PERGAMON

Neural Networks 15 (2002) 761–774

Neural Networks

www.elsevier.com/locate/neunet

2002 Special issue

Neuromodulation and plasticity in an autonomous robot

Olaf Sporns*, William H. Alexander

Department of Psychology, Indiana University, Bloomington, IN 47405, USA

Received 10 October 2001; revised 2 April 2002; accepted 2 April 2002

Abstract

In this paper we implement a computational model of a neuromodulatory system in an autonomous robot. The output of the neuromodulatory system acts as a value signal, modulating widely distributed synaptic changes. The model is based on anatomical and physiological properties of midbrain diffuse ascending systems, in particular parts of the dopamine and noradrenaline systems. During reward conditioning, the model learns to generate tonic and phasic signals that represent predictions and prediction errors, including precisely timed negative signals if expected rewards are omitted or delayed. We test the robot's learning and behavior in different environmental contexts and observe changes in the development of the neuromodulatory system that depend upon environmental factors. Simulation of a computational model incorporating both reward-related and aversive stimuli leads to the emergence of conditioned reward and aversive behaviors. These studies represent a step towards investigating computational aspects of neuromodulatory systems in autonomous robots. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Learning; Reinforcement learning; Temporal difference learning; Biorobotics; Value systems; Dopamine; Reward conditioning; Adaptive systems

1. Introduction

Throughout the nervous system, neuromodulators have a variety of functions ranging from the regulation of neuronal excitability and plasticity to effects on gene expression and structural modifications in neural circuits. In computational models, neuromodulation may be viewed as exerting an influence on neuronal response functions, learning rates, or other model parameters (Doya, 2000; Fellous & Linster, 1998; Hasselmo, 1995; Hasselmo, Wyble, & Fransen, 2002; Pennartz, 1996; Servan-Schreiber, Printz, & Cohen, 1990). Several models have focused on the potential functional roles of diffusely projecting neuromodulatory systems in influencing the magnitude and direction of synaptic plasticity, ultimately resulting in behavioral change. In this paper we present one possible computational implementation of such systems and we investigate their action during learning and autonomous behavior in a robot.

Adaptive behavior requires that the behaving system or agent, be it an organism or a robot, is sensitive to the consequences of its own actions. Different environmental stimuli and events have different saliency, defined here as their 'predictive power' or 'relevance' to the agent. Salient stimuli and events play special roles in influencing learning and plasticity. Whatever the mechanisms are that mediate plastic changes, they must be unsupervised (or self-supervised), allowing the agent to learn as a result of its own actions and in the absence of an external teacher. A number of computer simulations and robot models (Almassy, Edelman, & Sporns, 1998; Edelman et al., 1992; Friston, Tononi, Reeke, Sporns, & Edelman, 1994; Pfeifer & Scheier, 1999; Rucci, Tononi, & Edelman, 1997; Scheier & Lambrinos, 1996; Sporns, Almassy, & Edelman, 2000; Verschure, Wray, Sporns, Tononi, & Edelman, 1995) have used 'value systems' as internal mediators of environmental saliency. Value systems are entirely part of the neural network architecture of the agent. Their outputs serve to modulate neural activity or plasticity by delivering a diffuse, globally acting signal. Typically, value systems become active after the occurrence of specific sensory stimuli, often as a result of behavioral actions of the agent. Their response pattern is phasic and short-lasting, essentially serving as a timing signal for synaptic modification. While some aspects of value may be viewed as 'innate' and the result of evolutionary adaptation (such as the immediate effects on behavior of food or noxious stimuli), other aspects are 'acquired' and the result of an agent's experience and behavior. In computational models, innate value is

* Corresponding author. Tel.: +1-812-855-2772.
E-mail address: osporns@indiana.edu (O. Sporns).

PII: S0893-6080(02)00062-X

Nomenclature

s_i	cell activation
$A(t)$	total synaptic input
c_{ij}	synaptic weight
Ω	cell persistence
ϕ	nonlinear cell response function (ρ: slope, θ: activation threshold)
β	behavioral threshold
ε	synaptic decay rate
η	learning rate
V	value signal
F	nonlinear postsynaptic function (κ, φ_1, φ_2, ξ_1 and ξ_2: parameters determining shape of F).

determined by fixed (hard-wired) connections, while acquired value is the result of synaptic modifications within the value system itself (Friston et al., 1994; Sporns et al., 2000).

The multiple diffuse ascending systems of the vertebrate brain have anatomical and physiological characteristics that render them good candidates for mediating neuromodulatory effects on synaptic plasticity. They show phasic activation in response to a variety of salient sensory stimuli, ranging from food rewards (Ljungberg, Apicella, & Schultz, 1992) to stimuli attracting attention, signaling novelty or triggering aversive responses (Aston-Jones et al., 1991; Jacobs, 1986). Their anatomical projections reach widespread areas of the brain, including large parts of the cerebral cortex. Their immediate physiological effects include modulation of the 'signal-to-noise' ratio of cortical neuronal activity (Hasselmo, Linster, Ma, & Cekic, 1997) as well as modulation of synaptic efficacy (Bear & Singer, 1986; Hasselmo & Barkai, 1995). Several of these systems maintain structural connections that suggest they might interact at the level of the midbrain. In addition, they project to overlapping regions of cortex suggesting interactions between different neurotransmitter systems at the level of their projection targets.

The function of the mammalian midbrain dopamine system in reward conditioning has been studied extensively in recent years (Schultz, 1998). One of its components, the ventral tegmental area (VTA), contains dopaminergic neurons projecting to widespread cortical areas, including frontal and prefrontal cortex. The majority of VTA dopamine neurons show phasic activation in response to food rewards. Their response pattern undergoes characteristic changes in the course of learning. Phasic activation following primary reward does not occur when the reward is reliably preceded by other reward predicting stimuli. These 'acquired' phasic responses occur immediately at the onset of stimuli that are predictive of rewards. In other words, dopamine responses are 'transferred' to conditioned, reward-predicting stimuli and become attenuated or disappear entirely for completely predicted primary rewards. If a fully predicted reward does not occur, dopamine neurons exhibit a transient depression of their baseline discharge rate at the time of the expected occurrence of the reward. This last finding suggests that the dopamine system has access to information concerning the timing of sensory inputs relative to the occurrence of reward. Several computational models of the midbrain dopamine system have been proposed (Montague, Dayan, & Sejnowski, 1996; Schultz, 1998; Schultz, Dayan, & Montague, 1997; Suri & Schultz, 2001), forging a strong connection between dopaminergic responses and temporal difference learning (Sutton & Barto, 1990). Neuromodulators other than dopamine may be involved in mediating the effects of aversive stimuli. Noradrenergic neurons of the mammalian locus coeruleus show responses to a wide variety of salient sensory stimuli, including those of an aversive nature. Noradrenergic neurons exhibit phasic discharge patterns (Aston-Jones, Rajkowski, Kubiak, & Alexinsky, 1994) and plastic changes related to behavior (Aston-Jones, Rajkowski, & Kubiak, 1997), although the predictive aspect of their responses is less clear.

In this paper, we focus on the question of how neuromodulatory systems operate in the course of relatively unconstrained behavior executed by an autonomous agent. We first present a neural implementation of a neuromodulatory system mediating reward that shares many structural and functional characteristics with the mammalian midbrain dopamine system. In addition, we implement a second neuromodulatory component mediating the effects of aversive stimuli on learning and plasticity. We demonstrate the operation of these two components of a neuromodulatory system in simple robot experiments, emphasizing the relationship between behavior and neural change.

2. Methods

2.1. General

All experiments reported in this paper were carried out using neural simulations implemented in Matlab 6.0 (Mathworks, Natick, MA), run on Linux 6.2 workstations (ASL, Newark, CA) and interfaced with autonomous

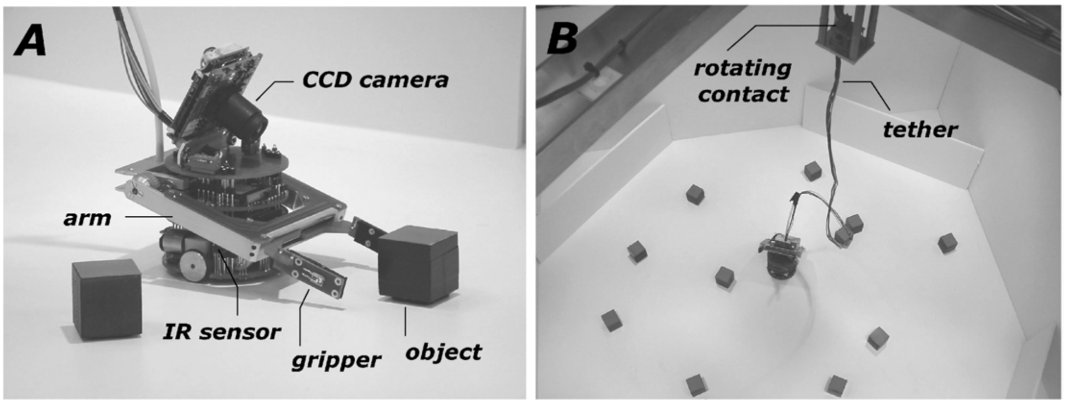


Fig. 1. (A) Monad, as configured for the experiments reported in this paper. (B) Robot environment. Environment consists of an enclosed platform (approx. 90 × 90 cm^2) containing objects. Monad's tether, consisting of power and communication lines, allows unrestricted movements. Monad is pictured in navigational mode, with arm/gripper module raised to allow IR sensor readings. Multiple objects are distributed at random throughout the environment.

khepera robots (K-Team, S.A., Préverenges, Switzerland), equipped with a color CCD video turret and a gripper module. Basic Matlab scripts and serial line communication modules for Linux are available from the authors upon request. Some additional material including short movies of robot behavior can be found at php.indiana.edu/~osporns/lab.htm.

