hippocampal activation and (b) hippocampal activation reflects (only partially in the S-D model) the mismatch between expected and actual environmental inputs.

Recently, Ambros-Ingerson, Granger, and Lynch (1990) proposed a model of Layers I and II of olfactory paleocortex that can generate a series of different output responses to individual stimuli. After learning a set of olfactory stimuli, the network's first response to a stimulus indicates its category, and subsequent responses indicate increasingly specific encodings of the stimulus. Although so far the relationship between hippocampal function and cortical learning has not been specified in this network, it is possible that this model offers an alternative scheme for the backpropagation method employed in the S-D model. Some data suggest that learning in olfactory paleocortex is under hippocampal regulation. First, the system works in synchrony with a theta rhythm sample input, which might arise from the hippocampus. Second, Eichenbaum, Fagan, Mathews, and Cohen (1988) found that olfactory discrimination was impaired after hippocampal system lesions when explicit comparisons among multiple odor stimuli (and presumably cortical learning under hippocampal control) were required, but not when learning about stimuli individually was demanded.

Human Amnesia

The application of the brain-mapped S-D model might be extended to some human cognitive processes. Lesions of the medial temporal lobe, which include the hippocampus, result in a severe impairment in acquiring some types of new knowledge (anterograde amnesia) without significantly affecting old memories (limited retrograde amnesia; Milner, 1966). Although paradigms such as paired associated learning and delayed recall show dramatic impairments in amnesic patients, priming for previously learned words appear to be intact (Cohen, 1984; Graf, Shimamura & Squire, 1985; Warrington & Weiskrantz, 1970, 1974). Although not conclusive, some evidence suggests that acquisition of classical conditioning is also spared (Daum, Channon, & Canavan, 1989; Weiskrantz & Warrington, 1979). To account for these results, Cohen and Squire (1980) proposed that whereas HL impair declarative memory, procedural memory remains mostly undisturbed. Declarative memory refers primarily to memory accessible to conscious recollection, such as paired associate learning and delayed recall, whereas procedural memory mainly refers to perceptual-motor skills, including classical conditioning and priming.

In the brain-mapped S-D model, the dichotomy between impaired and preserved memories is addressed in terms of cortical versus cerebellar learning: Inasmuch as associative cortex necessitates hippocampal activation to accomplish a degree of learning, cerebellar areas demand hippocampal inhibition to limit excessive learning. Therefore, whereas according to Cohen and Squire (1980) procedural memory remains unchanged, according to the S-D model cerebellar learning is also altered by HL.

According to the brain-mapped S-D model, anterograde am-

nesia is explained as the failure to acquire cortical CS–CS associations (as suggested by Schmajuk, 1989) and configural stimuli (as suggested in the present article) in the absence of the hippocampus. Retrograde amnesia is absent because associations previously stored in cortical areas are not modified by HL. A given classical conditioning paradigm is intact to the extent that it does not depend on hippocampal integrity (see Table 1). Finally, intact priming is interpreted in terms of the activation of previously formed, and still available, cortical associations that facilitate the preferential retrieval of the primed words over others.

Summary and Conclusion

In the present article we have addressed the question of hippocampal participation in classical conditioning. In order to provide unequivocal predictions of different hippocampal manipulations on associative learning, we proposed a neural network that describes stimulus configuration. Hippocampal function was defined by mapping relevant nodes and connections in the neural network onto the hippocampus.

First, we introduced the S-D model, a multilayer network that (a) describes behavior in real time, (b) incorporates a layer of "hidden" units positioned between input and output units, (c) incorporates inputs that are connected to the output directly and indirectly through the hidden-unit layer, and (d) implements a biologically plausible backpropagation procedure to train its hidden-unit layer.

Next, we mapped nodes and connections in the S-D network onto cerebellar, cortical, and hippocampal circuits. We hypothesized that (a) CS_i–US and CN_g–US associations are stored in cerebellar regions and (b) stimulus configuration takes place in association cortex. Regarding hippocampal function, we theorized that the hippocampus computes and broadcasts (a) a signal proportional to the aggregate prediction of the US to cerebellar areas in order to control the associations formed with the US and (b) error signals to association cortex in order to modulate stimulus configuration.

Subsequently, with the network mapped onto the brain, we formally studied neural activity in different brain regions and the effect of lesions of these areas. The precise predictions provided by the model were contrasted with experimental results, and the model correctly described the effect of hippocampal and cortical lesions, as well as neural activity in hippocampus and medial septum, in many conditioning paradigms.

Finally, we suggested that improved descriptions of hippocampal manipulations might be provided by a neural network in which the hippocampus (a) regulates cerebellar learning by interacting with both US and CS cerebellar inputs and (b) modulates stimulus configuration by interacting with the association cortex.

References

Adey, W. R. (1966). Neurophysiological correlates of information transaction and storage of brain tissue. In E. Stellar & J. M. Sprague (Eds.), *Progress in physiological psychology* (Vol. 1, pp. 1–43). San Diego, CA: Academic Press.

Ambros-Ingerson, J., Granger, R., & Lynch, G. (1990). Simulation of paleocortex performs hierarchical clustering. *Science, 247,* 1344–1348.

Andersson, E., & Armstrong, D. M. (1987). Complex spikes in Purkinje cells in the lateral vermis (b zone) of the cat cerebellum during locomotion. *Journal of Physiology* (London), *385,* 107–134.

Andersson, E., Gorwicz, M., & Hesslow, G. (1987). Inferior olive excitability after high frequency climbing fiber activation in the cat. *Experimental Brain Research, 67,* 523–532.

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

Bennet, T. L. (1975). The electrical activity of the hippocampus and processes of attention. In R. L. Isaacson & K. H. Pribram (Eds.), *The hippocampus* (pp. 71–99). New York: Plenum Press.

Berger, T. W. (1984). Long-term potentiation of hippocampal synaptic transmission affects rate of behavioral learning. *Science, 224,* 627–630.

Berger, T. W., Bassett, J. L., & Weikart, C. (1985, November). *Hippocampal–cerebellar interactions during classical conditioning.* Paper presented at the annual Meeting of The Psychonomic Society, Boston, MA.

Berger, T. W., Clark, G. A., & Thompson, R. F. (1980). Learning-dependent neuronal responses recorded from limbic system brain structures during classical conditioning. *Physiological Psychology, 8,* 155–167.

Berger, T. W., & Orr, W. B. (1983). Hippocampectomy selectively disrupts discrimination reversal conditioning of the rabbit nictitating membrane response. *Behavioral Brain Research, 8,* 49–68.

Berger, T. W., Rinaldi, P. C., Weisz, D. J., & Thompson, R. F. (1983). Single-unit analysis of different hippocampal cell types during classical conditioning of rabbit nictitating membrane response. *Journal of Neurophysiology, 50,* 1197–1219.

