Pergamon

0306-4522(93)E0013-G

Neuroscience Vol. 59, No. 2, pp. 229–243, 1994
Elsevier Science Ltd
IBRO
Printed in Great Britain
0306-4522/94 $6.00 + 0.00

VALUE-DEPENDENT SELECTION IN THE BRAIN: SIMULATION IN A SYNTHETIC NEURAL MODEL

K. J. FRISTON,* G. TONONI,* G. N. REEKE JR,† O. SPORNS* and G. M. EDELMAN*‡

*The Neurosciences Institute, Suite 10, 3377 North Torrey Pines Court, La Jolla, CA 92037, U.S.A.
†The Rockefeller University, New York, NY 10021, U.S.A.

Abstract—Many forms of learning depend on the ability of an organism to sense and react to the adaptive value of its behavior. Such value, if reflected in the activity of specific neural structures (neural value systems), can selectively increase the probability of adaptive behaviors by modulating synaptic changes in the circuits relevant to those behaviors. Neuromodulatory systems in the brain are well suited to carry out this process since they respond to evolutionarily important cues (innate value), broadcast their responses to widely distributed areas of the brain through diffuse projections, and release substances that can modulate changes in synaptic strength.

The main aim of this paper is to show that, if value-dependent modulation is extended to the inputs of neural value systems themselves, initially neutral cues can acquire value. This process has important implications for the acquisition of behavioral sequences. We have used a synthetic neural model to illustrate value-dependent acquisition of a simple foveation response to a visual stimulus. We then examine the improvement that ensues when the connections to the value system are themselves plastic and thus become able to mediate acquired value. Using a second-order conditioning paradigm, we demonstrate that auditory discrimination can occur in the model in the absence of direct positive reinforcement and even in the presence of slight negative reinforcement. The discriminative responses are accompanied by value-dependent plasticity of receptive fields, as reflected in the selective augmentation of unit responses to valuable sensory cues. We then consider the time-course during learning of the responses of the value system and the transfer of these responses from one sensory modality to another. Finally, we discuss the relation of value-dependent learning to models of reinforcement learning. The results obtained from these simulations can be directly related to various reported experimental findings and provide additional support for the application of selectional principles to the analysis of brain and behavior.

Evolution has endowed certain organisms with several means to sense the adaptive value of their behavior. According to the theory of neuronal group selection,[11,12,14] evolutionarily selected value systems modulate synaptic changes in multiple brain regions to provide various constraints for the selection of adaptive behaviors in somatic time. In this article, we use a synthetic neural model to extend our previous work on value and value systems as they relate to the brain.[12,38,48] Our main goal here is to address in detail how value systems themselves can be modified and extended by experience.

The central idea of the theory of neuronal group selection is that selective processes operate in the nervous systems of individuals to enhance adaptive behavior despite the absence of predetermined categories and fixed rewards in the environment. The main principles governing these somatic selective processes are conceptually similar to those that operate in evolution, but their substrate (developmentally established repertoires of interconnected neuronal groups) and basic mechanisms (modification of synaptic strengths) differ from those of evolution. Specifically, the theory proposes that brain function is mediated by: (i) selectional events occurring among interacting cells in the developing embryo to form large repertoires of variant neural circuits; (ii) further selectional events occurring among populations of synapses to enhance those neuronal responses having adaptive value for the organism; and (iii) re-entrant signals, exchanged via parallel and reciprocal connections, that serve through synaptic selection to integrate response patterns among functionally segregated brain areas in an adaptive fashion. These processes are said to be sufficient to account for a variety of brain functions ranging from perception to intricate motor responses.[14]

Inasmuch as somatic selectional systems do not operate according to a predefined program or syntax, they must be constrained by evolutionarily selected biases (innate values) incorporated in the phenotype. While a full discussion of the concept of values is beyond the scope of this paper, some crucial properties of candidate value systems are considered here in detail. In this paper, we use the word "value" with

‡To whom correspondence should be addressed.

Abbreviations: **AI**, auditory area; **ACe**, central nucleus of the amygdala; **Ain**, auditory input; CS, conditioned stimulus; CR, conditioned response; **LHA**, lateral hypothalamic area; **SC**, oculomotor map; TD, temporal difference; US, unconditioned stimulus; VAL, value system; **Vin**, visual input; **V1**, visual area.

reference to neuronal responses in the following sense. The value of a global pattern of neuronal responses to a particular environmental situation (stimulus) is reflected in the capacity of that response pattern to increase the likelihood that it will recur in the same context. In this respect, value is analogous to "adaptive fitness" in evolutionary selection, where the adaptive fitness of a phenotype is defined in terms of its propensity to be represented in subsequent generations. Thus, value plays a role in neuronal selection similar to that which adaptive fitness plays in evolutionary selection. Inasmuch as value systems themselves are subject to evolutionary constraints, the relationship between value and adaptive fitness is complex. Value is subject to the overall constraint that it must, *ex post facto*, act to increase adaptive fitness. Although evolutionary processes cannot select for valuable neuronal responses in somatic time, they can select for mechanisms that subserve such neuronal selection. In this paper, we discuss the relationship between value and adaptive fitness specifically in terms of the interaction between acquired and innate value.

We propose that the increased probability of valuable neuronal responses is mediated by particular structures in the nervous system that we call "value systems," which operate through selective consolidation of synaptic changes. The value of a neuronal event can be operationally defined in terms of the activity it effectively evokes in such value systems. For neural value systems to constrain somatic selection, they should possess a number of structural and functional properties. They should be responsive to evolutionarily or experientially salient cues. They should broadcast their responses to wide areas of the brain and release substances that can modulate changes in synaptic strength. In addition, value systems should be capable of a transient response to sustained input, inasmuch as it is changes in circumstances (environmental or phenotypic) that are important for successful adaptation. There is substantial evidence[16,24,27-29,35,50,51] to indicate that the aminergic and cholinergic neuromodulatory systems possess such properties.

In our previous theoretical work,[13,39,48] value took the form of a global signal that modulated changes in synaptic strength to reinforce adaptive behaviors. It was assumed that the neural systems subserving value had been selected during evolution to signal autonomic consequences of behaviors relevant for the homeostasis of the organism. In these simulations, the sensory inputs eliciting value were fixed; i.e. the modeled value systems specified only innate value. In the present paper, we present a theoretical analysis of how value itself may be acquired. We hypothesize that acquired value arises from value-dependent and experience-dependent plasticity in the afferents to value systems themselves. As a result, whenever an adaptive behavior is acquired through value-dependent modulation of synaptic changes, certain neuronal activity patterns that reliably precede this behavior become themselves capable of eliciting value. In this way, such activity patterns can reinforce or stabilize other antecedent patterns. Through this "bootstrap into the past", successive patterns of neuronal activity can be linked together and assembled into complicated, adaptive behavioral sequences.

Using a synthetic neural model,[39] we explore the role of innate and acquired value in the acquisition of adaptive and convergent behavior and extend our previous work on visual tracking[13,39] and operant conditioning in the context of visual integration.[48] We simulate a simple organism having neural circuits constituting a visual area, an auditory area, and oculomotor connections and explore foveation of a visual stimulus and the acquisition of discriminative eye movements to different auditory tones. After experience, the simulated organism acquired foveation through value-dependent plasticity in sensorimotor maps. Addition of value-dependent plasticity in the connections from the sensorimotor maps to the value system itself was shown to significantly improve behavioral performance. This plasticity also allowed learning of a simulated auditory discrimination task when a visual stimulus was used as a secondary reinforcer, even when the visual stimulus proper did not elicit any intrinsic or innate value. On the basis of these results, we then examined the transfer of value system responses between stimuli during learning. (Although we use the term "learning" for the acquisition of simple behaviors in the model, true learning involves mechanisms and interactions at all levels of the system.) In interpreting the results, we make some experimental predictions, consider brain structures and transmitter systems that could mediate value-dependent learning, and review our findings in the light of comparable experiments in animals. Finally, using a formal analysis, we relate value-dependent learning to temporal difference models of reinforcement learning.

SIMULATIONS

The synthetic neural models and simulations were chosen to provide a clear illustration of how value systems and value-dependent learning might be implemented in the brain. They were also designed to relate various theoretical predictions to findings in the experimental literature. To distinguish between real brain areas and simulated areas, the names of the latter appear in bold characters.