2.2. Robot design

The robotic system described in this paper is named 'Monad', after the individual elements of being, the building blocks of the universe of metaphysics, central to the philosophies of Giordano Bruno and Gottfried Wilhelm Leibniz. Monad (Fig. 1(A)) consisted of a mobile circular platform ($\varnothing \approx 5$ cm) capable of translation and rotation limited to an approximate speed of 5 cm/s. An arm/gripper module allowed a one degree-of-freedom gripper to be raised and lowered, as well as closed and opened, thus permitting physical contact with objects. Rotational movements of the robot wheels as well as movements of the arm/gripper were triggered by the activation of neural units within the simulation. Robot sensors consisted of a color CCD camera (fl 3 mm, 59.8 × 42.6° field of view) mounted on top of the mobile platform and angled forward, 8 infrared (IR) sensors mounted around the periphery of the platform, and resistivity sensors on the inner surfaces of the gripper. The color camera continuously transmitted RGB images (320 × 240 pixels), which were used as input to the visual part of the neural simulation. A low-resolution gray level image was derived from the RGB signal and used as input to the visual approach system. The 8 IR sensors acquired infrared reflectance readings at 20 ms intervals that were converted into motor signals using a fixed motor map. Synaptic weights were set so as to effectively steer Monad away from obstacles (see below). The resistivity sensors recorded the conductivity across the surface of objects, a signal that served as a measure of object 'taste' (low conductivity = appetitive taste, high conductivity = aversive taste). Resistivity was recorded as a scalar variable with 8-bit resolution; for simplicity, all readings were converted to binary outputs, with high conductivity activating an aversive taste receptor (T_{av}) and low conductivity activating an appetitive taste receptor (T_{ap}).

2.3. Robot environment

Monad was tethered via a flexible wire bundle and a rotating contact, which was mounted directly above an environmental enclosure (Fig. 1(B); 90 × 90 cm^2), with white floor and walls, illuminated by DC-powered halogen lights and containing various stimulus objects. These objects were black 1 in. cubes, with a single colored face at the top. Objects were visually indistinguishable except for their color, which was red or blue. The black surfaces of the objects were either electrically non-conductive or conductive, a physical property analogous to 'taste'. Objects were sufficiently light to be easily manipulated by Monad's gripper and were either presented manually by the experimenter or placed at random positions within the environment at the beginning of an experiment.

2.4. Robot behavior

At each point in time, Monad was in one of several behavioral modes, forming a simple behavioral hierarchy. By default and in the absence of overt visual targets, Monad was traversing the environment at fairly constant speed ($\approx$3–5 cm/s) while sensing IR reflectance and avoiding obstacles or walls (*mode = navigate*). If a high-contrast visual target was detected, the target was approached under the guidance of the visual approach system (*mode = approach*). This system translated the activity of units in a visual map into motor (speed) commands relayed to Monad's two high-precision DC motors, via a fixed motor map. While approaching a target, all steering movements were under visual control and no IR sensor readings were taken. Approach terminated if a visual target loomed large (i.e. was physically close) in the center of Monad's visual field (fovea). Before learning, Monad attempted to establish physical contact with all targets located in close physical proximity (*mode* = *interact*), by lowering the arm and closing the gripper. After obtaining sensor readings of the conductivity of an object, the object was released (*mode* = *withdraw*) by opening the gripper and returning the arm to a raised position. Then, Monad turned away from the released object to resume navigation or approach another target.

The behavioral sequence outlined earlier constituted Monad's 'default' or innately specified behavioral pattern. All objects of high-visual contrast were approached and 'tasted', regardless of their visual appearance. When objects were encountered, Monad emitted different unconditioned responses (UR) depending upon the nature of the unconditioned stimulus (US), i.e. 'taste'. If an object was found to be appetitive, a prolonged gripping response ensued. If an object was found to be aversive, the gripping response was immediately terminated. After learning (essentially amounting to an instrumental conditioning paradigm), the visual appearance (color) of objects was sufficient to trigger conditioned responses, a reward-related conditioned response ($\mathrm{CR_R}$) consisting of immediate approach and gripping for appetitive stimuli, and an aversive conditioned response ($\mathrm{CR_V}$) consisting of immediate withdrawal without gripping for aversive stimuli. These conditioned responses were triggered by activation of motor units M_{ap} and M_{av}, respectively, as soon as their activation difference exceeded a behavioral threshold β ($\beta = 0.3$). In all robot experiments, connections driving $M_{\mathrm{ap}}/M_{\mathrm{av}}$ activation were subject to value-dependent learning (see below).

2.5. Neural simulation

All visual images and sensor readings were relayed to a neural simulation, and motor commands were initiated from the simulation and relayed to Monad via simple serial line commands. Neural and behavioral states were continuously recorded and saved for off-line display and analysis. In some cases, digital video recording of the environment was carried out in parallel. On average, a single simulation cycle required about 250 ms of CPU time.

All neural units were implemented using a continuous firing rate model with a single saturating non-linearity, according to

$$s_i(t+1) = \phi[A(t) + \Omega s_i(t)]$$

where $s_i(t)$ is the activity of unit i at time t, $A(t)$ is the total synaptic input to unit i at time t, Ω is the unit's temporal persistence ($0 < \Omega < 1$), and ϕ is a saturating nonlinear function, given as $\phi = \tanh(\rho[A(t) + \Omega s_i(t)])$ if $[\cdot] > \theta$ and $\phi = 0$ otherwise, with ρ denoting the slope of the function and θ acting as an activation threshold. $A(t)$ was calculated as the linear sum of all inhibitory and excitatory inputs, i.e. $\sum c_{ij} s_j(t)$. (For parameter values see Fig. 3.)

Schematic diagrams of the neural model and its constituent networks (including parameter values) are shown in Figs. 2–4. Fig. 2 shows segregated sensorimotor circuits that governed robot navigation and approach (Fig. 2(A) and (B)), as well as the robot's visual system (Fig. 2(C)). IR sensors are connected to motor units driving the two wheels of Monad via a fixed motor map (Fig. 2(A)). A raw image delivered by Monad's CCD camera provides input to the visual approach system, after being converted to a low-resolution format, thresholded and contrast-inverted (Fig. 2(B)). The resulting visual array is converted to steering motions of the two wheels using two symmetrical fixed motor maps. The central (foveal) part of the raw image is converted to red, green, blue and yellow (red + green/2) arrays, which are further processed using red–green and blue–yellow center–surround convolutions. The resulting neural arrays (R + G − , B + Y −) are used to drive color-selective units in C_{red} and C_{blue} that discount visual topography and report the presence or absence or red or blue color within the visual field. We note that the emphasis of the present model was on the computational aspects of the neuromodulatory system. Therefore, no attempt was made to implement complex categorization or sensorimotor mappings (for earlier work on visual categorization see Almassy et al., 1998; Krichmar, Snook, Edelman, & Sporns, 2000). Stimulus categories were limited to the object's color, specifically 'red' and 'blue' and category-dependent behavioral outputs were simple 'appetitive' or 'aversive' behaviors.

The robot's neuromodulatory system had two main components, mediating effects of appetitive (reward) and aversive stimuli, respectively. The structure of the reward component is shown in Fig. 3. Color selective units ($C_{\mathrm{red}}/C_{\mathrm{blue}}$) provided sensory inputs to a network transforming this input into a continuous temporal representation. As a result of excitatory and inhibitory interactions within this network (essentially forming a delay chain), stimulus-specific units $D_{\mathrm{red}}/D_{\mathrm{blue}}$ within this network became active after a specific amount of time had elapsed since the onset of their preferred stimulus. These units had fairly broad temporal tuning with significant mutual overlap in terms of their 'temporal receptive field' (see Figs. 5 and 8, below, for example activity traces). A more realistic implementation of such a network would consist of intermixed populations of neurons that show complex spatio-temporal activation patterns. D_{red} units projected to two subcomponents of the reward system mediating different aspects of reward stimuli through two distinct sets of modifiable connections. Both of these subcomponents consisted of an 'integrator' unit (S_{O1} and S_{O2}, respectively), which activated a feedforward inhibitory unit (S_{I1} and S_{I2}, respectively) and a phasic response unit (S_{S1} and S_{S2}, respectively). Connection and cell parameters for S_{O}, S_{I}, and S_{S} were chosen to confer the desired dynamic properties of sustained or tonic response (S_{O}), powerful inhibition (S_{I}) and burst-like phasic responses (S_{S}). S_{S1} and S_{S2} emitted two phasic components of the reward-related neuromodulatory output signal.