Berger, T. W., Swanson, G. W., Milner, T. A., Lynch, G. S., & Thompson, R. F. (1980). Reciprocal anatomical connections between hippocampus and subiculum in the rabbit: Evidence for subicular innervation of regio superior. *Brain Research, 183,* 265–276.

Berger, T. W., & Thompson, R. F. (1978a). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response. I. The hippocampus. *Brain Research, 145,* 323–346.

Berger, T. W., & Thompson, R. F. (1978b). Neuronal plasticity in the limbic system during classical conditioning of the rabbit nictitating membrane response. II. Septum and mammillary bodies. *Brain Research, 156,* 293–314.

Berger, T. W., & Thompson, R. F. (1982). Hippocampal cellular plasticity during extinction of classically conditioned nictitating membrane behavior. *Behavioral Brain Research, 4,* 63–76.

Berger, T. W., Weikart, C. L., Bassett, J. L., & Orr, E. B. (1986). Lesions of the retrosplenial cortex produce deficits in reversal learning of the rabbit nictitating membrane response: Implications for potential interactions between hippocampal and cerebellar brain systems. *Behavioral Neuroscience, 100,* 802–809.

Berry, S. D., & Thompson, R. F. (1979). Medial septal lesions retard classical conditioning of the nictitating membrane response in rabbits. *Science, 205,* 209–211.

Berthier, N. E., & Moore, J. W. (1986). Cerebellar Purkinje cell activity related to the classically conditioned nicitating membrane response. *Experimental Brain Research, 63,* 341–350.

Bilkey, D. K., & Goddard, G. V. (1985). Medial septal facilitation of hippocampal granule cell activity is mediated by inhibition of inhibitory interneurons. *Brain Research, 361,* 99–106.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 104,* 3–21.

Buchanan, S. L., & Powell, D. A. (1982). Cingulate cortex: Its role in Pavlovian conditioning. *Journal of Comparative and Physiological Psychology, 96,* 755–774.

Carpenter, G., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing, 37,* 54–115.

Carpenter, G., & Grossberg, S. (1988). Neural dynamic of category learning and recognition: Attention, memory of consolidation, and amnesia. In J. Davis, R. Newburgh, & E. Wegman (Eds.), *Brain structure, learning, and memory* (pp. 233–290). Boulder, CO: Westview Press.

Clark, G. A., McCormick, D. A., Lavond, D. G., & Thompson, R. F. (1984). Effects of lesions of cerebellar nuclei on conditioned behavioral and hippocampal neuronal responses. *Brain Research, 291,* 125–136.

Cohen, N. J. (1984). Preserved learned capacity in amnesia: Evidence for multiple memory systems. In L. R. Squire & N. Butters (Eds.), *The neuropsychology of memory.* New York: Guilford Press.

Cohen, N. J., & Squire, L. R. (1980). Preserved learning and retention of pattern analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210,* 207–209.

Daum, I., Channon, S., & Canavan, A. G. M. (1989). Classical conditioning in patients with severe memory problems. *Journal of Neurology, Neurosurgery, and Psychiatry, 52,* 47–51.

Davis, K. D., & Dostrovsky, J. O. (1986). Modulatory influences of red nucleus stimulation on the somatosensory responses of cat trigeminal subnucleus oralis neurons. *Experimental Neurology, 91,* 80–101.

Donegan, N. H., Gluck, M. A., & Thompson, R. F. (1989). Integrating behavioral and biological models of classical conditioning. In R. D. Hawkins & G. H. Bower (Eds.), *Computational models of learning in simple neural systems* (pp. 109–156). San Diego, CA: Academic Press.

Eichenbaum, H., Fagan, A., Mathews, P., & Cohen, N. J. (1988). Hippocampal system dysfunction and odor discrimination learning in rats: Impairment or facilitation depending on representational demands. *Behavioral Neurosciences, 102,* 331–339.

Eichenbaum, H., Kuperstein, M., Fagan, A., & Nagode, J. (1987). Cue-sampling and goal-approach correlates of hippocampal unit activity in rats performing an odor-discrimination task. *Journal of Neuroscience, 7,* 716–732.

Foy, M. R., & Thompson, R. F. (1986). Single unit analysis of Purkinje cell discharge in classical conditioned and untrained rabbits. *Society of Neuroscience Abstracts, 12,* 518.

Frey, P. W., & Ross, L. E. (1968). Classical conditioning of the rabbit eyelid response as a function of interstimulus interval. *Journal of Comparative and Physiological Psychology, 65,* 246–250.

Garrud, P., Rawlins, J. N. P., Mackintosh, N. J., Goodal, G., Cotton, M. M., & Feldon, J. (1984). Successful overshadowing and blocking in hippocampectomized rats. *Behavioural Brain Research, 12,* 39–53.

Gormezano, I., Kehoe, E. J., & Marshall, B. S. (1983). Twenty years of classical conditioning research with the rabbit. *Progress in Psychobiology and Physiological Psychology, 10,* 197–275.

Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming among modalities and priming across category levels: Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 386–396.

Grastyan, E., Lissak, K., Madarasz, I., & Donhoffer, H. (1959). Hippocampal electrical activity during the development of conditioned reflexes. *Electroencephalography and Clinical Neurophysiology, 11,* 409–430.

Gray, J. A., & McNaughton, N. (1983). Comparison between the behavioural effects of septal and hippocampal lesions: A review. *Neuroscience and Biobehavioral Reviews, 7,* 119–188.

Grossberg, S. (1975). A neural model of attention, reinforcement, and discrimination learning. *International Review of Neurobiology, 18,* 263–327.

Holland, P. C. (1990). Forms of memory in Pavlovian conditioning. In J. L. McGaugh, N. M. Weinberger, & G. Lynch (Eds.), *Brain organization and memory* (pp. 79–105). New York: Oxford University Press.

Ito, M. (1984). *The cerebellum and neural control.* New York: Raven Press.

James, G. O., Hardiman, M. J., & Yeo, C. H. (1987). Hippocampal lesions and trace conditioning in the rabbit. *Behavioural Brain Research, 23,* 109–116.

Jarrard, L. E. (1983). Selective hippocampal lesions and behavior: Effects of kainic acid lesions on performance of place and cue tasks. *Behavioral Neuroscience, 97,* 873–889.

Kehoe, E. J. (1986). 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 layared network model of associative learning: Learning to learn and configuration. *Psychological Review, 95,* 411–433.

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.

Krnjevic, K., Ropert, N., & Casullo, J. (1988). Septohippocampal desinhibition. *Brain Research, 438,* 182–192.

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

Lehky, S. R., & Sejnowski, T. J. (1988). Network model of shape-from-shading: Neural function arises from both receptive and projective fields. *Nature, 333,* 452–454.