The model

A two-dimensional visual input (**Vin**, a model retina 16×16 pixels in size, where 1 pixel corresponds to 1° of visual angle) was relayed to a visual area (**V1**, Fig. 1) consisting of 16×16 units representing local neuronal groups rather than single neurons. An auditory area (**A1**) received an ordered mapping from a one-dimensional input (**Ain**), which rep-

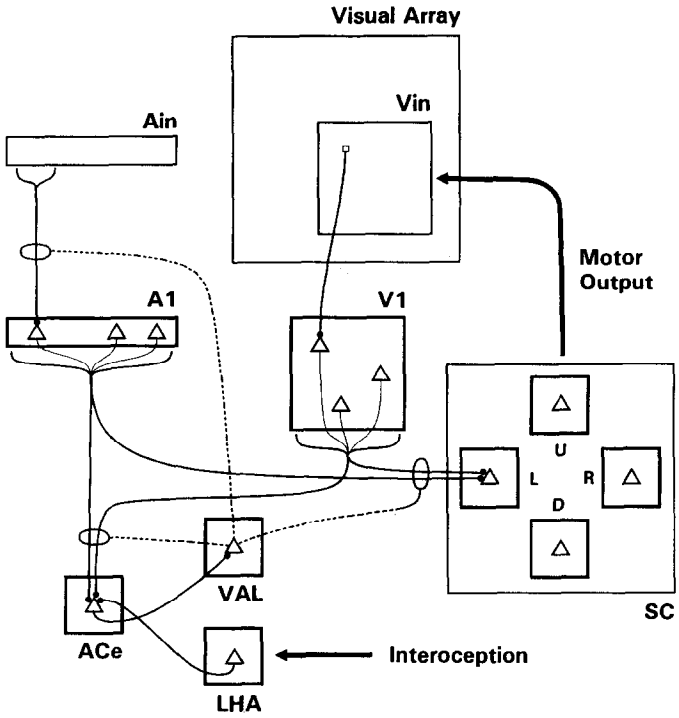

Fig. 1. Schematic layout of simulated areas and connections. Thin boxes enclose sensory inputs and motor outputs; heavier boxes enclose neural areas. Triangles represent neurons or groups of neurons; solid lines ending in filled circles represent tracts of excitatory connections and their synaptic terminals. Dashed lines represent efferents from **VAL** responsible for modulating changes in strength of encircled connections. The visual array corresponds to a bounded visual scene (32 × 32 pixels in size). The visual input area, **Vin**, is a model retina which receives stimulation from a (16 × 16) portion of the full visual array. This sampling of the visual scene changes in accord with simulated oculomotor output (from **SC**). Heavy arrow (upper right) indicates pathway by which motor cells in **SC** (U = up, R = right, D = down, L = left) cause **Vin** to move. Heavy arrow (bottom) indicates pathway by which simulated inputs triggering innate values excite area **LHA**. See text for names of areas and Table 1 for other details.

resented auditory input in frequency space. Receptive fields of the 16 **A1** units were initially Gaussian with a full width at half maximum of 3.29 units and a maximum response of 0.6 to the preferred frequency presented with unit intensity. All visual (**V1**) and auditory (**A1**) units projected to a simple motor map (**SC**) responsible for generating horizontal and vertical eye movements.[39] Behaviors leading to innately valuable changes elicited activity in a unit called **LHA** designed to represent the lateral hypothalamic area or equivalent nuclei. These behaviors can be thought of as fixed action patterns[7] emitted in response to a releasing stimulus.[6,20] Alternatively, in an experimental setting, these behaviors are equivalent to unconditioned responses to unconditioned stimuli (US; e.g., food or juice rewards). The areas **LHA**, **V1**, and **A1** sent efferents to a structure called **ACe**, corresponding to the central nucleus of the amygdala. **ACe**, which acted as a site of convergence for both innate and potentially acquired values, in turn projected to a unit called **VAL**, whose activity was able to affect the plasticity of all the connections in the simulated brain. The **VAL** unit can be thought of as modelling the activity of cells of origin of the cholinergic system (substantia innominata, nucleus basalis of Meynert), or of the meso-corticolimbic dopaminergic system (ventral tegmental area and nucleus accumbens). See Fig. 1 for further details.

In the actual simulations, the model was tested in two stages. The first stage proceeded along the lines of our previous work[13,39] and addressed the role of innate and acquired value in the acquisition of an orienting response to a spot of light presented in the periphery of the visual field. If successful foveation occurred within 2° of the center of the stimulus, **LHA** was activated. Shortly after foveation, the spot disappeared and then reappeared in the periphery of vision. The second stage was explicitly designed to model a number of relevant second order operant conditioning experiments in rats and non-human primates.[8,18,28,40] This involved presenting a simulated high, middle, or low frequency tone for 16 iterations. If, by the time the tone was over, a discriminative oculomotor response (moving the eye upwards for high tones and downwards for low tones) had occurred, the visual stimulus appeared in the periphery of vision and could then be foveated to obtain the primary reward. No response, or an incorrect response, resulted in a new trial in which tones were presented, which began after a short inter-trial

Table 1. Stimulation parameters

Area	Number of units	Dynamics α mean, S.D.	ω	μ	Functional description
Vin	512 (16 × 16)	0, 0	0	0	Visual input (1 unit = 1°)
Ain	16	0, 0	0	0	Auditory input
V1	512 (16 × 16)	0, 0	0	0.01	Visual retinotopic map
A1	16	0, 0	0	0.01	Auditory tonotopic map
SC	4	0, 0.04	0.4	0.06	Oculomotor map
LHA	1	0, 0	0.92	0	Inputs with innate value
ACe	1	0.5,0	0	0	Limbic structure
VAL	1	0, 0	0	0	Diffuse ascending system

Connection	Number per unit	δ	η	Initial c_{ij}	g_k	Connectivity Description
Vin → V1	1	0	0.1	1	1	Retinotopic mapping
Ain → A1	16	1	0.1	0.6	1	Tonotopic mapping (FWHM = 3.29)
V1 → SC	256	1	0.1	0.2	0.1	Complete and non-ordered
A1 → SC	16	1	0.1	0.2	0.1	Complete and non-ordered
V1 → ACe	256	0.2	0.1	0.01	0.1	Complete and non-ordered
A1 → ACe	16	0.2	0.1	0.01	0.1	Complete and non-ordered
LHA → ACe	1	0	0.1	1	0.16	No plasticity
ACe → VAL	1	0	0.1	1	1	No plasticity

interval. Finally, to test its robustness, the correct discriminative response was confronted with negative reinforcement (simulated by using negative value; see below), and the second stage was repeated.

Dynamics

We used Cortical Network Simulator[39] to simulate the neuronal system. Each unit was taken to correspond to a neuronal group of hundreds to thousands of densely interconnected neurons,[11,12] and each iteration corresponded to about 100 ms of simulated time. The response (s_i) of each unit (i) to its inputs was calculated as:

$$s_i(t+1) = \psi\{\Sigma g_k c_{ij} \cdot s_j(t) + \alpha_i(t) + \omega s_i(t)\} \cdot \sigma\{D_i(t)\}$$

$$D_i(t+1) - D_i(t) = \mu[s_i(t) - D_i(t)]. \quad \text{(Eqn 1.1)}$$

s_j is the activity of unit j connected to unit i with connection strength, c_{ij} and g_k is a constant, common to all connections between one area and another. k is a subscript that identifies the set of all connections between any two areas. α_i is spontaneous activity or noise—an independent random number uncorrelated over time which is selected for each unit from a Gaussian distribution with a constant mean and variance in a given area (see Table 1). ω is a coefficient of persistence which is a constant for all cells in a given area—see Table 1. $\psi\{\cdot\}$ is a piecewise linear approximation to an increasing sigmoidal function that limits s_i to the range [0, 1] (this approximation was chosen simply for computational expediency). $\sigma\{\cdot\}$ is a polynomial approximation to a decreasing sigmoidal function of the form $\sigma\{x\} = 1 - 2x^2 + x^4$ when $0 < x < 1$ and 1 ($x < 0$) or 0 ($x > 1$) otherwise. D_i is a depression term that simulates adaptation during sustained periods of activity. The rate of adaptation is determined by μ, which is constant for all cells in a given area. See Table 1 for values of all parameters used in the simulations.

The sensory units (**Vin**, **Ain**, and **LHA**) all responded according to Eqn 1.1 except that the afferent input term $\Sigma g_k \cdot c_{ij} \cdot s_j(t)$ was simply replaced by an appropriate sensory input, with values in the range [0, 1]. While inputs to **Vin** and **Ain** represented visual and auditory sensory input, respectively, the activity of **LHA** (s_{LHA}) was designed to simulate responses to signals that would result from a reward following a certain behavior. Whenever such behavior was emitted in the model, s_{LHA} was set to unity. s_{LHA} decay was set so that it would fall to negligible levels after about 60 iterations or 6 s of simulated time (i.e. $\omega_{LHA} = 0.92$ in Eqn 1.1 giving a half-life $t_{1/2} = 8.66$ iterations). The latency of actually evoked LHA responses from mechanoreceptor stimulation in the proximal stomach was found to be about 370 ms.[52]

The activity of **VAL** (modeling a simple value system) reflected the change in afferent input from **ACe**.

$$s_{VAL}(t) = s_{ACe}(t) - s_{ACe}(t-1). \quad \text{(Eqn 1.2)}$$

This time derivative of **ACe** activity was meant to emulate the phasic responses of dopaminergic and cholinergic neurons to external stimuli (e.g., those predicting appetitive reward); such responses are phasic and transient, with time-courses of the order of 100–200 ms.[10,28,40] It should be noted that both the response dynamics of the **VAL** unit and the postsynaptic effects of **VAL** activity (viz. the modulation of changes in synaptic strength in other model areas) are specific for this unit. They may be thought of as the result of evolutionary adaptations giving rise to neural value systems with such properties.