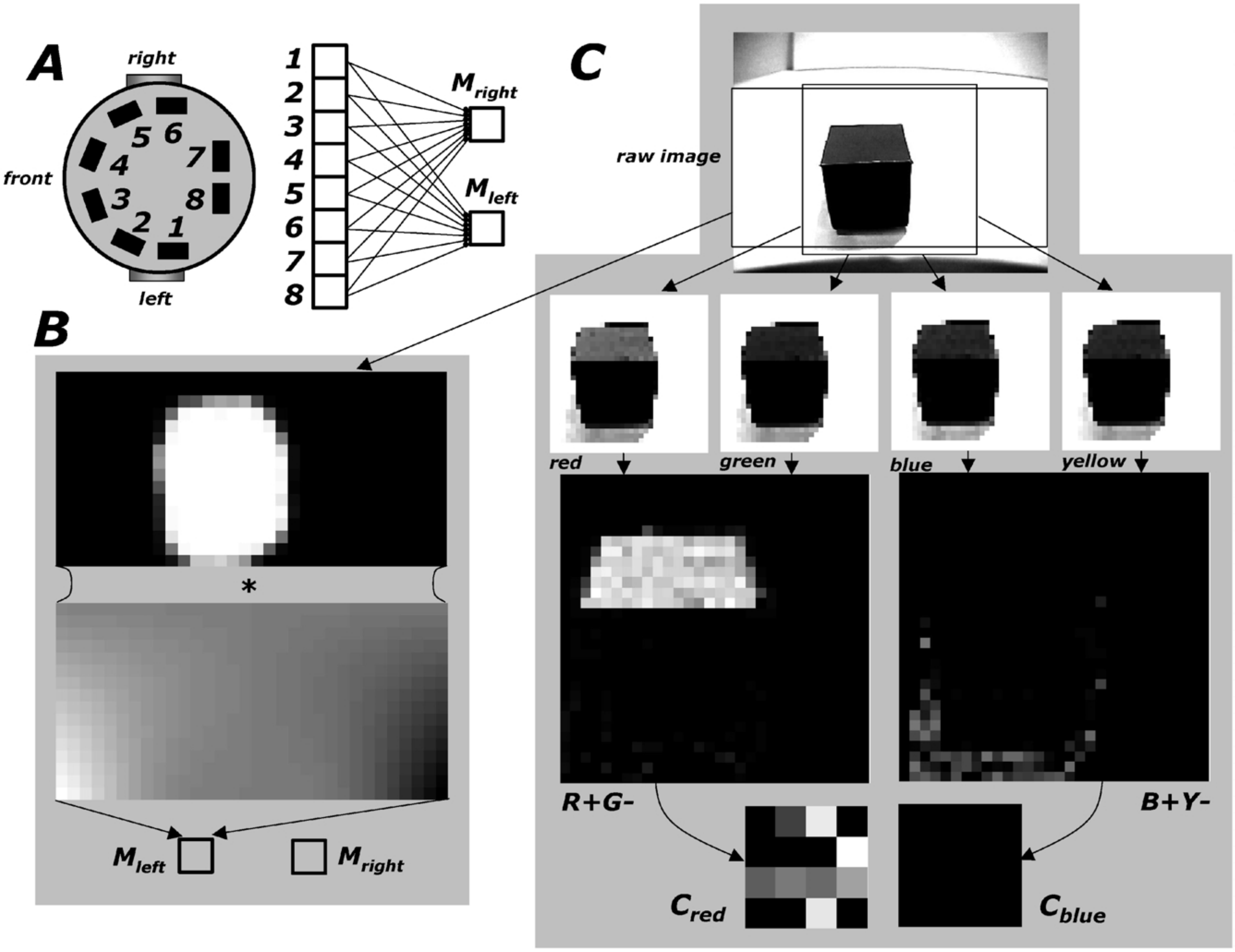
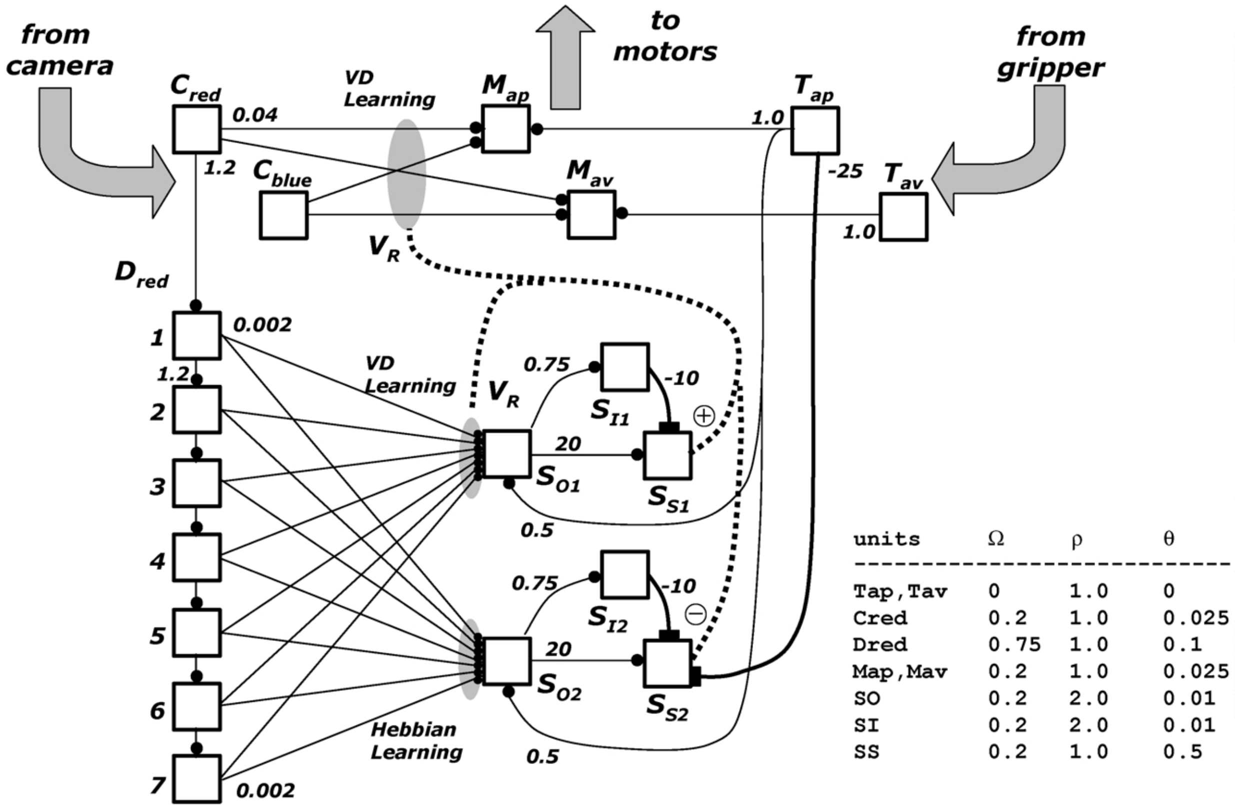


Fig. 2. (A) Schematic top view of Monad with positions of 8 IR sensors. Sensor readings are converted to motor (speed) signals to M_{left} and M_{right} via a fixed motor map. (B) Input to the visual approach system is provided after converting a large portion of the raw image (shown in panel C) into a low-resolution, high-contrast, black-and-white visual array. This array governs motor speeds through two fixed and symmetric motor maps, one for activating M_{left} and one for M_{right}. The motor map shown is for M_{left}. (C) Separate color channels derived from the central (foveal) portion of the raw image are processed by red–green and blue–yellow center–surround units (R + G − , and B + Y − , respectively). Their outputs drive color-selective units C_{red} and C_{blue} that discount spatial information (translation invariance).

Initially, the connections linking temporal delay units and $S_{\mathrm{O1}}/S_{\mathrm{O2}}$ were weak, but in the course of learning they became strengthened in specific patterns and capable of driving responses in $S_{\mathrm{O1}}/S_{\mathrm{O2}}$. $S_{\mathrm{O1}}/S_{\mathrm{O2}}$ were also activated by the primary reward (appetitive taste), through a strong 'innate' and excitatory connection. Due to their combined excitatory and inhibitory inputs, S_{S1} and S_{S2} both showed a phasic response profile, a short burst of activation followed by inhibition from S_{I1} and S_{I2}. The S_{S2} phasic response is inhibited ('cancelled') by the simultaneous occurrence of a primary reward, due to an 'innate' and inhibitory connection from T_{ap} to S_{S2}. Essentially, activation in S_{S1} and S_{S2} constituted a temporal derivative of increases (but not decreases) in $S_{\mathrm{O1}}/S_{\mathrm{O2}}$ activation. The firing level of S_{S1} was taken to be proportional to an increase in the level of neuromodulator released at projection targets over a stationary baseline. In turn, the firing level of S_{S2} was taken to be proportional to a decrease in the level of neuromodulator below the same stationary baseline. The overall level of neuromodulator released by the reward system, taken to be a 'value signal', was calculated as $V_{\mathrm{R}} = S_{\mathrm{S1}} - \lambda S_{\mathrm{S2}}$, with λ setting the overall difference in gain between the two components ($\lambda = 1$ in this paper). V_{R} was a signed scalar variable, which was positive if the level of neuromodulator increased above baseline and negative if it decreased below baseline (i.e. we assume that baseline = 0).