Loechner, K. J., & Weisz, D. J. (1987). Hippocampectomy and feature-positive discrimination. *Behavioral Brain, Research, 26,* 63–73.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Mackintosh, N. J. (1983). *Conditioning and associative learning.* Oxford, England: Clarendon Press.

McCormick, D. A., Clark, G. A., Lavond, D. G., & Thompson, R. F. (1982). Initial localization of the memory trace for a basic form of learning. *Proceedings of the National Academy of Sciences, 79,* 2731–2735.

McCormick, D. A., Steinmetz, J. E., & Thompson, R. F. (1985). Lesions of the inferior olivary complex cause extinction of the classically conditioned eyeblink response. *Brain Research, 359,* 120–130.

Micco, D. J., & Schwartz, M. (1972). Effects of hippocampal lesions upon the developments of Pavlovian internal inhibition in rats. *Journal of Comparative and Physiological Psychology, 76,* 371–377.

Milner, B. R. (1966) Amnesia following operation on temporal lobes. In C. W. N. Whitty & O. L. Zangwill (Eds.), *Amnesia* (pp. 109–133). London: Butterworths.

Mishkin, M., & Petri, H. L. (1984). Memories and habits: Some implications for the analysis of learning and retention. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (pp. 281–296). New York: Guilford Press.

Mizumori, S. J. Y., McNaughton, B. L., Barnes, C. A., & Fox, K. B. (1989). Preserved spatial coding in hippocampal CA1 pyramidal cells during reversible suppression of CA3 output: Evidence for pattern completion in hippocampus. *Journal of Neuroscience, 9,* 3915–3928.

Moore, J. W., & Blazis, D. E. J. (1989). Simulation of a classically conditioned response: A cerebellar neural network implementation of the Sutton-Barto-Desmond model. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of plasticity* (pp. 187–207). San Diego, CA: Academic Press.

Moore, J. W., & Stickney, K. J. (1980). Formation of attentional-associative networks in real time: Role of the hippocampus and implications for conditioning. *Physiological Psychology, 8,* 207–217.

Moore, J. W., & Stickney, K. J. (1982). Goal tracking in attentional-associative networks: Spatial learning and the hippocampus. *Physiological Psychology, 10,* 202–208.

Moore, J. W., Yeo, C. H., Oakley, D. A., & Russell, I. S. (1980). Conditioned inhibition of the nictitating membrane response in decorticate rabbits. *Behavioural Brain Research, 1,* 397–409.

Morris, R. G. M., Garrud, P., Rawlins, J. N. P., & O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature, 297,* 681–683.

Moyer, J. R., Deyo, R. A., & Disterhoft, J. F. (1990). Hippocampectomy disrupts associative learning of the trace conditioned eye-blink response in rabbits. *Behavioral Neuroscience, 104,* 243–252.

Oakley, D. A., & Russell, I. S. (1972). Neocortical lesions and Pavlovian conditioning. *Physiology and Behavior, 8,* 915–926.

Oakley, D. A., & Russell, I. S. (1973). Differential and reversal conditioning in partially neodecorticate rabbits. *Physiology and Behavior, 13,* 221–230.

Oakley, D. A., & Russell, I. S. (1975). Role of cortex in Pavlovian discrimination learning. *Physiology and Behavior, 15,* 315–321.

Orr, W. B., & Berger, T. W. (1985). Hippocampectomy disrupts the topography of conditioned nictitating membrane responses during reversal learning. *Journal of Comparative and Physiological Psychology, 99,* 35–45.

Parker, D. B. (1985). *Learning-logic* (Tech. Rep. TR-47). Cambridge, MA: Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552.

Port, R. L., Mikhail, A. A., & Patterson, M. M. (1985). Differential effects of hippocampectomy on classically conditioned rabbit nictitating membrane response related to interstimulus interval. *Behavioral Neuroscience, 99,* 200–208.

Port, R. L., & Patterson, M. M. (1984). Fimbrial lesions and sensory preconditioning. *Behavioral Neuroscience, 98,* 584–589.

Port, R. L., Romano, A. G., & Patterson, M. M. (1986). Stimulus duration discrimination in the rabbit: Effects of hippocampectomy on discrimination and reversal learning. *Physiological Psychology, 14,* 124–129.

Port, R. L., Romano, A. G., Steinmetz, J. E., Mikhail, A. A., & Patterson, M. M. (1986). Retention and acquisition of classical trace conditioned responses by rabbits with hippocampal lesions. *Behavioral Neuroscience, 100,* 745–752.

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

Rickert, E. J., Bent, T. L., Lane, P., & French, J. (1978). Hippocampectomy and the attenuation of blocking. *Behavioral Biology, 22,* 147–160.

Rickert, E. J., Lorden, J. F., Dawson, R., Smyly, E., & Callahan, M. F. (1979). Stimulus processing and stimulus selection in rats with hippocampal lesions. *Behavioral and Neural Biology, 27,* 454–465.

Robinson, G. B. (1986). Enhanced long-term potentiation induced in rat dentate gyrus by coactivation of septal and entorhinal inputs. *Brain Research, 379,* 56–62.

Robinson, G. B., Port, R. L., & Berger, T. W. (1989). Kindling facilitates acquisition of discriminative responding but disrupts reversal learning of the rabbit nictitating membrane response. *Behavioural Brain Research, 31,* 279–283.

Robinson, G. B., & Racine, R. J. (1986). Interactions between septal and entorhinal inputs to the rat dentate gyrus: Facilitation effects. *Brain Research, 379,* 63–67.

Ross, R. T., Orr, W. B., Holland, P. C., & Berger, T. W. (1984). Hippocampectomy disrupts acquisition and retention of learned conditional responding. *Behavioral Neuroscience, 98,* 211–225.

Rudell, A. P., Fox, S. E., & Ranck, J. B. (1980). Hippocampal excitability phase-lock to theta rhythm in walking rats. *Experimental Neurology, 68,* 87–96.

Rudy, J. W., & Sutherland, R. J. (1989). The hippocampal formation is necessary for rats to learn and remember configural discriminations. *Behavioural Brain Research, 34,* 97–109.

Rumelhart, D. E., Hinton, G. E., & Williams, G. E. (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: Bradford Books, MIT Press.

Salvatierra, A. T., & Berry, S. D. (1989). Scopolamine disruption of septo-hippocampal activity and classical conditioning. *Behavioral Neuroscience, 103,* 715–721.

Schmajuk, N. A. (1984a). A model of the effects of hippocampal lesions on Pavlovian conditioning. *Abstracts of the 14th Annual Meeting of the Society for Neuroscience, 10,* 124.

Schmajuk, N. A. (1984b). Psychological theories of hippocampal function. *Physiological Psychology, 12,* 166–183.