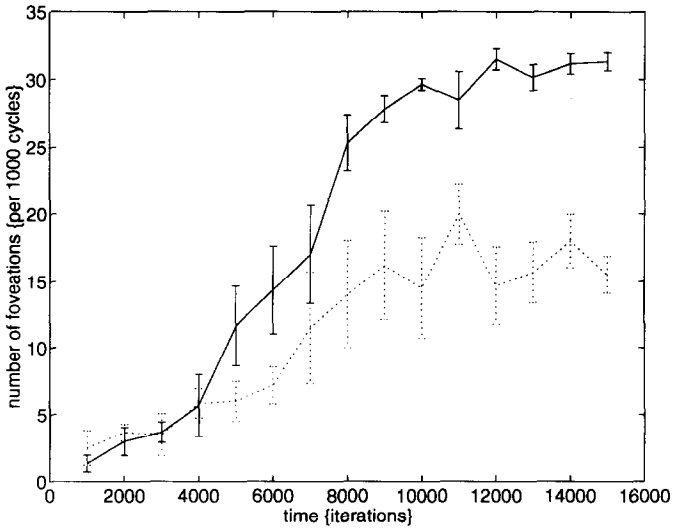

Fig. 2. "Learning" curves for acquisition of the foveation response to a visual stimulus. Performance is expressed as the mean and standard error (over six runs) of the number of foveations per 1000 iterations of the neural model (100 s of simulated time). The solid curve is for the intact system with adaptive or acquired value. The dashed curve was obtained with **V1** disconnected from **ACe**. Convergence is essentially complete after about 10,000 iterations.

Value-dependent changes in synaptic strength took the following form:

$$h_{ij}(t+1) - h_{ij}(t) = \delta_k \cdot \sigma\{c_{ij}(t)\} \cdot s_i(t) \cdot s_j(t) - \eta_k h_{ij}(t),$$

and

$$c_{ij}(t+1) - c_{ij}(t) = s_{VAL} \cdot h_{ij}(t+1). \quad \text{(Eqn 1.3)}$$

h_{ij} is an associative term that represents a trace of the product of pre- and postsynaptic activity. $\sigma\{\cdot\}$ is the same sigmoid function as in Eqn 1.1. δ_k is a parameter controlling the rate of synaptic change and η_k is the decay rate, where k again denotes all the connections from one area to another. The connections given the greatest plasticity were: (i) those mediating sensorimotor integration (**V1 → SC, Ain → A1, A1 → SC**)

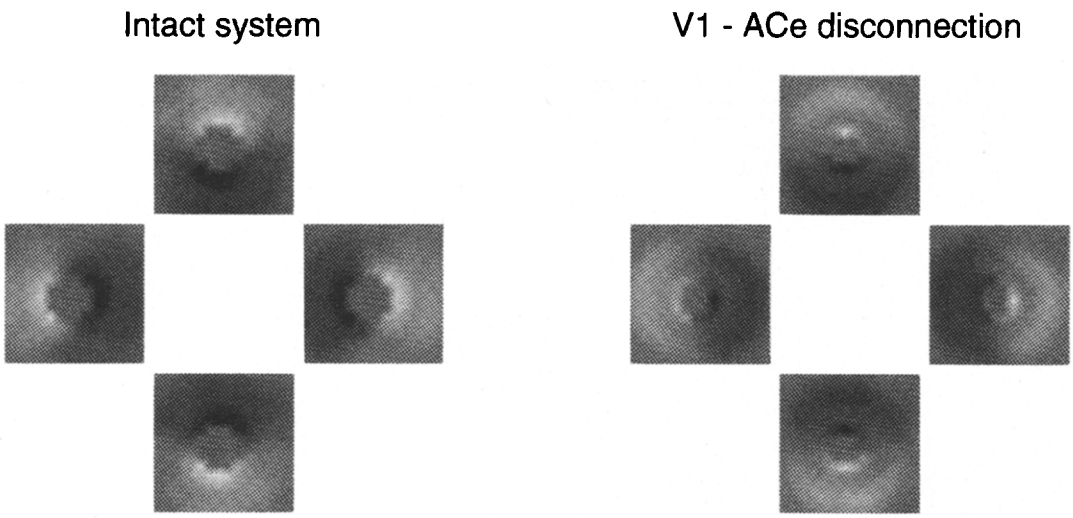

Fig. 3. Connection strengths (averaged over six runs) between the visual system (**V1**) and **SC** units after 15,000 training iterations. Connections within each box, each represented by a single pixel, are mapped according to the location of the source of the connection in **V1**. Top box in each array displays connection strengths to up unit minus connections strengths to down unit, bottom box displays complementary connections strengths to down unit minus those to up unit, and similarly for left and right units. Connection strength differences are displayed on a gray scale in which positive differences are light and negative differences are dark. Left array: intact system. Right array: **V1 → ACe** disconnected during training. The ordered gradient-like mappings result in appropriate saccade-like movements according to the position of the stimulus in retinotopic space.

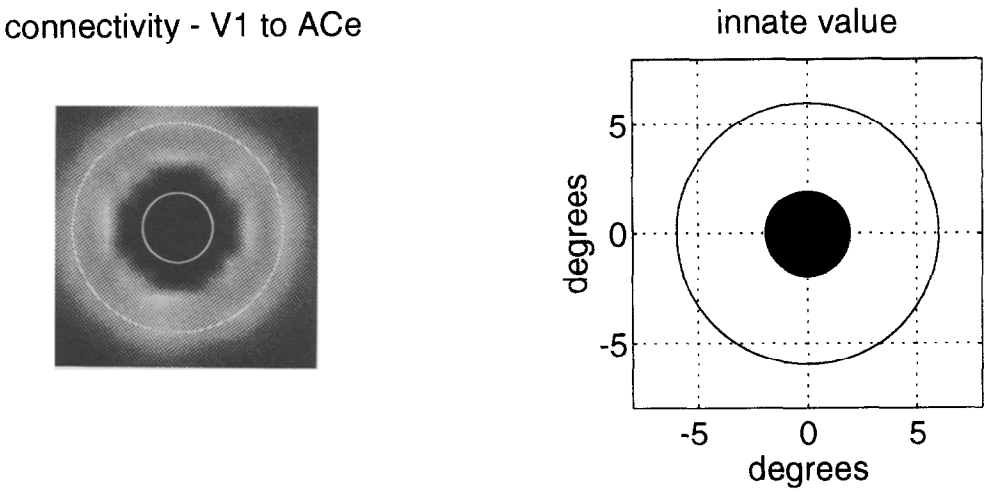

Fig. 4. Left: connection strengths (averaged over six runs) established by value-dependent learning from visual area **V1** to **ACe** using the conventions of Fig. 3. These connections define which retinotopic positions have acquired the potential to elicit value. **Right**: black dot indicates positions in retinotopic space that are associated with a priori or innate value. Note that learned value does not develop in the central region associated with innate value. The circles in both diagrams are at 2° (delineating the region where innate value is present) and 6° (indicating the initial positions of visual targets).

and (ii) those mediating acquired value (**V1** → **ACe**, **A1** → **ACe**). Connections defining innate value (**LHA** → **ACe** and **ACe** → **VAL**) were not plastic. Table 1 contains the actual parameters used.

In each cycle of the simulation, the variables were computed in the order in which the equations are presented above. First the s_i including s_{VAL} were updated synchronously, and then the new depression term (D_i) was computed using Eqns 1.1 and 1.2. Following this, h_{ij} was updated and then c_{ij} using Eqn 1.3.

