ELSEVIER

Brain Research Reviews 21 (1996) 219–245

BRAIN RESEARCH REVIEWS

Full-length review

The ascending neuromodulatory systems in learning by reinforcement: comparing computational conjectures with experimental findings

Cyriel M.A. Pennartz *

Physics of Computation, California Institute of Technology, Mail Code 139-74, Pasadena, CA 91125, USA

Accepted 27 November 1995

Abstract

A central problem in cognitive neuroscience is how animals can manage to rapidly master complex sensorimotor tasks when the only sensory feedback they use to improve their performance is a simple reinforcing stimulus. Neural network theorists have constructed algorithms for reinforcement learning that can be used to solve a variety of biological problems and do not violate basic neurophysiological principles, in contrast to the back-propagation algorithm. A key assumption in these models is the existence of a reinforcement signal, which would be diffusively broadcast throughout one or several brain areas engaged in learning. This signal is further assumed to mediate up- and downward changes in synaptic efficacy by acting as a multiplicative factor in learning rules. The biological plausibility of these algorithms has been defended by the conjecture that the neuromodulators noradrenaline, acetylcholine or dopamine may form the neurochemical substrate of reinforcement signals. In this commentary, the predictions raised by this hypothesis are compared to anatomical, electrophysiological and behavioural findings. The experimental evidence does not support, and often argues against, a general reinforcement-encoding function of these neuromodulatory systems. Nevertheless, the broader concept of evaluative signalling between brain structures implied in learning appears to be reasonable and the available algorithms may open new avenues for constructing more realistic network architectures.

Keywords: Acetylcholine; Dopamine; Long-term potentiation; Memory; Noradrenaline; Supervised learning; Synaptic plasticity; Temporal difference learning

Contents

Abbreviations: ACh, acetylcholine; AChE, acetylcholinesterase; AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid;; ARP, associative reward-penalty; BCM, Bienenstock, Cooper and Munro; ChAT, choline acetyltransferase; CS, conditioned stimulus; DA, dopamine; LC, locus coeruleus; LTD, long-term depression; LTP, long-term potentiation; NA, noradrenaline;; NBM/SI, nucleus basalis of Meynert/substantia innominata; NMDA, *N*-methyl-D-asparate; 6-OHDA, 6-hydroxy-dopamine; SNpc, substantia nigra, pars compacta; TD, temporal difference; US, unconditioned stimulus; VTA, ventral tegmental area.

* Corresponding author. Present address: Graduate School Neurosciences Amsterdam, Netherlands Institute for Brain Research, Meibergdreef 33, 1105 AZ Amsterdam-Zuidoost, The Netherlands. Fax: (31) (20) 696-1006; E-mail: pennartz%nibr@nluug.nl

SSDI 0165-0173(95)00014-3

1. Introduction

One can find almost as many attitudes towards neural network theories of cognitive brain function as there are people working in the field. At one end of the spectrum, sceptics have argued that some of the most basic premises of network theory (e.g., the implementation of learning by synaptic plasticity) are yet to be vindicated by experimental findings. This view holds that the presently available network models are, at best, metaphorical in their imitation of how brain structures learn. At the other end of the spectrum, theorists claim that some network models may indeed capture some essential features of how biological neural nets learn and perform. Despite the recognized paucity of evidence for some basic premises, they maintain that systems level neuroscience can be brought to a higher level of understanding by advancing biologically plausible principles of learning, synaptic interaction and dynamic regulation of neural activity. No matter which attitude is adopted, it is hard to dispute that in computational neuroscience hypotheses on learning are formulated in rigourous mathematical terms, making them less equivocal than non-formal, intuitive approaches.

The present commentary addresses the general computational problem of how does the vertebrate brain utilizes rewarding environmental feedback in associating sensory stimuli with adaptive behavioural responses. More specifically, one previously proposed computational hypothesis poses that chemically distinct 'reinforcement signals' are necessary for efficient behavioural adaptation. We will examine the proposal that ascending neuromodulatory systems in the central nervous system (CNS) may fulfil the role of such reinforcement signals.

2. Supervised learning and reinforcement learning

In neural network theory, two main classes of learning strategies in network models have been distinguished: unsupervised and supervised learning. In unsupervised learning, synaptic weights are commonly changed according to correlations present in the inputs that are fed into a network [22,86,106,113]. Alternatively, some approaches inspired by statistical mechanics implement unsupervised learning by embedding correlative information into an associative memory network by presetting synaptic weights to fixed values [90]. The unsupervised model for development of stimulus selectivity in visual cortex [22,13] has been relatively successful in that a number of its predictions have been partially confirmed by electrophysiological research in the field of long-term potentiation and long-term depression [26,32,55,122,147].