Schmajuk, N. A. (1986). *Real-time attentional models for classical conditioning and the hippocampus.* Unpublished doctoral dissertation, University of Massachusetts.

Schmajuk, N. A. (1987). SEAS: A dual memory architecture for computational cognitive mapping. In *Proceedings of the Ninth Annual Conference of the Cognitive Science Society* (pp. 644–654). Hillsdale, NJ: Erlbaum.

Schmajuk, N. A. (1989). The hippocampus and the control of information storage in the brain. In M. Arbib & S. I. Amari (Eds.), *Dynamic interactions in neural networks: Models and data* (pp. 53–72). New York: Springer Verlag.

Schmajuk, N. A. (1990). Role of the hippocampus in temporal and spatial navigation: An adaptive neural network. *Behavioral Brain Research, 39,* 205–229.

Schmajuk, N. A., & DiCarlo, J. J. (1991a). Neural dynamics of hippocampal modulation of classical conditioning. In M. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Neural network models of conditioning and action* (pp. 149–180). Hillsdale, NJ: Erlbaum.

Schmajuk, N. A., & DiCarlo, J. J. (1991b). A neural network approach to hippocampal function in classical conditioning. *Behavioral Neuroscience, 105,* 82–110.

Schmajuk, N. A., & Moore, J. W. (1985). Real-time attentional models for classical conditioning and the hippocampus. *Physiological Psychology, 13,* 279–290.

Schmajuk, N. A., & Moore, J. W. (1988). The hippocampus and the classically conditioned nictitating membrane response: A real-time attentional-associative model. *Psychobiology, 46,* 20–35.

Schmajuk, N. A., & Moore, J. W. (1989). Effects of hippocampal manipulations on the classically conditioned nictitating membrane response: Simulations by an attentional-associative model. *Behavioural Brain Research, 32,* 173–189.

Schmajuk, N. A., Spear, N. E., & Isaacson, R. L. (1983). Absence of overshadowing in rats with hippocampal lesions. *Physiological Psychology, 11,* 59–62.

Schmaltz, L. W., & Theios, J. (1972). Acquisition and extinction of a classically conditioned response in hippocampectomized rabbits (*Oryctolagus cuniculus*). *Journal of Comparative and Physiological Psychology, 79,* 328–333.

Sears, L. L., & Steinmetz, J. E. (1990). Acquisition of classical condi-

tioned-related activity in the hippocampus is affected by lesions of the cerebellar interpositus nucleus. *Journal of Neuroscience Research, 27,* 681–692.

Seidenberg, M. S., & McClelland, J. L. (1990). A distributed development model of word recognition and naming. *Psychological Review, 96,* 523–568.

Semple-Rowland, S. L., Bassett, J. L., & Berger, T. W. (1981). Subicular projections to retrosplenial cortex in the rabbit. *Society for Neurosciences Abstracts, 7,* 886.

Skelton, R. W. (1988). Bilateral cerebellar lesions disrupt conditioned eyelid responses in unrestrained rats. *Behavioral Neuroscience, 102,* 586–590.

Smith, M., & Gormezano, I. (1965). Effects of alternating classical conditioning and extinction sessions on the conditioned nictitating membrane response of the rabbit. *Psychonomic Science, 3,* 91–92.

Solomon, P. R. (1977). Role of the hippocampus in blocking and conditioned inhibition of rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology, 91,* 407–417.

Solomon, P. R., & Moore, J. W. (1975). Latent inhibition and stimulus generalization of the classically conditioned nictitating membrane response in rabbits (*Oryctolagus cuniculus*) following dorsal hippocampal ablation. *Journal of Comparative and Physiological Psychology, 89,* 1192–1203.

Solomon, P. R., Vander Schaaf, E. R., Thompson, R. F., & Weisz, D. J. (1986). Hippocampus and trace conditioning of the rabbit's classically conditioned nictitating membrane response. *Behavioral Neuroscience, 100,* 729–744.

Spence, K. W. (1936). The nature of discrimination learning. *Psychological Review, 43,* 427–449.

Squire, L. R., Shimamura, A. P., & Amaral, D. G. (1989). Memory and the hippocampus. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of plasticity* (pp. 208–239). San Diego, CA: Academic Press.

Steinmetz, J. E., Logan, C. G., & Thompson, R. F. (1988). Essential involvement of mossy fibers in projecting the CS to the cerebellum during classical conditioning. In C. Woody, D. Alkon, and J. McGaugh (Eds.), *Cellular mechanisms of conditioning and behavioral plasticity* (pp. 143–148). New York: Plenum Press.

Stewart, M., & Fox, S. E. (1990). Do septal neurons pace the hippocampal theta rhythm? *Trends in Neuroscience, 13,* 163–168.

Sutherland, R. J., & Rudy, J. W. (1989). Configural association theory: The role of the hippocampal formation in learning, memory, and amnesia. *Psychobiology, 17,* 129–144.

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

Tesauro, G. (1990). Neural models of classical conditioning: A theoretical viewpoint. In S. J. Hanson & C. R. Olson (Eds.), *Connectionist modelling and brain function: The developing interface* (pp. 74–104). Cambridge, MA: MIT Press.

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

Thompson, R. F. (1989). Neural circuit for classical conditioning of the eyelid closure response. In J. H. Byrne & W. O. Berry (Eds.), *Neural models of Plasticity* (pp. 160–177). San Diego, CA: Academic Press.

Vertes, R. P. (1982). Brain stem generation of hippocampal EEG. *Progress in Neurobiology, 19,* 159–186.

Vinogradova, O. S. (1975). Functional organization of the limbic system in the process of registration of information: Facts and hypothesis. In R. L. Isaacson & K. H. Pribram (Eds.), *The hippocampus* (pp. 3–69). New York: Plenum Press.

Vinogradova, O. S., Brazhnik, E. S., Karanov, A. M., & Zhadina, S. D. (1980). Neuronal activity of the septum following various types of deafferentation. *Brain Research, 187,* 353–368.

Warrington, E. K., & Weiskrantz, L. (1970). The amnesic syndrome: Consolidation or retrieval? *Nature, 228,* 628–630.

Warrington, E. K., & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia, 12,* 419–428.

Weikart, C., & Berger, T. W. (1986). Hippocampal lesions disrupt classical conditioning of cross-modality reversal learning of the rabbit nictitating membrane response. *Behavioural Brain Research, 22,* 85–90.

Weisendanger, R., & Weisendanger, M. (1982). The corticopontine system in the rat. *Journal of Comparative Neurology, 208,* 227–238.

Weiskrantz, L., & Warrington, E. K. (1979). Conditioning in amnesic patients. *Neuropsychologia, 17,* 187–194.

Weiss, C., Houk, J. C., & Gibson, A. R. (1990). Inhibition of sensory responses of cat inferior olive neurons produced by stimulation of red nucleus. *Journal of Neurophysiology, 64,* 1170–1185.