EXPERIMENTS AND RESULTS

Stage 1: *roles of innate and acquired value*

Acquisition of foveation behavior. These simulations used a circular visual stimulus with a Gaussian luminance profile (2.35° full width at half maximum). Whenever the initially random, spontaneous stochastic activity of SC units caused foveation to within 2° or less, s_{LHA} was set to unity and eight iterations later the stimulus was removed. Following an inter-trial interval of eight iterations, the stimulus was again

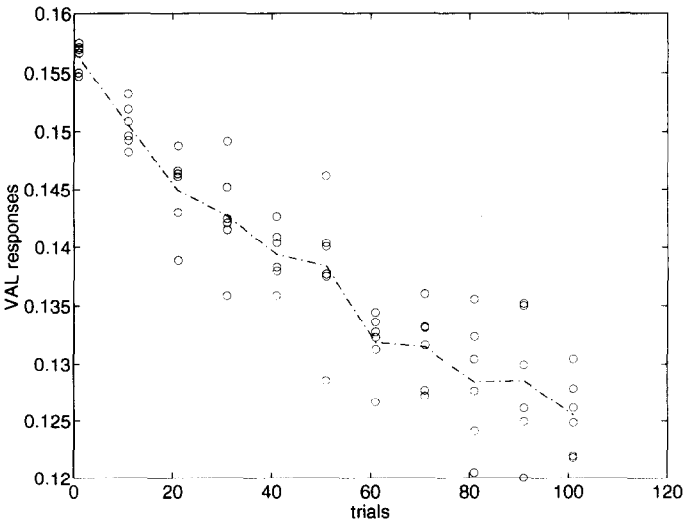

Fig. 5. Adaptation in the activity level of the value unit, s_{VAL}, designated VAL responses, at the time of foveation. Note that these data are expressed as a function of trials, rather than iterations. The iterations (time) required for each trial decreases with learning as fewer iterations of the model are required for foveation to occur. Individual data points are from six separate runs; each point represents the mean response over 10 consecutive trials. The decline in s_{VAL} illustrates the learning-dependent decrease of the value response.

presented at a random location 6° from the center of retinotopic space. Eye movements were scaled such that maximal activity in a **SC** unit resulted in an angular velocity of 1° per iteration, in the appropriate direction. This meant that no movement could foveate a peripheral stimulus in a single iteration. Despite this constraint, and despite the fact that only those visual stimuli located within the central 2° elicited innate value, the frequency of foveations increased rapidly with the emergence of serial eye movements that brought the stimulus progressively closer to the fovea. Figure 2 shows the improvement of performance expressed as the number of foveations per 1000 iterations. These data were obtained from six runs of 15,000 iterations each. To obtain the lower curve in Fig. 2, **ACe** was disconnected from **V1** and thus the curve reflected the action of innate value alone. The quantitative improvement with innate plus acquired value over innate value alone is clearly evident. Nevertheless, the results also show that, under certain circumstances, innate value is sufficient for some degree of adaptive behavior. As shown in stage 2 below, in certain experimental paradigms there is a more profound dissociation in the qualitative aspects of acquisition of such behavior with and without acquired value.

Value-dependent plasticity in sensorimotor mappings. Successful foveation of an arbitrarily positioned stimulus requires the formation of an ordered sensorimotor mapping under the constraint of value. In the model, this requires functional specialization of **SC** units with respect to luminance contrast and retinotopic position. The pattern of connection strengths from **V1** to **SC** units that emerges during value-dependent learning is presented in Fig. 3. It is this change that mediates the adaptive behavior depicted in Fig. 2 and it ensures that the output of **SC** units is a nonlinear but monotonic function of stimulus position. After we disconnected **V1** and **ACe**, changes in connection strengths were smaller and were limited to the immediate pericentral region.

Value-dependent plasticity of the connections to the value system itself (acquired value). In the model, the acquisition of value depends on value-dependent associative changes in afferents to the value system itself. These changes are shown in Fig. 4. Connections from **V1** to **ACe** are progressively enhanced, first around the fovea and then in the periphery. Through these enhanced connections, stimulus positions that elicit saccade-like movements to the center, or to retinotopic locations with established connections to **ACe** (those that have already acquired value), come themselves to activate **ACe** and thus they acquire the potential to elicit value. As the activity of **VAL** depends on an increase in **ACe** activity, value-dependent modulation of plasticity is greatest when an eye movement trajectory passes from a position that has no **V1** → **ACe** connections to a region that does. For the most part, this is what occurs when an adaptive movement occurs by chance. In this way, value becomes most effective at the point at which movement is incorporated into a learned sequence.

Decrease of value responses during learning. The transfer (see below) of value system responses to earlier components of a behavioral sequence means that late components progressively lose the capacity to elicit value. Figure 5 demonstrates this point: as the acquisition of foveation behavior proceeds, s_{VAL}, determined at the point of foveating the stimulus, decreases. Empirical evidence[28] for the progressive loss of dopaminergic neuron responses is reviewed in the Discussion.

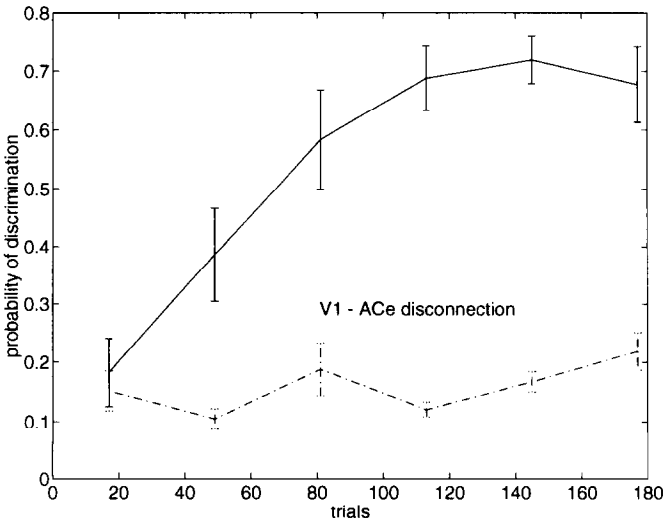

Fig. 6. Discrimination learning using a visual stimulus to reinforce a correct oculomotor response to a simulated high or low tone. Performance is expressed as the mean and standard error (over six runs) of the fraction of correct responses (averaged over 32 consecutive trials of each individual run). Solid line: simulation with intact nervous system. Broken line: **V1** → **ACe** connections cut.

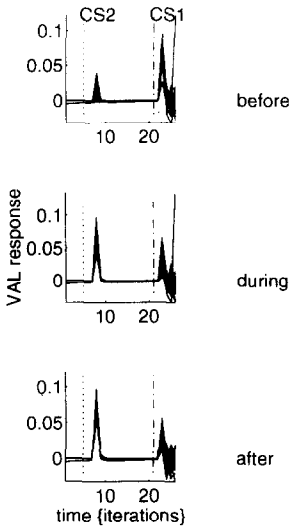

Fig. 7. Transfer of value response from visual stimulus onset to auditory stimulus onset. The activity of the value unit, s_{VAL} (VAL response), is plotted as a function of iterations (time), locked to tone onset in each trial (from eight iterations before tone onset to four iterations after the appearance of the visual stimulus). All data were taken from trials giving a correct discrimination. Before: traces from the first eight correct discriminations; during: traces from eight correct trials halfway through the simulation; after: traces from the last eight trials. CS_2 marks the tone onset and CS_1 marks the appearance of the visual stimulus. The tone progressively acquires value as indicated by the increasing **VAL** response.

Stage 2: Second order conditioning

Learning a discrimination task without direct reinforcement. After the system had acquired foveation behavior, the simulated equivalent of a pure high, middle, or low frequency tone was presented for 16 iterations (1.6 s of simulated time). If, at the end of that time, the oculomotor system had generated an upwards or downwards eye movement through 2° or greater when exposed to high or low tones, respectively ("correct" responses), the visual stimulus appeared at a random position 6° from the center of the fovea. After foveation, or after an incorrect discriminative response, there was an inter-trial interval of eight iterations and a new trial presenting the tones began. This was repeated for 15,000 iterations. Note that the visual stimulus is presented only after a correct discriminative response.

One can consider this task to be a second-order conditioning experiment in learning in which the peripheral spot is the CS_1, foveation is the CR_1, and the tone the CS_2 which cues a discriminative eye movement (CR_2). The results show that, by virtue of its associated acquired value, the peripheral visual stimulus was able to reinforce a correct auditory discrimination (despite the fact that initially eye movements were emitted by chance). The learning curves depicting probability of a correct discrimination as a function of trials for the intact system and after disconnection of **ACe** from **V1** are shown in Fig. 6. Clearly, disconnecting **ACe** from **V1** eliminated both acquired value and discrimination learning. The results of this simulated lesion study are similar to the experimental findings of Gaffan and Harrison[18] reviewed below.