In a separate set of experiments, a second component of the neuromodulatory system is added, responsive to aversive stimuli. Fig. 4 shows the neural networks comprising this system, which is similar in its design to the positive subcomponent of the reward-mediating neuromodulatory system. All neuromodulatory components receive sensory inputs to S_{O1}, S_{O2} and S_{O3} from D_{red} and D_{blue} units (D_{red} units not shown in Fig. 4 for clarity). The output of the aversive neuromodulatory system V_{V} consists



units	Ω	ρ	θ
Tap,Tav	0	1.0	0
Cred	0.2	1.0	0.025
Dred	0.75	1.0	0.1
Map,Mav	0.2	1.0	0.025
SO	0.2	2.0	0.01
SI	0.2	2.0	0.01
SS	0.2	1.0	0.5

Fig. 3. Schematic diagram of the reward component of Monad's neuromodulatory system. Neural units and networks are indicated as boxes, excitatory connections are shown as thin lines connecting boxes, inhibitory connections are shown as thick lines, the output signals of the neuromodulatory system are indicated as thick hatched lines, and their modulatory targets are indicated as ellipsoid shaded areas. Some feedback inhibitory connections between D_{red} units that are used for generating temporally specific responses are not shown for clarity. Parameter values for neural units are listed on the right of the figure. Parameters ρ and θ refer to the slope and activation threshold of the nonlinear response function ϕ ($\phi = \tanh$ for all units shown in Figs. 3 and 4). Both sets of S_{O}, S_{I} and S_{S} units had identical unit parameters. Connection weights c_{ij} are indicated as positive or negative numbers. Parameter values for synaptic modification were: connections $C_{\mathrm{red}} \rightarrow M_{\mathrm{ap}}$: $\varepsilon = 0.001$, $\eta = 0.25$, $\kappa = 0.5$, $\varphi_1 = 40$, $\varphi_2 = 40$, $\xi_1 = 3$ and $\xi_2 = 5$; connections $D_{\mathrm{red}} \rightarrow S_{\mathrm{O1}}$: $\varepsilon = 0.001$, $\eta = 0.15$, $\kappa = 0.5$, $\varphi_1 = 60$, $\varphi_2 = 60$, $\xi_1 = 1$ and $\xi_2 = 3$; connections $D_{\mathrm{red}} \rightarrow S_{\mathrm{O2}}$: $\varepsilon = 0.001$, $\eta = 0.15$, $\kappa = 0.5$, $\varphi_1 = 6$, $\varphi_2 = 6$, $\xi_1 = 1$ and $\xi_2 = 4$. See text for a more detailed description of anatomy and physiology.

of the activity of S_{S3} alone; no temporal prediction similar to S_{S2} was implemented. The overall value signal is calculated as $V = V_{\mathrm{R}} + V_{\mathrm{V}}$, i.e. as the linear sum of the levels of the two reward-related and aversive neuromodulatory systems.

Two types of learning were employed in the present model, value-dependent learning and Hebbian learning. Through value-dependent learning, the value signal V influenced synaptic modification in sensorimotor connections linking color-selective units in $C_{\mathrm{red/blue}}$ to motor units in $M_{\mathrm{ap/av}}$, as well as in connections linking the temporal delay units $D_{\mathrm{red/blue}}$ to S_{O1} and S_{O3}. Thus, value exerted a dual influence, by directly modifying behavior through changes in sensorimotor linkages, and by modifying the response characteristics of parts of the neuromodulatory system itself. Connections that were subject to value-dependent learning were updated according to a ternary learning rule:

$$c_{ij}(t+1) = (1-\varepsilon)c_{ij}(t) + \eta s_j(t) F(s_i(t)) V$$

where $s_j(t)$ is the presynaptic activation, $s_i(t)$ is the postsynaptic activation, c_{ij} is the connection weight from unit j to unit i, ε, the incremental decay rate of connection weight per iteration, η, the learning rate, $F(\cdot)$ the nonlinear function applied to postsynaptic activity, V, the value signal. $F(\cdot)$ determined if a connection weight increased or decreased depending upon the level of postsynaptic activity. $F(\cdot)$ was a continuous saturating function ($-1 < \mathrm{F}(\cdot) < 1$), here modeled as

$$F = \kappa(1 - \tanh(\varphi_1 s_i(t) - \xi_1)) + \tanh(\varphi_2 s_i(t) - \xi_2).$$

Parameter values κ, φ_1, φ_2, ξ_1 and ξ_2 (see Fig. 3) were set by the experimenter solely to determine the shape of $F(\cdot)$ and were not thought to have direct physiological analogs. Given the parameter values given in the legends to Figs. 3 and 4, the shape of F modeled the observed dependence of the sign of synaptic modification on postsynaptic activity (cf. Bear, Cooper, & Ebner, 1987; Bienenstock, Cooper, & Munro, 1982). For very low levels of postsynaptic activity, F tends to be very close to zero and synaptic changes are very small. For low to intermediate levels of postsynaptic activity, F takes on negative values, resulting in synaptic weakening. High values of postsynaptic activity result in

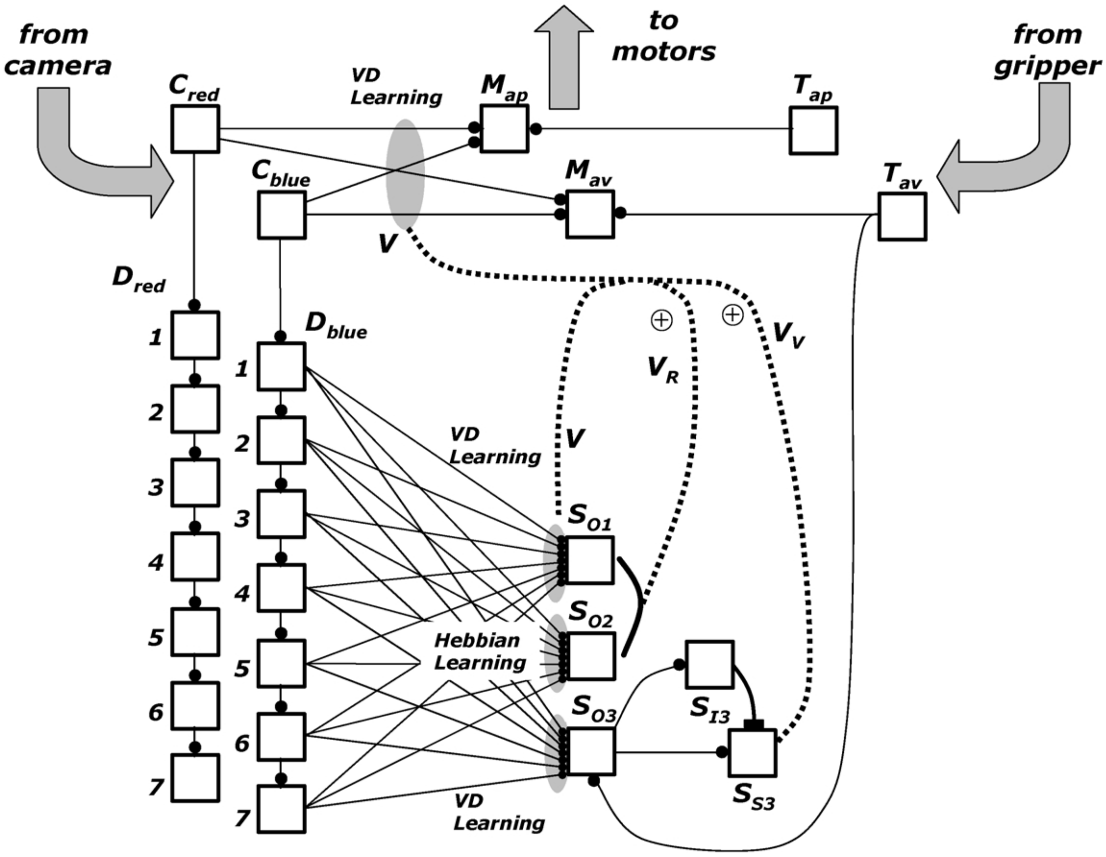


Fig. 4. Schematic diagram of the aversive component of Monad's neuromodulatory system (compare Fig. 3). Connections linking D_{red} units with S_{O1}, S_{O2} and S_{O3} are not shown for clarity, but are present within the model and subject to plastic change. Also, S_{I1}, S_{S1}, S_{I2}, and S_{S2} (see Fig. 3) are not shown, but are present in the model. Parameter values are unchanged from Fig. 3. Parameter values for synaptic modification in connections linking C_{blue} to M_{ap} and D_{blue} to S_{O1} and S_{O2} are identical to the ones shown in Fig. 3. For connections $D_{\mathrm{red/blue}} \rightarrow S_{\mathrm{O3}}$ parameter values were $\varepsilon = 0.001$, $\eta = 0.15$, $\kappa = 0.5$, $\varphi_1 = 60$, $\varphi_2 = 60$, $\xi_1 = 1$ and $\xi_2 = 3$. See text for a more detailed description of anatomy and physiology.

positive values for F (saturating at $F = 1$) and synaptic strengthening. For plastic connections that are not value-dependent (i.e. connections between $D_{\mathrm{red}}/D_{\mathrm{blue}}$ and S_{O2}), the value component of the earlier equation is set to 1, and thus synaptic modification depends entirely on pre- and post-synaptic activity only (Hebbian learning).

3. Results

The functional characteristics of the neuromodulatory system were explored in three sets of experiments. In the first set, objects associated with reward (appetitive taste) were presented manually to the robot, in order to ensure reproducible timing of reward delivery within individual learning trials. In the second set of experiments, objects were placed at random throughout the environment and all robot behavior and learning proceeded in a fully self-guided and autonomous fashion. This tended to degrade the consistent timing of reward delivery. In the third set of experiments, a component of the neuromodulatory system mediating aversive stimuli was added and both appetitive and aversive objects were used.