Weiss, M., & Pellet, J. (1982a). Raphe–cerebellum interactions. I. Effects of cerebellar stimulation and harmaline administration on single unit activity. *Experimental Brain Research, 48,* 163–170.

Weiss, M., & Pellet, J. (1982b). Raphe–cerebellum interactions. II. Effects of midbrain raphe stimulation and harmaline administration on single unit activity of cerebellar cortical cells in the rat. *Experimental Brain Research, 48,* 171–176.

Weisz, D. J., Clark, G. A., & Thompson, R. F. (1984). Increased responsivity of dentate granule cells during nictitating membrane response conditioning in rabbit. *Behavioural Brain Research, 12,* 145–154.

Welsh, J. P., & Harvey, J. A. (1989). Cerebellar lesions and the nictitating membrane reflex: Performance deficits of the conditioned and unconditioned response. *Journal of Neuroscience, 9,* 299–311.

Wible, C. G., Findling, R. L., Shapiro, M. W., Lang, E. J., Crane, S., & Olton, D. S. (1986). Mnemonic correlates of unit activity in the hippocampus. *Brain Research, 399,* 97–110.

Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, S–R versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review, 86,* 44–60.

Widrow, B., & Hoff, M. E. (1960). Adaptive switching circuits. Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4. 96–104.

Wilklund, L., Björklund, A., & Sjölund, B. (1977). The indolaminergic innervation of the inferior olive. 1. Convergence with the direct spinal afferents in the areas projecting to the cerebellar anterior lobe. *Brain Research, 131,* 1–21.

Wyss, J. M., & Sripanidkulchai, K. (1984). The topography of the mesencephalic and pontine projections from the cingulate cortex. *Brain Research, 293,* 1–15.

Yeo, C. H., Hardiman, M. J., Moore, J. W., & Russell, I. S. (1984). Trace conditioning of the nictitating membrane response in decorticate rabbits. *Behavioural Brain Research, 11,* 85–88.

Zipser, D., & Andersen, R. A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. *Nature, 331,* 679–684.

Zipser, D., & Rumelhart, D. E. (1990). Neurobiological significance of new learning models. In E. Schwartz (Ed.), *Computational neuroscience* (pp. 192–200). Cambridge, MA: MIT Press.

Appendix A

A Formal Description of the S-D Model

The S-D (Schmajuk-DiCarlo) model is depicted in Figures 1 and 2. Figure 4 shows a proposed mapping of the S-D model onto hippocampal, cortical, and cerebellar areas of the brain.

Short-Term Memory (STM) Traces

The STM trace of CS_i, X_i, is defined by

$$d(X_i)/dt = -K_i X_i + K_2(K_3 - X_i)CS_i, \quad \text{(A1)}$$

where $-K_i X_i$ represents the passive decay of the STM of CS_i, K_2 represents the rate of increase of X_i, and constant K_3 is the maximum possible value of X_i. K_3 can be regarded as the total number, or the percentage, of cells or membrane active sites that can be excited. Therefore, $(K_3 - X_i)$ represents the total number, or the percentage, of inactive sites that can be excited (see Grossberg, 1975).

We assume that the input representing context has a constant STM trace, $X_{context} = 0.5$.

Neural Activity

Hidden unit j is activated by the STM of different CS_is in proportion to their connections with the hidden unit:

$$sum_j = \Sigma_i VH_{ij} X_i, \quad \text{(A2)}$$

where VH_{ij} represents the association between the STM trace X_i and hidden unit j.

The output of the hidden units is a sigmoid given by

$$an_j = K_4(sum_j^n/\beta_1^n + sum_j^n). \quad \text{(A3)}$$

Whereas Rumelhart et al. (1986) assumed that the output of neural units was active even in the absence of any input, we assume that the hidden units are active only when inputs are present, that is $an_j = 0$ if $sum_j = 0$.

The output of the input units is given by

$$as_i = K_5 X_i. \quad \text{(A4)}$$

We assumed $K_4 > K_5$, which gives an advantage to the hidden units over the direct inputs to establish associations with the US.

Input–Output Associations

Changes in the associations between input i and the US, VS_i, are given by a modified delta rule

$$d(VS_i)/dt = K_6 as_i(1 - |VS_i|)EO_i. \quad \text{(A5)}$$

By Equation A5, VS_i changes only when as_i is active and the output error EO_i is not zero. Equation A5 is a Hebbian rule (VS_i changes with concurrent presynaptic activity, as_i, and postsynaptic activity, EO_i). The term $(1 - |VS_i|)$ bounds VS_i ($-1 \le VS_i \le 1$).

Output error EO_i is given by

$$EO_i = US - as_i VS_i - B_i, \quad \text{(A6)}$$

where $B_i = \Sigma_{h \neq i} as_h VS_h + \Sigma_j an_j VN_j$, that is, the aggregate prediction of the US based on every node associated with the US, different from as_i, including hidden-unit–output connections. By Equation A6, EO_i is linearly controlled by the individual local prediction of the US, $as_i VS_i$, and the aggregate prediction of the US, B_i.

Hidden-Unit–Output Associations

Changes in the association between hidden unit j and the US, VN_j, are given by a modified delta rule

$$d(VN_j)/dt = K_6 an_j(1 - |VN_j|)EO_i, \quad \text{(A7)}$$

By Equation A7, VN_j changes only when an_j is active and EO_i is not zero. As in Equation A5, Equation A7 is a Hebbian rule (VN_j changes with concurrent presynaptic activity, an_j, and postsynaptic activity, EO_i). The term $(1 - |VN_j|)$ bounds VN_j ($-1 \le V_j \le 1$).

Output error EO_i is given by

$$EO_i = US - an_j VN_j - B_j, \quad \text{(A8)}$$

where $B_j = \Sigma_{h \neq j} an_h VN_h + \Sigma_i as_i VS_i$, that is, the aggregate prediction of the US based on every node associated with the US, different from an_j, including input–output connections. Notice that output errors EO_i and EO_j used in Equations A6 and A8 are identical.

Input–Hidden-Unit Associations

Changes in the association between input i and hidden unit j, VH_{ij}, are given by a modified delta rule

$$d(VH_{ij})/dt = K_7 as_i(1 - |VH_{ij}|)EH_j, \quad \text{(A9)}$$

where term $(1 - |VH_{ij}|)$ bounds VH_{ij} between -1 and 1. By Equation A9, VH_{ij} changes only when as_i is active and the hidden-unit error, EH_j, is not zero.