Transfer of value responses during learning. During discrimination learning, we observed the transfer of value-system responses from the CS_1 (appearance of the visual stimulus) to the CS_2 (tones) that predicted the CS_1 and acquired the capacity to elicit discrimina-

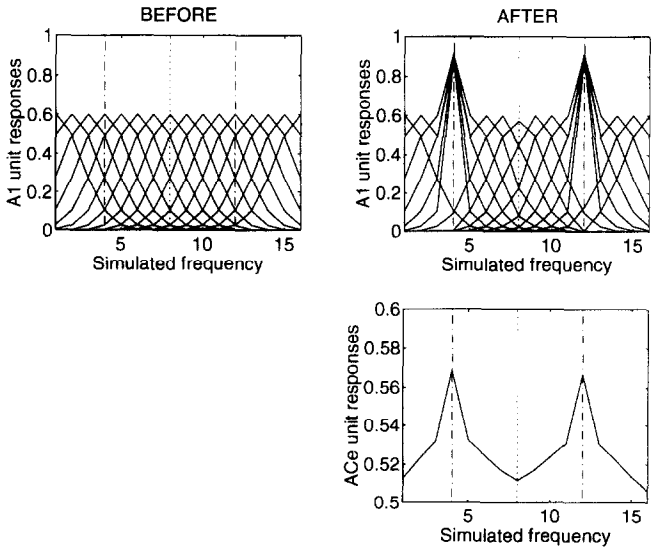

Fig. 8. Top: receptive field plasticity of **A1** units expressed as strength of response to pure tones of unit amplitude. The curves show shifts, as predicted, in center frequency and peak response from before (left) to after (right) value-dependent learning. The vertical lines indicate the low and high (–·–·–, CS_2) and middle (. . . . , neutral or control) frequencies used in the experiments. Bottom: Equivalent tuning curves for **ACe** showing that only those tones (high and low) which are predictive of value have established significant connections to **ACe**.

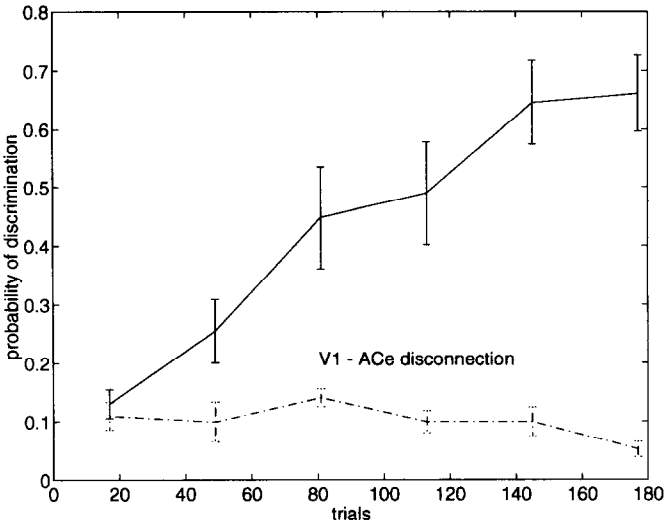

Fig. 9. Discrimination learning (as in Fig. 6) for the intact system (solid line) and with **V1 → ACe** disconnected (broken line) when the correct discriminative response was made mildly aversive. Note that in the lesioned system, the second-order response cannot be acquired, and there is instead a reduction in the probability of the desired discriminative response because of the aversive first-order value.

tive eye movements. Figure 7 shows the activity profile of s_{VAL} before, during, and after discrimination learning. Before learning, s_{VAL} responses are highest at the appearance of the CS_1; after learning the s_{VAL} response has been transferred to the earlier occurrence of the CS_2. The mechanism of this transfer is straightforward: initially the appearance of the visual stimulus causes increased activity in **ACe** and a value system response. Such activity modifies connections (i) **Ain → A1**, (ii) **A1 → SC** and (iii) **A1 → ACe**. This results in: (i) a plastic change in the receptive field properties of **A1** units (see below); (ii) synaptic change in the appropriate auditory–motor connections; and (iii) a potential for the tones to elicit value. By the time the visual stimulus appears, **ACe** has already been excited by the tone and the value (reflected by an increase in s_{ACe}) that is elicited by the visual stimulus is attenuated.

Value-dependent receptive field plasticity. An interesting consequence of value-dependent plasticity in afferents to sensory units in the model is that receptive field properties can change preferentially to sample cues having potential value. This can be shown in the simulations in terms of the tuning curves of **A1** and **ACe** units. In the discrimination learning experiment, both high and low tones are potentially valuable in the sense that they both predict the visual stimulus, given the correct response. As the middle tone does not specify any valuable response, it serves as a control. The results show that **A1** and **ACe** units preferentially respond to either the high or low tones but not to the middle tone. Figure 8 shows the tuning curves of **A1** units before and after learning. Both a shift in tuning curves and a sharpening of frequency selectivity are evident. The equivalent tuning curves for **ACe** demonstrate that only the high and low tones have acquired the potential to elicit value. Receptive field plasticity of this sort has been studied by Weinberger and colleagues in cat auditory cortex.[32,50]

Learning a discrimination task despite negative reinforcement. To test the robustness of these responses, the second stage of simulated training was repeated, but with the discriminative response receiving mild negative reinforcement. This was reflected in negative value: if discrimination occurred, **ACe** received an input that mirrored **LHA** input but was negative in sign (s_{LHA} set to -0.3, exponential decay $t_{1/2} = 4.95$ iterations). This value was chosen to be substantially less than the input to **ACe** on foveating the spot ($s_{LHA} = 1.0$) but not to be trivially low. Under this paradigm, the system could in principle show two kinds of behavior. It could (i) passively avoid an aversive discriminative response and thus forgo the potential value of foveating the visual stimulus or, (ii) perform the auditory discrimination despite its temporary aversive effect in order to get to the potentially valuable visual stimulus. The actual solution obtained depends on acquired value: aversion after discrimination is easily offset by the acquired value of a peripheral visual stimulus and learning proceeds normally, if a little slowly. If acquired value is abolished and **ACe** is disconnected from **V1**, this is not possible and what is learned is determined by the immediate (and innately specified) consequences of an action. Figure 9 demonstrates this dissociation by comparing discrimination learning with and without **V1 → ACe** disconnection.

DISCUSSION

The present simulations have been concerned with several important aspects of neural value systems, in particular their role in constraining and accelerating

the selection of adaptive behaviors in somatic time. As in our previous theoretical work, we have shown that value-dependent learning can account for the development of adaptive behavior by modulating synaptic changes in sensory-motor mappings and sensory receptive fields. The main contribution of the present paper is the further demonstration that, without any additional assumptions, value-dependent learning can be usefully applied to the afferent connections to value systems themselves. Value itself thereby becomes adaptive in somatic time, with several important consequences.

In this discussion, we first consider the most critically salient features of value systems in the present model as well as some possible neurobiological substrates for value systems. We then review the results of the simulations and relate them to the experimental literature. Finally, we analyse the relationship of value-dependent learning to temporal difference models of reinforcement learning.

Value, value-dependent learning, and value systems

From a selectionist perspective, there are in general no programs, sets of instructions, or teachers explicitly controlling synaptic changes in neuronal systems.[11,12] There are, however, structures or constraints in the phenotype that reflect prior evolutionary selection for what we have called innate values.[39] Certain neural or behavioral events may acquire value if they predict events with innate value and therefore contribute to adaptive behavior and phenotypic fitness. In the present simulations, for instance, foveation (i.e. the behavior itself and the neuronal activity that brings it about) reflects acquired value, because foveation is likely to be followed by favorable consequences (i.e. reward, food ingestion).

Value-dependent learning refers to the way in which local synaptic changes in the nervous system can be influenced by global modulatory signals that are triggered by events associated with value, either innate or acquired. In general, these changes will be such that there is convergence towards adaptive behavior. For instance, in this and previous simulations (c.f. Fig. 19 in Ref. 39; Fig. 3 in Ref. 13), acquiring the ability to foveate an arbitrarily positioned stimulus requires the value-dependent formation of ordered sensorimotor maps. Similarly, in the present work, auditory discrimination learning requires appropriate connections from **Ain** to **A1** and from **A1** to **SC**. A consequence of broadcasting a global value signal to a large number of brain areas is that receptive field properties in sensory areas may change so as to preferentially sample cues with value. In the present simulations, this property appeared as adaptive changes in the tuning curves of **A1** units and it closely corresponded to experimental results obtained by Weinberger *et al.*[50] Their experiments demonstrate a CS-specific modification of frequency receptive fields in auditory cortex during conditioning. Tuning curves shift so that the new "best frequency" becomes that of the CS. Moreover, pairing of exogenous acetylcholine and a single tone results in a similar shift, with maximal change at the frequency paired with acetylcholine.[32]

We consider neural value systems to be brain structures that are particularly suited to mediate value-dependent learning (we discuss several candidates below). Such systems possess some important structural and functional characteristics, many of which are represented in a schematic way in the present model. Through diffuse projections, the value system **VAL** modulates synaptic changes in most areas of the simulated brain. **VAL** shows a transient response to sustained stimuli and it signals salient events, specified at first innately and then by progressive adaptation to the environment. Its afferent connections are subject to two selective mechanisms: (i) overall patterns of connections that are specified epigenetically during development can be selected by evolution over generations and mediate intrinsic or innate value; and (ii) particular connections can be selected by value-dependent synaptic changes within the organism's lifetime and mediate adaptive or acquired value.