3.1. Robot experiments with consistent timing of reward

Our first goal was to investigate the development of neural connectivity and activation patterns in Monad's neuromodulatory system in a task setting in which there was consistent relative timing of object vision and subsequent reward (appetitive taste), controlled by the experimenter. To this end, the experimenter placed objects in Monad's navigational path, at positions just outside of the visual field. This resulted in a fairly stereotypic behavioral sequence, beginning with initial visual acquisition, followed by guided visual approach, the establishment of physical contact with the gripper and the sensing of taste. After each trial the object was released and navigation through the environment resumed. Individual trials, from first visual contact to taste sensing, took about 1–2 s of real time, or 6–8 iterations. About 20–25 trials were conducted in a typical experiment lasting 1000 iterations (4–5 min). Fig. 5 shows average activation patterns of neural networks involved in visual sensing of objects, taste, motor action and neuromodulation. In all three sets of panels, a red object was approached, resulting in activation of red-selective units as well as triggering a temporal stimulus representation in

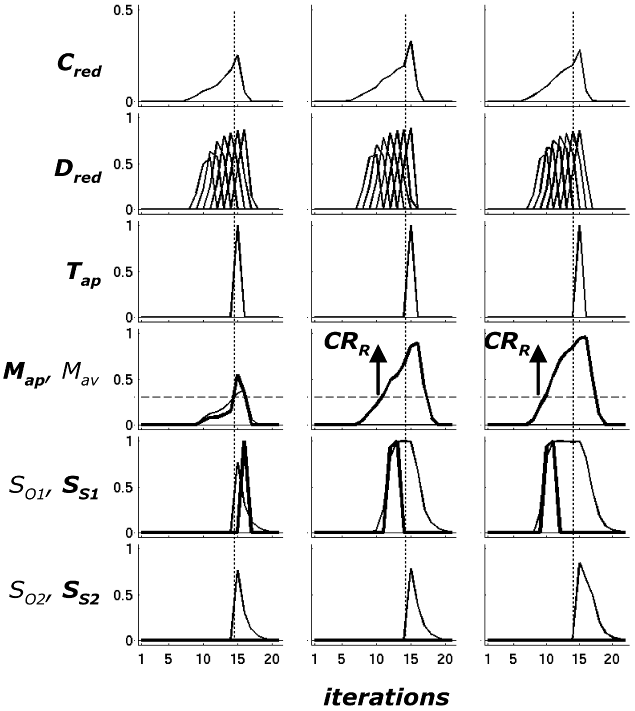


Fig. 5. Development of neural activation patterns for various sensory, motor and neuromodulatory networks. Data shown are from three representative trials with red objects associated with reward (appetitive taste). Panels on the left are recorded before learning, panels in the middle after approximately 500 iterations (≈10 learning trials), and panels on the right were obtained at the end of learning (1000 iterations, ≈20 trials). Top to bottom, plots show neural activity in C_{red}, D_{red}, T_{ap}, $M_{\mathrm{ap}}/M_{\mathrm{av}}$, $S_{\mathrm{O1}}/S_{\mathrm{S1}}$, and $S_{\mathrm{O2}}/S_{\mathrm{S2}}$. M_{ap}, S_{S1} and S_{S2} are shown as thick lines. In all cases, plots show neural activation averaged over all neural units within the designated network. Hatched line in panels for $M_{\mathrm{ap}}/M_{\mathrm{av}}$ indicates the behavioral threshold β ($\beta = 0.3$). A reward-related conditioned response ($\mathrm{CR_R}$) is triggered as soon as $M_{\mathrm{ap}} - M_{\mathrm{av}}$ exceeds β. Dotted vertical line at iteration 15 indicates the onset of appetitive taste.

Dred units. After approach was complete, the object was gripped and the primary reward (appetitive taste) was delivered, here at iteration 15. Before learning (Fig. 5, left panels), motor and neuromodulatory responses were not yet driven by color-selective cells. Instead, the primary reward (T_{ap}) triggered motor (M_{ap}) and $S_{\mathrm{O1}}/S_{\mathrm{O2}}$ activation. S_{S1} became active, reporting the positive temporal difference in S_{O1} due to the reward. S_{S2} activation was cancelled by the arrival of the primary reward. During learning (Fig. 5, middle panels), M_{ap} was activated by selectively strengthened inputs from C_{red}, while M_{av} was competitively inhibited. S_{O1} was triggered, not by the primary reward, but by activity in D_{red} units, as a result of value-dependent modifications of connections between D_{red} and S_{O1}. Thus, S_{S1} became active during visual approach, before objects were physically encountered. Due to long-lasting activation of S_{O1}, it did not become active in response to the primary reward. S_{O2} continued to be activated by the primary reward, but was also partially driven by strengthened connections from D_{red} units whose 'temporal receptive field' coincided with the timing of the reward. S_{S2} remained suppressed due to the actual delivery of the reward. After learning (Fig. 5, right panels), M_{ap} units were strongly driven by C_{red}. S_{O1} and S_{S1} were activated immediately after a red object was visually acquired, with S_{O1} remaining active for a prolonged time period. S_{O2} was driven by both primary reward and timing signals from D_{red}. S_{S2} remained suppressed.

Fig. 6 shows activation levels, averaged over several trials, of the reward component of the neuromodulatory system, as well as average value signals used in synaptic modification. Panels show $S_{\mathrm{O1}}/S_{\mathrm{S1}}$ and $S_{\mathrm{O2}}/S_{\mathrm{S2}}$ activations, as well as VR, both as an average over multiple trials and for individual trials. Before learning, $S_{\mathrm{O1}}/S_{\mathrm{S1}}$ were activated by the primary reward (delivered at iteration 12), accounting for the single spike in the value signal. After learning, $S_{\mathrm{O1}}/S_{\mathrm{S1}}$ were triggered by visual input from C_{red} and D_{red} units. S_{O2} responded to the primary reward in most trials (13 out of 15) and no S_{S2} activity resulted. In two trials, the reward appeared slightly delayed (as measured from the onset of red visual input) and some S_{S2} activation occurred. On average, the value signal consisted of a single spike temporally aligned to the onset of the reward-predicting stimulus (color red). After learning, if the reward was withheld (by presenting red objects to Monad and then quickly withdrawing them as the robot approached), $S_{\mathrm{O1}}/S_{\mathrm{S1}}$ activation was unaffected, but $S_{\mathrm{O2}}/S_{\mathrm{S2}}$ activation was entirely driven by the temporally specific expectation of the reward, mediated by selectively strengthened connections between D_{red} and S_{O2}. This reward expectation signal was not cancelled by an actual reward and thus contributed a negative spike to the value signal. Over time, this negative value signal would lead to weakening of connections between D_{red} and S_{O1}; in addition, connections between D_{red} and S_{O2} would tend to return to baseline, reflecting the change in the consistent timing between the reward-predicting stimulus and the primary reward. If, after learning, the reward was delayed (by pulling the object away from Monad during visual approach), S_{S2} activation occurred at the expected time of reward followed by a second S_{S1} activation to the primary reward once it was delivered. This produced a tri-phasic value signal with an initial positive spike due to the appearance of the reward-predicting stimulus (color red), followed by a negative spike due to reward omission at the expected time, followed by a second positive spike due to the final, now unpredicted, delivery of the primary reward. Over time, if the delay between the onset of the reward-predicting stimulus and the reward delivery changes to a new consistent value, connections between D_{red} and S_{O1} would remain largely unaffected (the predictive nature of the color 'red' is not changed), while connections between D_{red} and S_{O2} would become modified to reflect the new delay time. The negative value spike at the old delay time will disappear and the initially positive value spike at the new delay time will be attenuated and disappear as well.

Fig. 7(A) shows connection weights between temporal representation units D_{red} and $S_{\mathrm{O1}}/S_{\mathrm{O2}}$, obtained from four

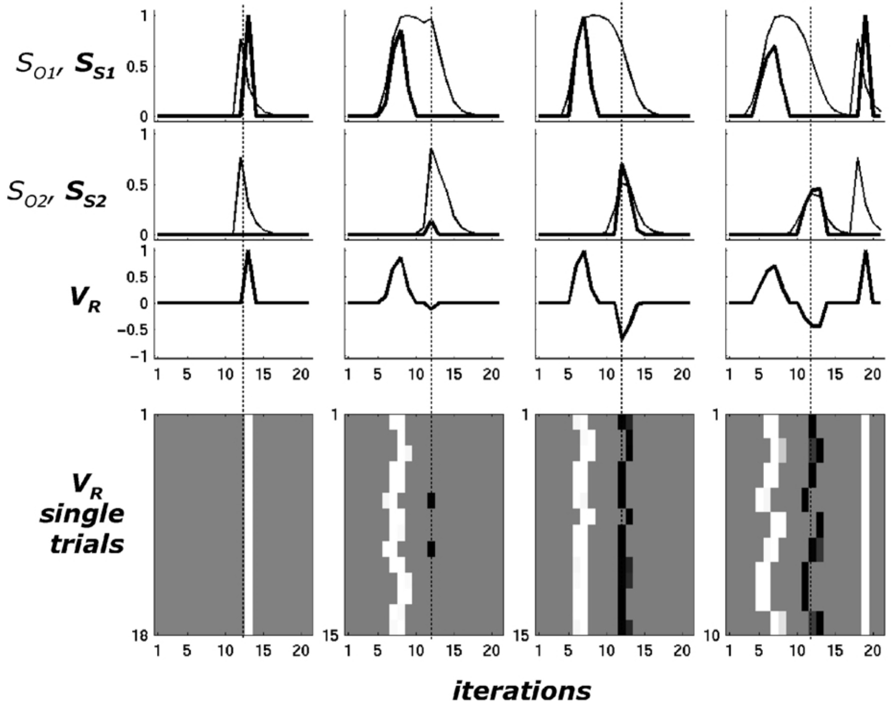


Fig. 6. Activation patterns of neuromodulatory networks, averaged over multiple trials before and after learning, as well as over trials for which a predicted reward was omitted or delayed. Top to bottom, plots show S_{O1}/S_{S1}, S_{O2}/S_{S2}, V_R and a raster plot of V_R obtained from individual trials. Dotted vertical line at iteration 12 indicates the actual or expected onset of appetitive taste.