Hidden-unit error, EH_j, is given by

$$EH_j = (1/(1 + e^{-K_8 an_j VN_j(US-B)})) - 0.5, \quad \text{(A10)}$$

where $B = \Sigma_i as_i VS_i + \Sigma_j an_j VN_j$, that is, the aggregate prediction of the US based on every node associated with the US. As explained in Appendix C, Equation A10 approximately describes the result of the interaction between medial septal and entorhinal cortex inputs to dentate gyrus, CA3, and CA1 hippocampal fields, as proportional to $\sqrt{|US - B|an_j VN_j}$.

The error signal used to train hidden units, described by Equation A10, is similar to the error signal for hidden units used by Rumelhart et al. (1986) in that it is proportional to (a) the value of the association of the hidden units with the output units and (b) the output error (in this case there is only one output error). Equation A10, however, differs from the original backpropagation hidden-unit error equation in that, instead of including the derivative of the activation function, it simply contains the activation function, an_j. Although the original error expression is not guaranteed to avoid the fatal problem of local minima (Rumelhart et al., 1986, p. 324), it is possible that the change introduced in the S-D model decreases the power of the network to circumvent local minima. In order to explore this possibility, we carried out several computer simulations. First, simulations using Equation A10 with either activation function an_j or its derivative yielded almost identical results. This similarity might be because $d(an_j) = (1 - an_j)\, an_j$ becomes $d(an_j) \propto an_j$ for small values of an_j, a condition generally met in our simulations. Second, simulations carried out with the original backpropagation network, using either the hidden-unit activation function or its derivative in the error term, also yielded similar results

(*Appendix A continues on next page*)

for several tasks, including the exclusive-or, parity, and encoder problems. It is important, however, that although similar results were obtained in these simulations, more hidden units were required to solve the tasks both with the S-D model and the original backpropagation network. Although in both cases the need for more hidden units might result from the use of an_j instead of $d(an_j)$, an additional factor in the S-D network might be the real-time characteristics of the system, that is, more combinations of inputs should be configured. As mentioned, because of the large number of neurons available in association cortices, the use of more hidden units is not a serious obstacle for a neurophysiologically viable system.

CR Generation

The CR output of the system is given by

$$CR = R_1[\Sigma_i as_i VS_i + \Sigma_j an_j VN_j], \quad \text{(A11)}$$

where $R_1[x] = 0$ if $x < K_9$, and $R_1[x] = x - K_9$, if $x > K_0$. K_9 is the behavioral threshold of the system.

The total output of the system, the nictitating membrane response (NMR), is given by

$$NMR = CR + UR \quad \text{(A12)}$$

(UR = unconditioned response).

Appendix B

Learning With Weights That Assume Only Positive Values

Because real synapses do not change from excitatory to inhibitory or vice versa, connectionistic models have been criticized for allowing weights to change from positive and negative and, conversely, from negative to positive. Pearce and Hall (1980) proposed a mathematical model of classical conditioning that incorporates CS–US associations that are only positive. Pearce and Hall proposed that *positive excitatory* CS–US associations increase whenever CS and US are presented together, whereas *positive inhibitory* CS–US associations increase whenever the CS is presented in the absence of the US. The net CS–US association is obtained by subtracting the positive inhibitory from the positive excitatory association. In a connectionist context, the Pearce-Hall scheme avoids the use of weights that change from positive to negative values and vice versa. However, because in the P-H model, positive excitatory and inhibitory associations increase but never decrease, they may reach unrealistically high values.

Figure B1 shows a network that also employs positive excitatory and inhibitory associations, but these always positive associations can increase and decrease. In the network illustrated in Figure B1, Node 1 receives input from the US. Node 2 receives one input proportional to CS_iV_i from Node N and another input proportional to B_i. B_i represents the aggregate prediction of the US, $B_i = \Sigma_{j \neq i} CS_j V_j$, based on all CSs other than CS_i. The output of Node 2 is proportional to $f[CS_iV_i + B_i]$, where $f[x] = (CS_iV_i + B_i)$ if $x > 0$, and $f[x] = 0$, if $x \leq 0$.

The output of Node 3 is proportional to $f[US - CS_iV - B_i]$, where $f[x] = (US - CSV - B_i)$ if $x > 0$, and $f[x] = 0$ if $x \leq 0$. The output of Node 4 is proportional to $f[CS_iV + B_i - US]$, where $f[x] = (CS_iV + B_i - US)$ if $x > 0$, and $f[x] = 0$ if $x \leq 0$. Therefore, Node 3 is active when the actual US is underpredicted ($US > CS_iV + B_i$) and Node 4 is active when the actual US is overpredicted ($CS_iV + B_i > US$). At a given time, either Node 3 or Node 4 is active, but they cannot both be active at the same time. We call the output of Node 3, $\theta_p = US - CS_iV - B_i$, and the output of Node 4, $\theta_n = CS_iV + B_i - US$. Notice that θ_p and θ_n are similar to the error expressions given by Equations A6 and A8.

Node E receives excitatory input from θ_p and inhibitory input from θ_n. Node I receives excitatory input from θ_n and inhibitory input from θ_p. The CS_i reaches both Nodes E and I through modifiable synapses, with strengths V_p and V_n, respectively. V_p and V_n are always positive and vary between 0 and 1. V_p increases whenever CS_i is active and θ_p is active; V_p decreases whenever CS_i is active and θ_n is active. V_n increases whenever CS_i is active and θ_n is active; V_n decreases whenever CS_i is active and θ_p is active. Node N computes the net output, CS_iV, by combining the output from Node E, CS_iV_p, with the output from Node N, CS_iV_n, according to $CS_i(V_p - V_m) = CS_iV$.

Similar results are obtained with the equations presented in Appendix A and the equations describing the network illustrated in Figure B1. For simplicity and speed in the simulations, in the present study we made use of the equations presented in Appendix A.

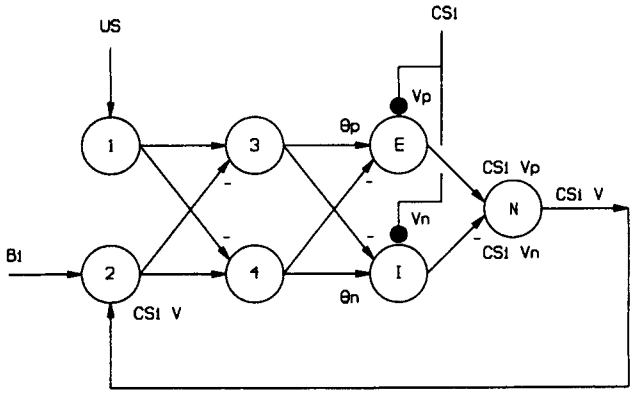

Figure B1. Neural implementation of excitatory (E) and inhibitory (I) associations using exclusively positive plastic connections. (US = unconditioned stimulus; CS_i = conditioned stimulus; B_i = aggregate prediction of the US; V = net CS_i–US association; V_p = positive excitatory CS_i–US association; V_n = positive inhibitory CS_i–US association; θ_p = theta population active when $US > CS_iV + B_i$; θ_n = theta population active when $US < CS_iV + B_i$. Arrows represent fixed synapses. Solid circles represent variable synapses.)