Innate and acquired value

As shown in our previous work, innate value (related to various protective reflexes, consummatory activities, and homeostatic needs[12]) is both necessary and sufficient to account for a significant degree of behavioral adaptation, both in complete simulations[39,48] and in a real-world device.[13] Being evolutionarily determined, however, innate value cannot be precisely tuned to a particular environment or to the individual needs of a specific phenotype in somatic time. Such tuning could be achieved, however, by the evolution of means that enable the acquisition of value in somatic time. In the present study, we demonstrate that allowing value-dependent plasticity in the inputs to the value system itself effectively represents one such means. The result is acquired value, i.e. value systems come to respond to an increasing variety of neural and behavioral events, events that reliably precede others that are innately valuable or have already acquired value.

The simulations carried out here reveal several advantageous properties of acquired value. First, when value-dependent synaptic changes were allowed in the connections to the value system itself, foveating behavior was acquired earlier and more reliably. Second, the simulations show that acquired value can be important for high-order conditioning. For example, the model was able to learn a discrimination task without direct reinforcement: when the peripheral visual stimulus that had acquired value in the first stage was used to reinforce discriminative responses to acoustic stimuli of different frequencies, acquired value was manifested by the connections from units in **V1** to **ACe**. Disconnecting **ACe** from

V1 eliminated both acquired value and discrimination learning. This simulated lesion study parallels an experiment in monkeys by Gaffan and Harrison[18] in which a visual discrimination task was reinforced using an auditory secondary reinforcer. Disconnection of the amygdala from the modality of the secondary reinforcer severely impaired discrimination. Third, the simulations show that, under certain circumstances, acquired value enables the model to override a temporary aversive stimulus in order to get to a potentially valuable situation. If the relevant afferents to the value system are eliminated or rendered not plastic, this is not possible. In the example above, if **V1** is disconnected from **ACe**, discrimination learning is determined only by the immediate and innately specified consequences of an action.

Candidate neural substrates for value systems

Neuromodulatory systems in the brain[23,31,35] are natural candidates for acting as value systems. While we do not suggest that any particular neurotransmitter system is the value system or is exclusively concerned with value, both cholinergic and aminergic systems seem to satisfy the major requirements. Considerable evidence suggests that monoaminergic and cholinergic neurotransmission can modulate enduring changes in synaptic strength.[22] There is evidence for the modulation of (i) experience-dependent changes in synaptic strength,[3,5,25,37] (ii) behavioral plasticity,[29,30,45,49] and (iii) long term potentiation of synaptic strength.[21,26] Cholinergic and monoaminergic systems have very diffuse projections.[31,35] Cholinergic and aminergic neurons respond to stimuli that have behavioral significance.[24,28]

Areas which project directly or indirectly to these neuromodulatory systems (e.g. LHA[47] and the amygdala[27]) can respond to stimuli coming from many sensory modalities.[36,42,43] In particular, there is evidence that the amygdala acts as a gateway through which salient events, both innate and learned,[27,34] may gain access to cholinergic[19] and dopaminergic cell groups and thereby influence learning.[8,15,24,27,29,34] In the model, ACe receives inputs not only from visual and auditory areas, but also from the LHA, which is implicated in many essential homeostatic functions.[33,41,47]

Learning-phase specificity of the responses of value systems

An important characteristic of value systems is their adaptation to sustained input, i.e. their tendency to respond preferentially to changes in their input. In the model, while **ACe** responds in a sustained way to its input, **VAL** only responds to changes in the input it receives from **ACe**. A consequence of the fact that the output of the value system is the time derivative of its input is that value-dependent modulation of plasticity is greatest when a behavior is incorporated into a learned sequence. As learning proceeds, early components of a behavioral sequence elicit value while late components lose this capacity. We have shown that this occurs in our simulations as indicated by a decrease in s_{VAL} at the point of foveation (Fig. 5). Experimental support for this notion comes from the adaptation of dopaminergic neurons: Ljungberg *et al.*[28] recorded unit activity in (cell groups) A8, A9, and A10 during operant conditioning of a reaction time task. Monkeys had to reach towards a lever when a light was illuminated. During acquisition, half the recorded dopaminergic neurons were phasically activated by a drop of liquid, delivered in order to reinforce the reaching movement. With established task performance, however, these neurons lost responses to this primary reward.

During discrimination learning in the present model, we observed the transfer of value-system responses from the conditioned reinforcer (visual stimulus, CS_1) to the conditioned stimulus (tone, CS_2) that predicted the CS_1. This resulted in acquisition of the capacity to elicit discriminative eye movements. In the experiment by Ljungberg *et al.*[28] described above, the loss of dopaminergic neuron responses to the primary reward was associated with an increasing response to the conditioned light stimulus.

Because such a transfer depends on plasticity in the connections between the modality of the discriminative stimulus and the amygdala (e.g., **A1** → **ACe** projections), the model suggests an interesting and somewhat counterintuitive experimental prediction: Transfer of unit responses in dopaminergic neurons, and in particular habituation of responses to a CS_1, should be abolished by disconnecting the amygdala from the modality of the discriminative CS_2. In the model, disconnection of **ACe** from **A1** was in fact found to abolish transfer of value responses and habituation to the light (results not shown).

Constraints on the value model

It is important to point out that the model has several limitations that require further comment. First, the link between foveation and reward is extremely simplistic. In the natural environment, many behaviors would precede and intervene between foveating a visual target and appetitive reward. We did not model these behaviors explicitly. The main reason for using simple behavioral contingencies was to emulate experimental conditioning paradigms and thus relate our findings to the experimental literature. We have assumed that value-dependent linking of behavioral sequences could also operate in a natural environment. Second, we did not consider value-dependent plasticity in connections between motor units (e.g., intrinsic connections within **SC**). This is clearly a very interesting area which we plan to pursue in terms of procedural learning and skill acquisition. Third, all the stimuli were either single points or tones. This simplifying device meant that all the sensory cues were uniquely identified in some sensory space and this obviated the complexities of

perceptual categorization and choice that have been explicitly addressed in previous work from our laboratory.[39,48] Provisional work using multiple visual stimuli of different colors shows that the current model can respond selectively to different wavelengths and conjunctions of wavelengths and that the results can be extended into this arena.

Relationship to temporal difference models of reinforcement learning

While the work presented here is primarily based on our previous theoretical work on value,[13,39,48] there are links with several theories of learning and reinforcement. An important characteristic of value systems is that their activity reflects the changes in their inputs. Because of this characteristic, the discharge of the value system is uncorrelated with its input (the derivative of a stationary stochastic process is uncorrelated with the process itself; see Ref. 9). As a consequence, runaway facilitation, which would be of no adaptive value, is avoided.

The use of the time derivative of convergent sensory signals is also a key aspect of temporal difference (TD) models of reinforcement learning.[46] TD models share with value-dependent learning the ability selectively to amplify behaviors that are initially generated by stochastic processes. This selection is based on reinforcement signals that are derived from the consequences of the total activity of the system. As has been found in TD models, stimuli that acquire value through value-dependent learning in the present model come to predict the occurrence of other valuable events. There are, however, some qualitative differences in the nature of this prediction which we describe in the Appendix. Unlike TD models, value-dependent learning requires no special apparatus to construct associative strengths (see Equation 2.1). The same rule for changing synaptic strengths is used for all types of connections, whether they pertain to acquiring value, to sensorimotor integration, or to the configuration of receptive fields. Most importantly, the notion of value is firmly rooted in evolutionary biology[12] and it has specific neurobiological correlates in both anatomy and physiology.

CONCLUSION

Several important properties of value acting in the nervous system are seen in its dynamic, context-sensitive character and its role as a constraint rather than as a precise or fixed set of instructions. Value is not an invariant that can be used to label a known world either in evolutionary or in somatic time. Inasmuch as the environment is unpredictable and open-ended and no two individuals are the same, the value of an event cannot in general be specified precisely *a priori*. This limits the usefulness of value descriptors that ignore either the history of the individual or the context in which they are exercised. On the other hand, this very limitation makes apparent the advantage in evolutionary terms of having value systems that are themselves adaptive in somatic time. In this paper, we have shown that this can be achieved with no further assumption than the requirement that connections to value systems themselves be under the same selectional constraints as those governing sensorimotor integration.

Evolutionary and somatic selection interact in interesting ways.[12] Given value systems with the appropriate anatomical and physiological characteristics, value can mediate its own acquisition during an organism's lifetime. During evolution, natural selection will favor value systems if their tendency to support acquired value and build up appropriate behavioral sequences leads to increases in adaptive fitness. Thus, while value systems constrain the selection of adaptive behavior in somatic time, they are also subject to selection in evolutionary time for those anatomical and neurophysiological characteristics that increase fitness.