representative learning experiments. Both sets of weights started at values near zero, but showed characteristic patterns of at the end of learning. The connections terminating on unit S_{O1} first showed strengthening for longer temporal delays (connections numbered 5–7). This is due to the initial phasic activation of S_{S1} by primary reward stimuli and the fact that the value signal derived from this S_{S1} activation was at first temporally coincident

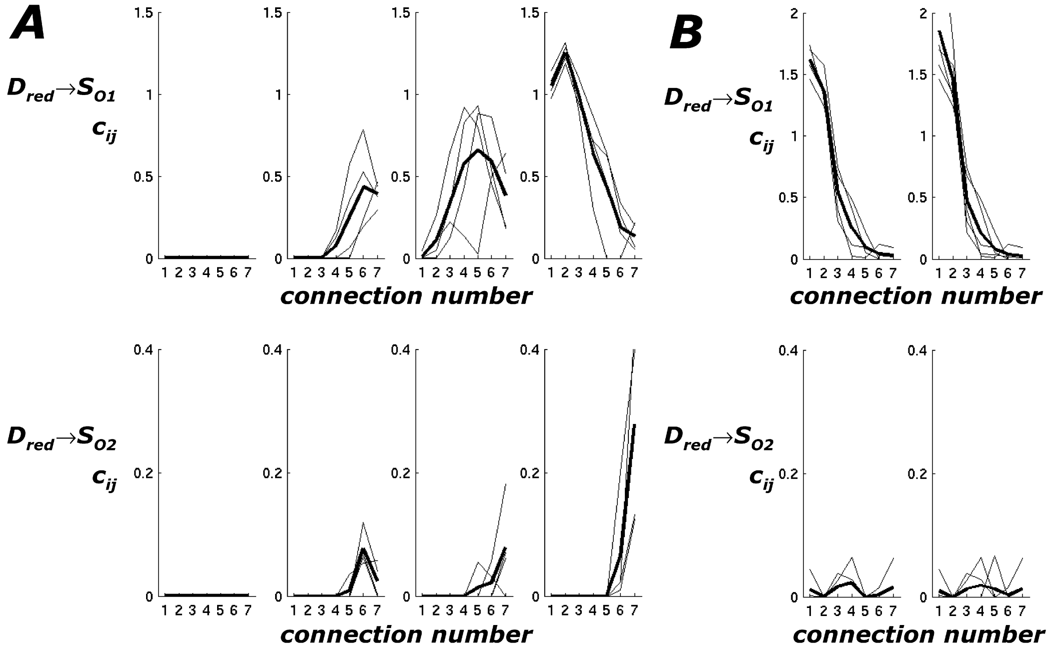


Fig. 7. Average synaptic weights for connections mediating reward-related neuromodulatory responses. (A) Connection weights (numbered 1–7) before, during (250 and 500 iterations) and after learning. Thin lines show weight profile for four individual learning experiments, thick line shows average profile. Top panels are for connections between D_{red} and S_{O1}, bottom panels are for connections between D_{red} and S_{O2}. (B) Weights after learning experiments involving fully autonomous behavior. Top and bottom panels are as in Fig. 7(A). Left panels were obtained after 1000 iterations, right panels after 2000 iterations.

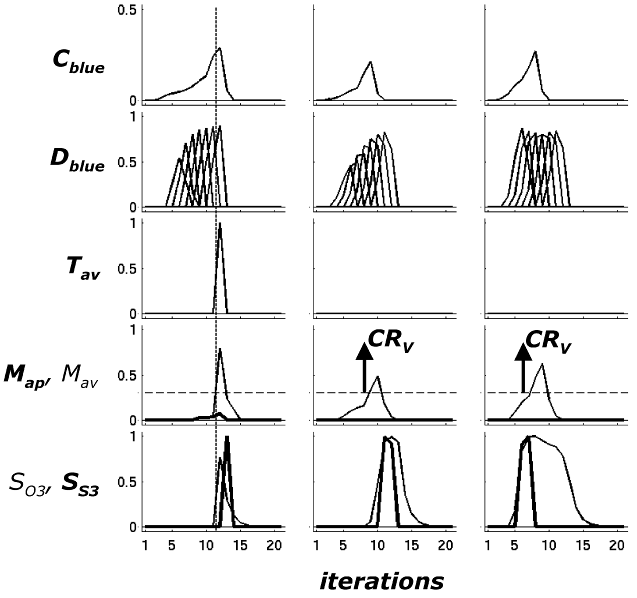


Fig. 8. Development of neural activation patterns for various sensory, motor and neuromodulatory networks. Data shown are from three representative trials with blue objects associated with aversive taste. The experiment was conducted by alternatingly presenting red and blue objects. Panels on the left are recorded before learning, panels in the middle after approximately 1000 iterations (≈ 10 learning trials), and panels on the right were obtained at the end of learning (3000 iterations, ≈30 trials). Top to bottom, plots show neural activity in C_{blue}, D_{blue}, T_{av}, $M_{\mathrm{ap}}/M_{\mathrm{av}}$, and $S_{\mathrm{O3}}/S_{\mathrm{S3}}$. M_{ap} and S_{S3} are shown as thick lines. In all cases, plots show neural activation averaged over all neural units within the designated network. Hatched line in panels for $M_{\mathrm{ap}}/M_{\mathrm{av}}$ indicates the behavioral threshold β ($\beta = 0.3$). An aversive conditioned response is triggered ($\mathrm{CS_V}$) as soon as $M_{\mathrm{av}} - M_{\mathrm{ap}}$ exceeds β. Dotted vertical line at iteration 15 indicates the onset of aversive taste. Note that T_{av} does not become activated if a CSV is triggered.

with the primary reward. As a result of the broad 'temporal receptive field' of D_{red} units, earlier temporal components gradually became capable of driving S_{O1} activation and caused S_{S1} to be active prior to the primary reward. The retrograde transfer of the temporal onset of the S_{O1} response was completed when the earliest onset of C_{red} and D_{red} triggered both S_{O1} and S_{S1} responses. The weight profile shows high synaptic weights for D_{red} units that became active immediately after the onset of C_{red} activity. The connections terminating on S_{O2} were modified using a Hebbian rule, without the modulatory action of the value signal. This ensured that the association between the primary reward and the appropriate temporal delay units remained 'fixed in time'. If the timing between the reward-predicting stimulus and the actual reward was consistent across trials, a weight pattern emerged that drove responses in S_{O2} at the expected time of reward. In Fig. 7(A), consistent timing of visual approach and appetitive taste resulted in the selective strengthening of connections 6 and 7.

3.2. Fully autonomous robot experiments

A critical factor in determining the outcome of robot experiments that do not consist of series of carefully timed trials, but allow fully autonomous behavior and exploration of the environment, was the overall density of objects associated with reward. If this density was low, behavioral sequences involving visual approach and contact with objects were fairly stereotypic, thus preserving the precise timing between first visual input and eventual reward delivery. In such cases, neural activation and synaptic patterns closely resembled those that emerged after manual training (see above). If reward objects, however, were more densely crowded, consistent timing between their first visual appearance and subsequent reward delivery was lost. Many objects were encountered 'suddenly', leading to immediate reward delivery as they were gripped and 'tasted'. Others were lost during approach as other, more salient, targets interfered with visual approach. Over four separate learning experiments (1000 iterations each), we recorded a total of 117 S_{S1} activations, 24 of them not followed by any reward ('lost objects'), 37 the result of unpredicted primary rewards, and the remaining 56 followed by rewards at delays of 1–6 iterations, with a fairly flat temporal profile. Corresponding to these behavioral data, the temporal patterns of the reward-mediating value signal V_{R} were complex and marked primarily by series of positive spikes (unpredicted rewards) while negative spikes are mostly absent (due to the lack of negative prediction errors). Connection weights, obtained after 1000 and 2000 iterations of autonomous behavior, are shown in Fig. 7(B). Weight patterns between D_{red} and S_{O1} resemble those obtained after manual training (see Fig. 7(A)), and were stable after 1000 iterations. However, connection patterns between D_{red} and S_{O2} were flat, reflecting the lack of consistent timing between onset of visual input and reward.