Appendix C

Modulation of Hippocampal Excitability by Medial Septal Inputs

This appendix derives an equation that describes the interaction between entorhinal and medial septal inputs to the hippocampus. Medial septal activity has a "modulatory" effect on dentate granule cells, as well as on CA1 and CA3 pyramidal cells. This modulation might be mediated by a GABAergic inhibition of inhibitory basket cell interneurons. Figure C1 shows a diagram of a neuronal arrangement in which a medial septal input inhibits the recurrent inhibition provided by a basket cell.

Using Grossberg's (1975) notation, entorhinal cortex input excites pyramidal and granule cells according to

$$d(\text{pyr}_i) = -K_{10}\text{pyr}_i + K_{11}\text{ent}_i - K_{12}\text{bas}_i\text{pyr}_i, \quad \text{(C1)}$$

where pyr_i is the activity of pyramidal (or granule) cell i, ent_i is the excitatory input from the entorhinal cortex, $\text{ent}_i = \text{an}_i\text{VN}_i$, and bas_i represents the inhibitory activity of basket cell i.

The activity of basket cell i is given by

$$d(\text{bas}_i) = -K_{13}\text{bas}_i + K_{14}\text{pyr}_i - K_{15}\theta\text{bas}_i, \quad \text{(C2)}$$

where θ represents the activity of the medial septum ($\theta = |\text{US} - B|$), which inhibits basket cell i.

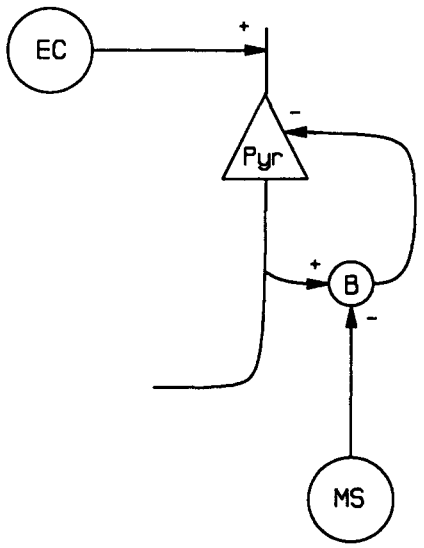

Figure C1. Medial septal modulation of inhibitory feedback of pyramidal cells (one aspect of the hippocampal intrinsic circuitry showing how medial septal inputs inhibit the recurrent inhibition provided by basket cells). (EC = entorhinal cortex; Pyr = pyramidal cell; B = basket cell; MS = medial septum.)

Combining Equations C1 and C2, the asymptotic output of the pyramidal (or granule) cell i is given by

$$\text{pyr}_i \propto \sqrt{\theta\text{ent}_i} \propto \sqrt{|\text{US} - B|\text{an}_i\text{VN}_i}. \quad \text{(C3)}$$

Equation C3 is positive when US > B and negative when US < B. Because the output of the same neuron cannot turn from excitatory to inhibitory, we assume that one neural population codes $\sqrt{(\text{US} - B)}\text{an}_i\text{VN}_i$ when $\theta > 0$ and a different one codes $\sqrt{(B - \text{US})}\text{an}_i\text{VN}_i$ when $\theta < 0$ (see Appendix B).

Because Equation C3 grows without limit with increasing values of $|\text{US} - B|$ we replace it with an Equation A10, which is bounded between −1 and 1. Figure C2 shows that the description of the activity of a pyramidal cell yielded by Equation C3 is well approximated by the hidden-unit error function, EH_j, generated by Equation A10. In both cases the output of the pyramidal cells, assumed to control cortical learning, is proportional to the product of the theta rhythm and the association of the hidden unit with the US.

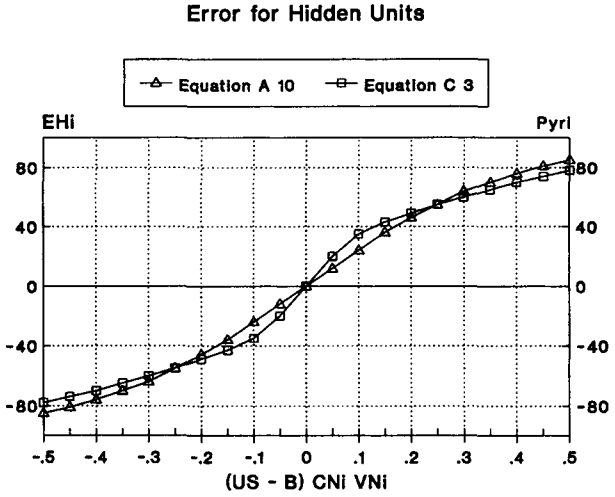

Figure C2. Comparison between the equation describing error to the hidden units used in the simulations and the equation describing pyramidal cell activity derived for the circuit shown in Figure C1. (The activity of a pyramidal cell [Pyr_i] yielded by Equation C3 is well approximated by the hidden-unit error function [EH_i] generated by Equation A10. In both cases the output of the pyramidal cells, which are assumed to control cortical learning, is proportional to the product of the theta rhythm and the association of hidden unit i with the US [unconditioned stimulus]. B = aggregate prediction of the US; CN_i = configural stimulus; VN_i = CN–US association.)

(*Appendix D follows on next page*)

Appendix D

A Formal Description of the Effects of Hippocampal, Cortical, and Cerebellar Lesions

Hippocampal Lesions Effects

Input–Output Associations

In agreement with the aggregate prediction hypothesis, after HL, changes in input–output associations, VS_i, are given by

$$d(VS_i)/dt = K_6 as_i(1 - |VS_i|)EO_i. \quad (D1)$$

The output error EO_i in Equation D1 is given by

$$EO_i = US - a_i VS_i, \quad (D2)$$

where the aggregate prediction B_i is no longer present.

Hidden-Unit–Output Associations

In agreement with the aggregate prediction hypothesis, after HL, changes in hidden-unit–output associations, VN_j, are given by

$$d(VN_j)/dt = K_6 an_j(1 - |VN_j|)EO_j, \quad (D3)$$

The output error EO_j in Equation D3 is given by

$$EO_j = US - a_j VN_j. \quad (D4)$$

Input–Hidden-Unit Associations

After HL, changes in input–hidden-unit associations, VH_{ij}, are given by

$$d(VH_{ij})/dt = K_7 as_i(1 - |VH_{ij}|)EH_j. \quad (D5)$$

The hidden-unit error EH_j is given by

$$EH_j = 0, \quad (D6)$$

and therefore $d(VH_{ij})/dt = 0$.