Acknowledgements—This work was carried out as part of the Institute Fellows in Theoretical Neurobiology research program at The Neurosciences Institute, which is supported by the Neurosciences Research Foundation. The Foundation received major support for this research from the J.D. and C.T. MacArthur Foundation, the Lucille P. Markey Charitable Trust, and Sandoz Pharmaceutical Corporation. KJF and OS are W. M. Keck Foundation Fellows.

REFERENCES

1. Barto A. G., Sutton R. S. and Anderson C. W. (1983) Neuronlike adaptive elements that can solve difficult learning and control problems. *IEEE Transactions Syst. Man Cybern.* **SMC-13,** 834–846.
2. Barto A. G., Sutton R. S. and Watkins C. J. C. H. (1990) Learning and sequential decision making. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel M. and Moore J.), pp. 539–602. MIT Press, Cambridge.
3. Bear M. F. and Singer W. (1986) Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature* **320,** 172–176.
4. Borisenko A. I. and Tarapov V. E. (1968) *Vector and Tensor Analysis with Applications.* Dover, New York, NY.
5. Brocher S., Artola A. and Singer W. (1992) Agonists of cholinergic and noradrenergic receptors facilitate synergistically the induction of long-term potentiation in slices of rat visual cortex. *Brain Res.* **573,** 27–36.
6. Buka S. L. and Lipsitt L. P. (1991) Newborn sucking behavior and its relation to grasping. *Infant Behav. Dev.* **14,** 59–67.
7. Camhi J. M. (1984) *Neuroethology. Nerve Cells and the Natural Behavior of Animals.* Sinauer Associates, MA.
8. Cador M., Robbins T. W. and Everitt B. J. (1989) Involvement of the amygdala in stimulus–reward associations: Interaction with the ventral striatum. *Neuroscience* **30,** 77–86.
9. Cox D. R. and Miller H. D. *The Theory of Stochastic Processes.* Chapman & Hall, New York, NY.

10. DeLong M. R., Crutcher M. D. and Georgopoulos A. P. (1983) Relations between movement and single cell discharge in the substantia nigra of the behaving monkey. *J. Neurosci.* **3,** 1599–1606.
11. Edelman G. M. (1978) Group selection and phasic reentrant signalling: a theory of higher brain function. In *The Mindful Brain* (eds Edelman G. M. and Mountcastle V. B.), pp. 51–100. MIT Press, Cambridge, MA.
12. Edelman G. M. (1987) *Neural Darwinism.* Basic Books, New York, NY.
13. Edelman G. M., Reeke G. N., Gall W. E., Tononi G., Williams D. and Sporns O. (1992) Synthetic neural modelling applied to a real-world artifact. *Proc. natn. Acad. Sci. U.S.A.* **89,** 7267–7271.
14. Edelman G. M. (1993) Neural Darwinism: selection and reentrant signalling in higher brain function. *Neuron* **10,** 115–125.
15. Everitt B. J., Cador M. and Robbins T. W. (1989) Interactions between the amygdala and ventral striatum in stimulus-reward associations using a second-order schedule of sexual reinforcement. *Neuroscience* **30,** 63–75.
16. Fibiger H. C. and Phillips A. G. (1986) Reward, motivation, cognition: psychobiology of the meso-telencephalic dopamine systems. In *Handbook of Physiology: The Nervous System. Vol. IV, Section I,* pp. 647–675. American Physiological Society, Bethesda, MD.
17. Freeman A. S. and Bunney B. S. (1987) Activity of A9 and A10 dopaminergic neurons in unrestrained rats: further characterization and effects of cholecystokinin. *Brain Res.* **405,** 46–55.
18. Gaffan D. and Harrison S. (1987) Amygdalectomy and disconnection in visual learning for auditory secondary reinforcement by monkeys. *J. Neurosci.* **7,** 2285–2292.
19. Grove E. A. and Nauta W. J. H. (1984) Light microscopic evidence for striatal and amygdaloid input to cholinergic cell group CH4 in the rat. *Neurosci. Abstr.* **10,** 7.
20. Hailman J. P. (1969) How instinct is learned. *Sci. Am.* **221,** 98–108.
21. Harley C. (1991) Noradrenergic and locus coeruleus modulation of the perforant path-evoked potential in rat dentate gyrus supports a role for the locus coeruleus in attentional and memorial processes. *Prog. Brain Res.* **88,** 307–322.
22. Hemmings H. C., Nestler E. J., Walaas S. I., Ouimet C. C. and Greengard P. (1987) Protein phosphorylation and neuronal function: DARPP-32, an illustrative example. In *Synaptic Function* (eds Edelman G. M., Gall W. E. and Cowan W. M.), pp. 213–249. Wiley, New York.
23. Jacobs B. L. and Azmitia E. C. (1992) Structure and functional of the brain serotonin system. *Physiol. Rev.* **72,** 165–228.
24. Kapp B. S., Wilson A., Pascoe J. P., Supple W. and Whalen P. J. (1990) A neuroanatomical systems analysis of conditioned bradycardia in the rabbit. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel M. and Moore J.). MIT Press, Cambridge, MA.
25. Kasamatsu T. (1991) Adrenergic regulation of visuocortical plasticity: a role of the locus coeruleus system. *Prog. Brain Res.* **88,** 599–616.
26. Klancnik J. M. and Phillips A. G. (1991) Modulation of synaptic efficacy in the dentate gyrus of the rat by electrical stimulation of the median raphe nucleus. *Brain Res.* **557,** 236–240.
27. LeDoux J. E. (1990) Information flow from sensation to emotion: Plasticity in the neural computation of stimulus value. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel M. and Moore J.), pp. 3–52. MIT Press, Cambridge, MA.
28. Ljungberg T., Apicella P. and Schultz W. (1992) Responses of monkey dopamine neurones during learning of behavioral reactions. *J. Neurophysiol.* **67,** 145–163.
29. McGaugh J. (1992) Neuromodulatory systems and the regulation of memory storage. In *Neuropsychology of Memory* (eds Squire L. and Butters N.), pp. 386–401. Guildford Press, New York, NY.
30. van Neerven J., Pompeiano O. and Collewijn H. (1991) Effects of GABAergic and noradrenergic injections into the cerebellar flocculus on vestibular-ocular reflexes in rabbit. *Prog. Brain Res.* **88,** 405–498.
31. Mesulam M. M., Mufson E. J., Levey A. L. and Wainer B. H. (1983) Cholinergic innervation of the cortex by basal forebrain: Cytochemistry and cortical connections of the septal area, diagonal band nucleus, nucleus basalis (substantia innominata) and hypothalamus in the rhesus monkey. *J. comp. Neurol.* **214,** 170–197.
32. Metherate R. and Weinberger N. M. (1989) Acetylcholine produces stimulus-specific receptive field alterations in cat auditory system. *Brain Res.* **480,** 372–377.
33. Mora F., Rolls E. T. and Burton M. J. (1976) Modulation during learning of the responses of neurones in the lateral hypothalamus to the sight of food. *Expl Neurol.* **53,** 508–519.
34. Murray E. A. (1991) Contributions of the amygdala complex to behavior in macaque monkeys. *Prog. Brain Res.* **87,** 167–180.
35. Oades R. D. and Halliday G. M. (1987) Ventral tegmental (A10) system: Neurobiology. 1. Anatomy and connectivity. *Brain Res. Rev.* **12,** 117–165.
36. Ono T., Nakamura K., Fukuda M. and Kobayashi T. (1992) Catecholamine and acetylcholine sensitivity of rat lateral hypothalamic neurons related to learning. *J. Neurophysiol.* **67,** 265–279.
37. Rauschecker J. P. (1991) Mechanisms of visual plasticity: Hebb synapses, NMDA receptors, and beyond. *Physiol. Rev.* **71,** 587–614.
38. Reeke G. N. and Edelman G. M. (1987) Selective neural networks and their implications for recognition automata. *Int. J. Supercomputer Appl.* **1,** 44–69.
39. Reeke G. N., Finkel L. H., Sporns O. and Edelman G. M. (1990) Synthetic neural modelling: a multilevel approach to the analysis of brain complexity. In *Signal and Sense. Local and Global Order in Perceptual Maps* (eds Edelman G. M., Gall W. E. and Cowan W. M.), pp. 607–707. Wiley, New York, NY.
40. Richardson R. T. and Delong M. R. (1986) Nucleus basalis of Meynert neuronal activity during a delayed response task in monkey. *Brain Res.* **399,** 364–368.
41. Rolls E. T., Burton M. J. and Mora F. (1980) Neurophysiological analysis of brain-stimulation reward in the monkey. *Brain Res.* **194,** 339–357.
42. Sessler F. M., Cheng J. T. and Waterhouse B. D. (1986) Effects of endogenous monoamines on lateral hypothalamic neuronal responses to iontophoretically applied acetylcholine and systematic changes in osmotic and blood pressure. *Soc. Neurosci. Abstr.* **12,** 1392.
43 Shimizu N., Take S., Horzi T. and Oomura Y. (1992) *In vivo* measurement of hypothalamic serotonin release by intracerebral microdialysis—significant enhancement by immobilization stress in rats. *Brain Res. Bull.* **28,** 727–734.