3.3. Reward and aversive conditioning

So far, all experiments have been conducted using only appetitive stimuli (red objects associated with appetitive taste). In order to investigate both reward and aversive neuromodulatory action, we added a second component to the neuromodulatory system (see Fig. 4) and presented red and blue objects, associated with appetitive and aversive taste, respectively. As described in Section 3.1, the experimenter placed objects manually, to ensure consistent timing between visual acquisition and taste. Neural activation patterns for networks associated with reward were very similar to those previously shown in Fig. 5. Fig. 8 shows average activation patterns of neural networks involved in vision, taste, motor action and neuromodulation following presentation of aversive objects. In each of the three sets of plots shown in Fig. 8, a blue object was approached, resulting in activity in C_{blue} as well as D_{blue} units. Before learning (Fig. 8, left panels), the blue object was approached and 'tasted', resulting in brief activations of T_{av}, M_{av}, and S_{O3}. S_{O3} activation triggered a phasic response in S_{S3}, temporally coincident with the delivery of the

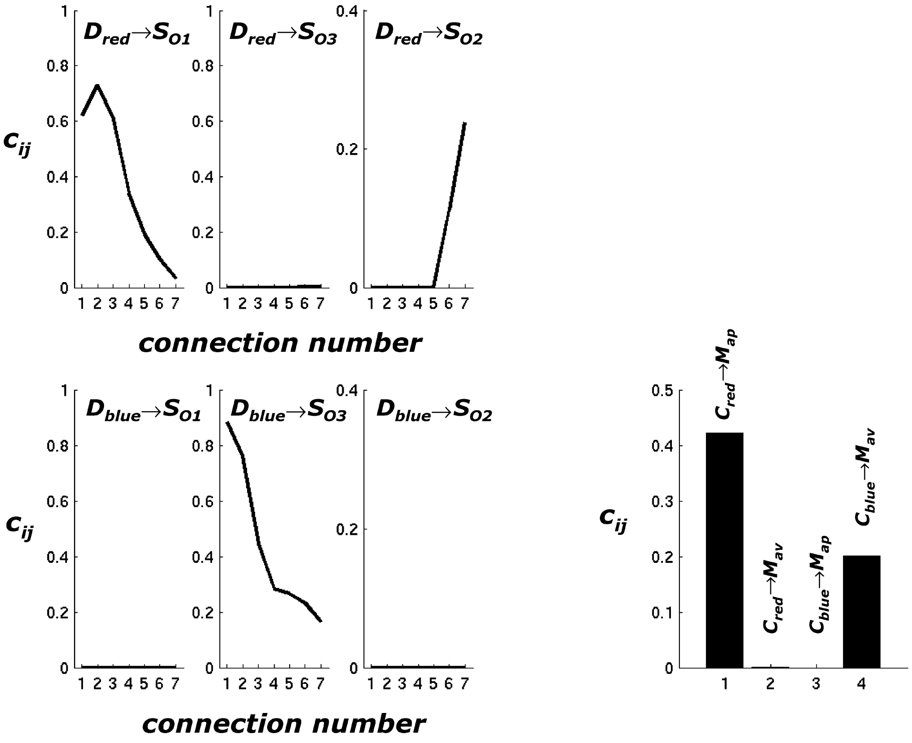


Fig. 9. Synaptic weights for connections mediating both reward-related and aversive neuromodulatory responses. Panels on the left show connections between $D_{\text{red}}/D_{\text{blue}}$ and S_{O1}, S_{O2} and S_{O3} as indicated. Connections are numbered 1–7. Plot on the lower right shows average connection weights between $C_{\text{red}}/C_{\text{blue}}$ and $M_{\text{ap}}/M_{\text{av}}$, as indicated. All data are from one representative run (length 3000 iterations).

primary aversive stimulus. As learning progressed (Fig. 8, middle and right panels), M_{av} was increasingly driven by strengthened connections from C_{blue}, resulting in the triggering of the CR_{V} prior to the completion of visual approach and gripping of the object. After triggering of the CR_{V}, the blue object was actively avoided and removed from Monad's visual field. Also, as a result of value-dependent modification of connections between D_{blue} and S_{O3}, S_{O3} became immediately activated as soon as a blue object entered the visual field and triggered a phasic response in S_{S3}, which was predictive of the expected aversive taste. S_{S3} responses continued to grow and moved retrogradely towards the earliest onset of D_{red}, despite the fact that the primary aversive stimulus was no longer encountered after the CR_{V} was established.

Fig. 9 shows patterns of connection weights between visual delay units $D_{\text{red}}/D_{\text{blue}}$ and components of the neuromodulatory system (S_{O1}, S_{O2}, S_{O3}), as well as motor units capable of triggering conditioned responses (M_{ap}, M_{av}). Integrator units for the two neuromodulatory components mediating reward and aversive conditioning developed strengthened connections with the appropriate set of sensory inputs. In addition, consistent timing between the onset of visual input (color red) and reward resulted in selective strengthening of connections between D_{red} and S_{O2}. Value-dependent modification of sensorimotor connections between $C_{\text{red}}/C_{\text{blue}}$ and $M_{\text{ap}}/M_{\text{av}}$ produced a pattern that allows color-selective units to elicit appropriate conditioned responses.

4. Discussion

Neuromodulatory systems play essential roles in linking behavior and neuroplasticity. In the present computational model, we implemented a biologically based neuromodulatory system in a behaving robot. The model captured several characteristic features of neuromodulatory systems involved in mediating the effects of rewarding and aversive sensory stimuli.

In the present paper as well as in previous work (Friston et al., 1994; Sporns et al., 2000) we conceptualized the functional effects of neuromodulators in plasticity as the action of 'value signals' used in changing synaptic connections. Value signals combine temporal specificity (they are phasic and short-lasting) with spatial uniformity (they affect widespread projection regions and act as a single global signal). Value enters into traditional Hebbian-type synaptic rules as a third factor, in addition to factors representing pre- and postsynaptic activity. Because of their phasic nature, value signals effectively gate plasticity, in addition to influencing its magnitude and direction (see below). Value affects plasticity more or less uniformly throughout the widespread cortical and subcortical regions to which value systems project. These and other properties of value systems are in close correspondence with known anatomical and physiological characteristics of various mammalian neuromodulatory systems, including the noradrenaline and the dopamine systems. Value signals fulfill a dual role in plasticity. They link appropriate sensory and motor units, ultimately resulting in adaptive behavioral change. They also change their own response properties, by modulating sensory afferents to components of value systems. Both of these roles are exemplified in the present model, through changes in sensorimotor connections as well as changes in inputs to the neuromodulatory system itself. Changes in the response properties of neuromodulatory systems underlie the distinction between components of value that are evolutionarily determined (hard-wired), or 'innate', and others that are 'acquired', or experience-dependent.

The temporal characteristics of the reward component of Monad's neuromodulatory system, shown in Figs. 5 and 6, are consistent with their actions in mediating different aspects of reward prediction. S_{S1} acts as a signal for unpredicted occurrences of primary reward or for the appearance of sensory stimuli that predict reward, but themselves are unpredicted. S_{S2} acts as a signal for (negative) errors in reward expectation. Most formulations of temporal difference learning represent the prediction error as a single (positive and negative) first derivative of the reward prediction (e.g. Montague et al., 1996; Suri & Schultz, 2001). Here, we distinguish the positive prediction error (S_{S1}), which is derived as a positive change in the prediction (S_{O1}), and the negative prediction error, which is derived separately through S_{O2}/S_{S2}. S_{S2} activation depends upon the development of specialized coincidence detectors associating a component of a temporal stimulus representation (a specific $D_{\mathrm{red}}/D_{\mathrm{blue}}$ unit) with the timed occurrence of a primary reward. This mechanism produces results that are overtly similar to other formulations of TD learning. In addition, it generates appropriate phasic signals in cases when delivery of the primary reward does not coincide with changes in sensory input (removal of the stimulus predicting the reward). The temporally specific negative prediction error required a different learning rule than the positive prediction error. Hebbian learning was used instead of value-dependent learning, in order to generate and maintain the temporal specificity of the S_{O2}/S_{S2} signal. Value-dependent learning, because of its temporally restricted action at the onset of predictive sensory stimuli, does not seem capable of supporting the generation of temporally specific association units. This result raises the possibility that the neural substrates for positive and negative prediction errors are anatomically and functionally segregated within the midbrain.

Both, reward and aversive components of the neuromodulatory system contained specialized units, which emitted phasic and tonic neural responses. Both types of responses are found in various neural structures and task contexts (reviewed in Suri, 2002; Suri & Schultz, 2001). In reward processing, phasic responses are typically elicited by primary rewards or by reward predicting stimuli. Phasic responses show plastic profiles, with responses disappearing completely for rewards that are fully predicted and instead appearing for reward-predicting stimuli. In our model, S_{S1} and S_{S3} represent reward or aversive stimulus prediction, or, in terms of temporal difference learning, encode positive errors in prediction. S_{S2} phasic activation carries information about whether expected rewards have actually occurred. In the brain, phasic anticipatory neural responses have been identified for many midbrain dopamine neurons (Schultz, 1998), as well as for noradrenergic neurons of the locus coeruleus (Aston-Jones et al., 1997). Tonic activation is observed for the 'integrator' units S_{O1} and S_{O3}, whose activity level after learning remains high for the time period between the occurrence of a predicting stimulus and the actual primary reward or aversive stimulus. Note that it is not necessary for the stimulus to be physically present in order to maintain tonic anticipatory activity (due to the persistent activity caused in the temporal delay units). S_{O1} and S_{O3} activation remains high even after short-lasting exposure to red or blue objects. For tonic anticipatory firing patterns to emerge, the phasic response component must participate (acting as a value signal) in synaptic changes in connections between temporal stimulus representations and the tonic response units. In the case of S_{O3}, the firing level remains high even if the actual aversive stimulus does not occur due to active avoidance by the robot. Tonic anticipatory activity has been observed in parts of the striatum and various cortical areas (Suri & Schultz, 2001). In temporal difference learning, tonic activation is associated with the amount of reward prediction (the temporal difference of which is used as the prediction error signal). In contrast to TD learning, tonic activation of S_{O1} and S_{O3} is primarily the result of neuroplasticity within neural afferents to these structures, which is dependent upon phasic value signals, but it is not primarily 'designed' to represent the prediction of reward accumulation. For example, it is possible for S_{O1} or S_{O3} signals to return to zero before a primary reward is delivered, without a change in S_{S1} or S_{S3} response properties.