Cortical Lesions Effects

After CL, VH_{ij} associations are zero.

Cerebellar Lesions Effects

After cerebellar lesions, VS_i and VN_j associations are zero.

Appendix E

Properties of the Hidden-Unit Layer

Number of Hidden Units

Figure E1 shows the effect of varying the number of hidden units on (a) the percentage of times that negative patterning is attained and (b) the average number of trials needed to achieve negative patterning. The criterion for negative patterning has been defined as at least 80% responding to CS_1 or CS_2 and less than 30% responding to CS_1–CS_2. Notice that, even when the exclusive-or problem (and therefore negative patterning) can be solved in a non-real-time model with only one

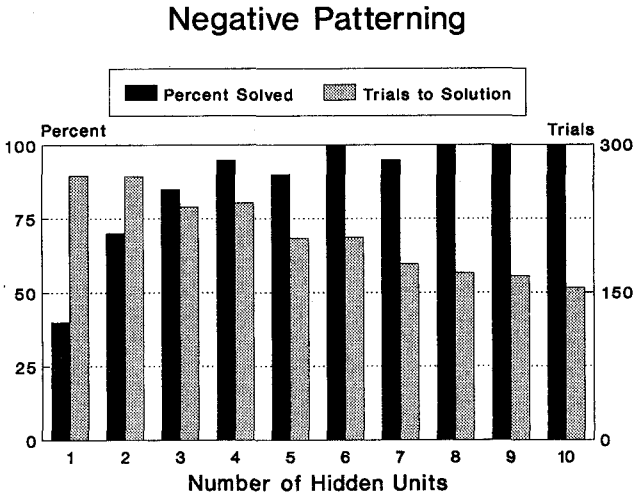

Figure E1. Performance of the S-D (Schmajuk-DiCarlo) model as a function of the number of hidden units. (Percentage of cases and average number of trials in which the system reached criterion in a negative patterning paradigm as a function of the number of hidden units incorporated into the system. Average trials to solution were computed only when solutions were reached in 300 trials or less.)

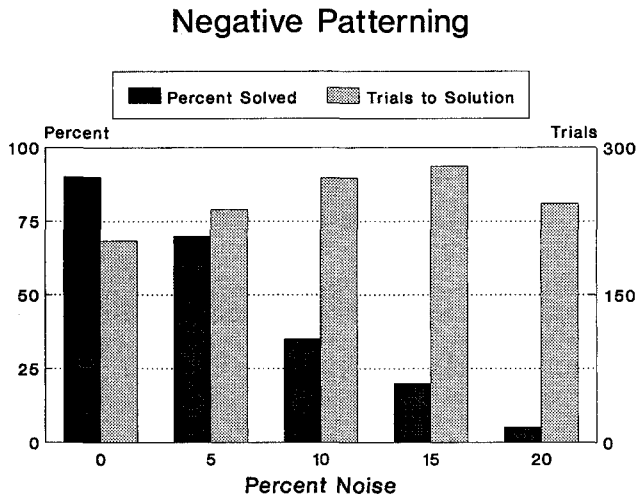

Figure E2. Performance of the S-D (Schmajuk-DiCarlo) model as a function of noise in the error signal. (Percentage of cases and average number of trials in which the system reached criterion in a negative patterning paradigm as a function of the noise added to the error signal used to train the hidden units. Percentage of noise was defined in relation to the range of possible error values [±0.5]. Average trials to solution were computed only when solutions were reached in 300 trials or less.)

hidden unit (Rumelhart et al., 1986, p. 321), in our model at least four units are required to solve the problem more than 75% of the time.

The results shown in Figure E1 are reminiscent of Lashley's (1950) principle of mass action. Lashley (1950) trained rats on complex maze tasks in which they made use of information from different sensory modalities (visual, somatosensory, auditory). After training, rats received different amounts of CL. Rats with extensive CL performed worse than rats with restricted CL, independently of the location of the lesion. This "mass action" effect is captured by the model, which predicts increasingly deficient performance with increasingly larger CL, that is, a smaller number of "cortical" hidden units.

Noise in Error Function for the Hidden Units

Medial septal activity has a "modulatory" effect on dentate granule cells, as well as on CA1 and CA3 pyramidal cells. In Appendix C we derived the asymptotic output of pyramidal and granule cells and showed how this activity is well approximated by the hidden-unit error function generated by Equation A10. In both cases the output of the pyramidal cells, assumed to control cortical learning, is proportional to the product of the theta rhythm and the association of the hidden unit with the US.

Because neural computations are intrinsically noisy, we studied the effect of adding white noise into the error function generated by Equation A10. Presumably, this noise represents the neural noise of pyramidal, granule, and basket cells. Figure E2 shows the percentage of cases and average number of trials in which the system reached criterion in a negative patterning paradigm, as a function of the noise added to the error signal used to train the hidden units. Solutions were defined as at least 80% responding to CS_1 and CS_2 and less than 30% responding to CS_1–CS_2, for a maximum of 300 trials. Although the number of trials to reach the solution is relatively constant (and dependent on the number of hidden units as just shown), the performance of the system degrades rapidly with noise greater than 5%.

Appendix F

Simulation Parameters

Our simulations assume that one simulated time step is equivalent to 10 ms. Each trial consists of 500 steps, equivalent to 5 s. Unless specified, the simulations assume 200-ms CSs, the last 50 ms of which overlap the US. CS onset is at 200 ms. Parameters are selected so that simulated asymptotic values of CR are reached in around 10 acquisition trials. Because asymptotic conditioned NM responding is reached in approximately 200 real trials (Gormezano, Kehoe, & Marshall, 1983), one simulated trial is approximately equivalent to 20 experimental trials.

Parameter values are $K_1 = 0.07$, $K_2 = 0.18$, $K_3 = 1$, $K_4 = 2.5$, $K_5 = 1.5$, $K_6 = 0.005$, $K_7 = 0.33$, $K_8 = 5$, $K_9 = 0.03$, $\beta_1 = .5$, $n = 1.5$. These parameter values are kept constant for all simulations for normal, HL, and CL cases. The initial values of VS_i and VN_j were 0. Because five hidden units are needed in order to attain reliable negative patterning (see Appendix E), this number of hidden units was used in all simulations. Input–hidden-unit association weights, VH_{ij}, were randomly assigned using a uniform distribution ranging between $\pm$ 0.25 in all simulations. Generalization factors are $\gamma_{\pm 1} = 0.9$, $\gamma_{\pm 2} = 0.6$, $\gamma_{\pm 3} = 0.3$, $\gamma_{\pm 4} = 0.15$, $\gamma_{\pm 5} = 0$, where subscripts indicate the number of arbitrary steps separating different frequencies from the training frequency.

Received October 12, 1990
Revision received July 3, 1991
Accepted July 18, 1991 ■