44. Siegel A. and Edinger H. (1981) Neural control of aggression and rage behavior. In *Handbook of the Hypothalamus: Behavioral Studies of the Hypothalamus* (eds Morgane P. J. and Panksepp J.), pp. 203–240. Dekker, New York, NY.
45. Sullivan R. M., McGaugh J. L. and Leon M. (1991) Norepinephrine-induced plasticity and one trial olfactory learning in neonatal rats. *Devl Brain Res.* **60,** 219–228.
46. Sutton R. S. and Barto A. G. (1990) Time derivative models of Pavlovian reinforcement. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel M. and Moore J.), pp. 497–538. MIT Press, Cambridge, MA.
47. Swanson L. W. (1987) The hypothalamus. In *Handbook of Chemical Neuroanatomy: Integrated Systems of the CNS* (eds Björklund A., Hökfelt T. and Swanson L. W.), pp. 1–124. Elsevier, Amsterdam.
48. Tononi G., Sporns O. and Edelman G. M. (1992) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex* **2,** 310–335.
49. Velley L., Cardo B., Kempf E., Mormede P., Nassif-Caudarella S. and Velly J. (1991) Facilitation of learning consecutive to electrical stimulation of the locus coeruleus: cognitive alteration or stress reduction. *Prog. Brain Res.* **88,** 555–570.
50. Weinberger N. W., Ashe J. H., Metherate R., McKenna T. M., Diamond D. M., Bakin J. S., Lennartz R. C. and Cassady J. M. (1990) Neural adaptive information processing: a preliminary model of receptive field plasticity in auditory cortex during Pavlovian conditioning. In *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel M. and Moore J.), pp. 91–138. MIT Press, Cambridge, MA.
51. White N. M. and Milner P. M. (1992) The psychobiology of reinforcers. *A. Rev. Psychol.* **43,** 443-471.
52. Yuan C. S. and Barber W. D. (1992) Hypothalamic unitary responses to gastric vagal input from the proximal stomach. *Am. J. Physiol.* **262,** G74–G80.

(*Accepted* 15 *September* 1993)

APPENDIX

Value-dependent learning provides a paradigm for the acquisition of adaptive behavior that does not require an external "teacher" to provide detailed error signals dependent on preestablished criteria of correct output, as employed in so-called "supervised" learning. In this Appendix, we compare and contrast value-dependent learning with reinforcement learning, which has emerged as a distinct alternative to supervised and unsupervised learning in neural network and control theory. One important class of reinforcement learning models comprises TD models.[1] The basic hypothesis of these models is that "reinforcement is the time derivative of a composite association combining innate (US) and acquired (CS) associations".[46] The similarity is evident to the present proposal for value-dependent learning, in which input to neuronal value systems (in the model, s_{ACe}) is differentiated to produce V, a global signal that modulates synaptic plasticity. The models differ in detail in that the TD model (as presented in Ref. 1) explicitly includes a specific formalism for predicting future reinforcement as a function of system inputs; in value-dependent learning, this function emerges implicitly, and more generally, as a consequence of the activity of neurons in value systems that have no special mechanisms adumbrated for this purpose. Here, we examine this key difference in some detail, using a continuous time formulation. (Our model may be considered a discrete-time approximation of this formulation.)

The input to a value system can be thought of as a potential to elicit value that varies according to the current state of the system and of the environment, which may be considered to define a location in an abstract, time-independent state space. An analogy can be drawn between an unchanging potential field (corresponding to these inputs) and the energy (corresponding to value) associated with movement in that field, which depends upon the field gradients and the direction of motion.[4] In what follows, let this potential be denoted by ϕ (in the model $\phi = s_{ACe}$). Furthermore let ϕ have innate and acquired components $\phi = \phi_i + \phi_a$ [in the model, $\phi_i = s_{LHA}$, $\phi_a = s_{(V1/A1)}$]. Using this distinction between the potential (ϕ) and value (V), we can consider reinforcement learning in the light of value learning.

In the TD model, the equality

$$\Delta C_i = \beta [\lambda (t+1) + \gamma \phi_p (t+1) - \phi_p (t)] \cdot \alpha_i \bar{x}_i \qquad \text{(Eqn 2.1)}$$

defines the update rule for C_i, which is the associative strength of US i, α_i, β and γ are positive constants, and ϕ_p is here called prediction and is $\Sigma C_j x_j$. $\bar{x}_i$ represents a trace of the i^{th} CS, and $\lambda (t)$ is the effectiveness of the US. The condition for the associative strengths to stabilize ($\Delta C_i = 0$) is:

$$\lambda (t+1) + \gamma \phi_p (t+1) = \phi_p (t)$$

or on repeated substitution:

$$\phi_p (t) = \lambda (t+1) + \gamma \lambda (t+2) + \gamma^2 \lambda (t+3) \ldots \gamma^n \lambda (t+n+1). \qquad \text{(Eqn 2.2)}$$

Because $\gamma < 1$, $\phi_p (t)$ represents a discounted sum of expected λ, or the effectiveness of unconditional stimuli that will be encountered in the future. The discounting depends on how fast γ^n decays. This interpretation of $\phi_p (t)$ as a predictor of innate associations allows the system to derive an estimate of reinforcement in the absence of a US. From the perspective of stochastic dynamic programming, the associative strengths $\phi_p (t)$ can be thought of as representing gradients of secondary reinforcement,[2] which intervene between sporadic unconditioned stimuli.

In the case of value-dependent selection, the potential to elicit innate value, ϕ_i, represents some valuable internal or autonomic state, which increases after the US and then decreases monotonically with time. Without loss of generality, $d\phi_i/dt = \lambda - \chi (t)$ where $\chi (t)$ is an arbitrary non-negative function of time and value is:

$$V = d\phi/dt = d(\phi_i + \phi_a)/dt = \lambda (t) - \chi (t) + d\phi_a (t)/dt.$$

The requirement for connectivity (c_{ij}) to stop changing is $V = 0$ or:

$$d\phi_a (t)/dt = \chi (t) - \lambda (t). \qquad \text{(Eqn 2.3)}$$

Compare this with the equivalent equation, in continuous time, for the TD model:

$$d\phi_p(t)/dt = (1-\gamma)\phi_p(t) - \lambda(t). \quad \text{(Eqn 2.4)}$$

Solutions of the differential equations 2.3 and 2.4 are:

$$\phi(t) = \int_t^{\infty} e^{-\theta(t,u)}\lambda(u)\,du. \quad \text{(Eqn 2.5)}$$

where:

$$\theta(t,u)_{\text{Value}} = \int_t^u \chi(\tau)/\phi_a(\tau)\,d\tau$$

$$\theta(t,u)_{\text{TD model}} = \int_t^u (1-\gamma)\,d\tau = (1-\gamma)(u-t).$$

$\theta(t,u) > 0$ in both cases and $\phi_a(t)$ represents a discounted prediction of $\lambda(t)$. Equation 2.5 can be thought of as a convolution of $\lambda(t)$ where (i) the convolution function (discounting function) changes with time and (ii) it runs from the present into the future.

The main difference between TD and value learning is in the nature of the discounting, which is fixed in the TD model but self-adjusting and dynamic in value learning. The nature of this adjustment means that discounting is greatest shortly after an US, when, assuming convergence has been reached, ϕ_i is falling fast and ϕ_a is low (note $\phi_i + \phi_a = \text{constant}$). Conversely, the effective prediction becomes more far-sighted with time elapsed since the last restoration of homeostasis (increase in ϕ_i). In other words, in an environment with sparse and infrequent unconditioned stimuli (innate value), the average prediction is more long-ranging. In a sense, reinforcement learning can be considered a special case of value learning, in which $\chi(t) = (1-\gamma)\phi_p(t)$.

For convergence to occur, V must asymptotically approach 0. Equivalently the sum (or more generally the interaction) of ϕ_i and ϕ_a is constant. This means that innate and acquired value should complement each other. This phenomenon is seen in Fig. 4 where connection strengths from **V1** to **ACe** are low where there is innate value and high where there is no innate value. This complementary interaction means there is a smooth progression from neuronal events with acquired value to events with innate value. Once established, and in the absence of changes in environmental contingencies that would affect innate or acquired value responses, this progression is exempt from further selective pressure in somatic time (because at later stages of learning, V tends to become small). A final experimental prediction ensues from this observation: in the absence of an expected reward, value system responses should show a decrease in activity at the time when the reward would normally be delivered.