The value signal used in this paper is a composite of reward-related and aversive components. It combines characteristics of a reinforcement signal and a saliency signal. A reinforcement signal typically acts to increase or decrease synaptic weights, depending upon the occurrence of a positive or negative prediction error. The reward-related component of the neuromodulatory system implemented in this paper delivers a reinforcement signal, which can be positive or negative. At the same time, positive phasic signals are emitted upon encounter of sensory inputs that are associated with, or predictive of, reward or aversive stimuli. This characteristic of the value signal is consistent with a more general saliency-based function (cf. Redgrave, Prescott, & Gurney, 1999). It seems likely that different neuromodulatory systems of the brain carry different types of signals, related to reinforcement, saliency or novelty. How these different neuromodulatory systems interact is largely unexplored. Possibilities include interconnections between different subcortical regions, or antagonistic or synergistic pharmacological effects at the level of their target regions. In our model, we did not yet incorporate anatomical connections linking the different components and assumed that their effects on synaptic plasticity add linearly. Future computational work is needed to investigate different modes of combinatorial functional coupling between different neuromodulatory systems.

Why implement neuromodulatory systems in robots? Robot learning is an active and rapidly progressing area of research (Nolfi & Floreano, 1999; Schaal, 2002; Sharkey, 1997; Touretzky & Saksida, 1997). In many of these approaches, learning and plasticity depend upon behavioral actions of an agent (animal, robot), which is embedded in an environment. One broad set of functional roles played by neuromodulators is in mediating the effects of behavior on plasticity. This pivotal role of neuromodulation provides a clear rationale for cross-level computational studies incorporating plasticity and behavior. Embodied systems can be studied using actual robots situated in an environment (Sporns, 2002) or using software agents interacting with simulated environments. Implementations of real robots are particularly useful as they allow a direct physical implementation of critical constraints on learning, including more or less complex stimuli, morphology and body structure, sensor properties and mechanics, and temporal dynamics of behavior. Robots are on their way to become sophisticated research tools in psychology and cognitive science (Weng et al., 2001), and they need to have internal mechanisms that can guide their behavioral plasticity and learning. In this context, biologically based computational models of neuromodulatory systems in robots may ultimately serve similar functional roles as their counterparts in animals.

References

Almassy, N., Edelman, G. M., & Sporns, O. (1998). Behavioral constraints in the development of neuronal properties: A cortical model embedded in a real world device. *Cerebral Cortex*, *8*, 346–361.

Aston-Jones, G., Chiang, C., & Alexinsky, T. (1991). Discharge of noradrenergic locus coeruleus neurons in behaving rats and monkeys suggests a role in vigilance. *Progress in Brain Research*, *88*, 501–520.

Aston-Jones, G., Rajkowski, J., & Kubiak, P. (1997). Conditioned responses of monkey locus coeruleus neurons anticipate acquisition of discriminative behavior in a vigilance task. *Neuroscience*, *80*, 697–715.

Aston-Jones, G., Rajkowski, J., Kubiak, P., & Alexinsky, T. (1994). Locus coeruleus neurons in the monkey are selectively activated by attended cues in a vigilance task. *Journal of Neuroscience*, *14*, 4467–4480.

Bear, M. F., Cooper, L. N., & Ebner, F. F. (1987). A physiological basis for a theory of synaptic modification. *Science*, *237*, 42–48.

Bear, M. F., & Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature*, *320*, 172–176.

Bienenstock, E. L., Cooper, L. N., & Munro, P. (1982). Theory for the development of neuron selectivity: Orientation selectivity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*, 23–48.

Doya, K. (2000). Metalearning, neuromodulation, and emotion. In G. Hatano, N. Okada, & H. Tanabe (Eds.), *Affective minds* (pp. 101–104). Amsterdam: Elsevier.

Edelman, G. M., Reeke, G. N., Gall, W. E., Tononi, G., Williams, D., & Sporns, O. (1992). Synthetic neural modeling applied to a real-world artifact. *Proceedings of the National Academy of Sciences USA*, *89*, 7267–7271.

Fellous, J.-M., & Linster, C. (1998). Computational models of neuromodulation. *Neural Computation*, *10*, 771–805.

Friston, K. J., Tononi, G., Reeke, N. G., Jr., Sporns, O., & Edelman, G. M. (1994). Value-dependent selection in the brain: Simulation in a synthetic neural model. *Neuroscience*, *59*, 229–243.

Hasselmo, M. E. (1995). Neuromodulation and cortical function: Modeling the physiological basis of behavior. *Behavioral Brain Research*, *67*, 1–27.

Hasselmo, M. E., & Barkai, E. (1995). Cholinergic modulation of activity-dependent synaptic plasticity in rat piriform cortex. *Journal of Neuroscience*, *15*, 6592–6604.

Hasselmo, M. E., Linster, C., Ma, D., & Cekic, M. (1997). Noradrenergic suppression of synaptic transmission may influence cortical signal-to-noise ratio. *Journal of Neurophysiology*, *77*, 3326–3339.

Hasselmo, M. E., Wyble, B. P., & Fransen, E. (2002). Neuromodulation in mammalian nervous systems. In M. Arbib (Ed.), *Handbook of brain theory and neural networks* (2nd ed). Cambridge, MA: MIT Press.

Jacobs, B. L. (1986). Single unit activity of locus coeruleus neurons in behaving animals. *Progress in Neurobiology*, *27*, 183–194.

Krichmar, J. L., Snook, J. A., Edelman, G. M., & Sporns, O. (2000). Experience-dependent perceptual categorization in a behaving real-world device. In J. A. Meyer, A. Berthoz, D. Floreano, H. Roitblat, & S. W. Wilson (Eds.), (pp. 41–50). *Animals to Animats 6: Proceedings of the Sixth International Conference on the Simulation of Adaptive Behavior*.

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*, 145–163.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive hebbian learning. *Journal of Neuroscience*, *16*, 1936–1947.

Nolfi, S., & Floreano, D. (1999). Learning and evolution. *Autonomous Robots*, *7*, 89–113.

Pennartz, C. M. A. (1996). The ascending neuromodulatory systems in learning by reinforcement: Comparing computational conjectures with experimental findings. *Brain Research Reviews*, *21*, 219–245.

Pfeifer, R., & Scheier, C. (1999). *Understanding intelligence*. Cambridge, MA: MIT Press.

Redgrave, P., Prescott, T. J., & Gurney, K. (1999). Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences*, *22*, 146–151.

Rucci, M., Tononi, G., & Edelman, G. M. (1997). Registration of neural maps through value-dependent learning: Modeling the alignment of auditory and visual maps in the barn owl's optic tectum. *Journal of Neuroscience*, *17*, 334–352.

Schaal, S. (2002). Robot learning. In M. Arbib (Ed.), *Handbook of brain theory and neural networks* (2nd ed). Cambridge, MA: MIT Press.

Scheier, C., & Lambrinos, D. (1996). Categorization in a real-world agent using haptic exploration and active perception. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, & S. W. Wilson (Eds.), (pp. 65–75). *From animals to animats: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*, Cambridge, MA: MIT Press.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*, 1–27.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–1599.

Servan-Schreiber, D., Printz, H., & Cohen, J. D. (1990). A network model of catecholamine effects: Gain, signal-to-noise ratio, and behavior. *Science*, *249*, 892–895.

Sharkey, N. E. (1997). The new wave in robot learning. *Robotics and Autonomous Systems*, 22.

Sporns, O. (2002). Embodied cognition. In M. Arbib (Ed.), *Handbook of* *brain theory and neural networks* (2nd ed). Cambridge, MA: MIT Press.

Sporns, O., Almassy, N., & Edelman, G. M. (2000). Plasticity in value systems and its role in adaptive behavior. *Adaptive Behavior*, *8*, 129–148.

Suri, R. E. (2002). TD models of reward predictive responses in dopamine neurons. *Neural Networks*, *15*, PII: S0893-6080(02)00046-1.

Suri, R. E., & Schultz, W. (2001). Temporal difference model reproduces anticipatory neural activity. *Neural Computation*, *13*, 841–862.

Sutton, R. S., & Barto, A. G. (1990). Time derivative models of Pavlovian reinforcement. In M. Gabriel, & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 539–602). Cambridge, MA: MIT Press.

Touretzky, D. S., & Saksida, L. M. (1997). Operand conditioning in skinnerbots. *Adaptive Behavior*, *5*, 219–247.

Verschure, P. F. M. J., Wray, J., Sporns, O., Tononi, G., & Edelman, G. M. (1995). Multilevel analysis of a behaving real world artifact: An illustration of synthetic neural modeling. *Robotics and Autonomous Systems*, *16*, 247–265.

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., & Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, *291*, 599–600.