In contrast, supervised learning is characterized by a comparison of network output to an independently determined 'desired output'. The outcome of this comparison results in the feedback of one or more error values (i.e., differences between actual and desired outputs) to the network. These error values instruct locally operating

mechanisms to induce an increase (potentiation) or decrease (depression) of synaptic weights. In artificial neural nets, an entity that calculates error values by comparing actual and desired outputs is referred to as a '*teacher*' and may be envisioned as a separate module operating outside the network. If this external entity feeds back only one global error signal to the entire network, the term '*critic*' is preferred above 'teacher', since the signal is merely *evaluative* and does not contain instructions for weight modifications in inputs to specific neurons [7,86]. The magnitude of the global error signal can be regarded as a 'mark' for the performance of the network, and this magnitude is decisive for the sign and amount of change in strength of all modifiable synapses. Reinforcement learning is usually held apart from supervised learning because the former is a selectional, not an instructional process and does not rely on learning from correct examples. In an instructional paradigm, the difference between the actual and desired activity of a neuron is used as a corrective feedback to that very neuron and commands changes in its synaptic inputs; the feedback is instructive as it contains the information as to how the network *should have* been responding on a neuron-to-neuron basis. In reinforcement learning, however, a network generally starts by sequentially producing many (often randomly generated) output patterns. Patterns resulting in low reward are selected against, because the algorithms are constructed in such a manner that a low-reward signal diminishes the probability that the same output will be generated on subsequent learning trials. The selection process is based on feedback signalling 'good' or 'bad' network performance, not on instructions for improving reponses of individual neurons. A scalar evaluative signal that is broadcast to an entire network during learning will be referred to as a *reinforcement signal*.

In biological neural nets, teaching or evaluative modules have so far not been convincingly demonstrated. However, this by no means implies that instructive or evaluative signalling between structures in the CNS would not take place. An animal in the process of developing behavioural strategies for optimal survival and reproduction often performs actions without a priori knowledge of the beneficial or deleterious consequences of those actions. As a result of the environmental changes caused by the animal's motor output, its brain receives a new set of sensory inputs, e.g., an altered visual perspective, a glimpse of a hidden prey or the taste of fresh meat. These new sensory inputs can be regarded as an environmental feedback which the animal may utilize to improve its performance. Depending on sensory modality, the feedback can be high-dimensional (e.g., complex auditory feedback used by birds to improve their species-specific song) or low-dimensional (e.g., a gastrointestinal sensation signalling nausea). For supervised learning by high-dimensional feedback in biological systems, the reader is referred to Knudsen [105].

This commentary mainly addresses the problem of learning based on low-dimensional environmental feedback, viz. external stimulants of the gustatory, olfactory, gastrointestinal and nociceptive systems. In this context 'low-dimensional' should not be taken to mean that the stimulus necessarily has few recognizable features, but that its impact on learning can be captured by its rewarding quality alone and hence can be represented by a single numerical value. These feedback stimuli are referred to as rewards or primary reinforcers. In the following discussion rewards or reinforcers are taken to include feedback stimuli of either positive (rewarding, appetitive) or negative (punishing, aversive) value. Primary reinforcers may inform an interconnected system of neuronal structures as to how appropriate or inappropriate its preceding motor output was. The environmental feedback does not reveal *how the animal should have been responding* in cases where its performance was far from optimal (except in a binary choice task). Thus, a change in the availability of reinforcers only provides a low-dimensional signal: it contains no specific instructions as to how to improve motor performance. In this sense, the environment can be regarded as a 'critic' emitting crude, evaluative signals to the animals that operate in it. Now that the problem of learning by reinforcement is being cast in this framework, one may wonder what forms of learning are not guided in any way by reinforcement. Although some reward-independent forms of learning such as latent learning and latent inhibition [119] have been identified, it is recognized that explicit or implicit reinforcement is present in most learning tasks for lower mammals. The notion of implicit reinforcement can be illustrated by learning in a Morris water maze, where a rat seeks a platform hidden under the water surface in order to escape a stressful condition, i.e., swimming around in a pool without visible routes to escape. Given the fact that the rat moves around in an aqueous environment, the investigator does not explicitly include a negative reinforcer; yet the animal is clearly motivated to escape to the platform.

The most widely used method of supervised learning in artificial neural nets is back-propagation [33,177,222]. This algorithm is most commonly used for training multilayer feed-forward networks. In this type of network, an input pattern is presented to the first layer of the net, which subsequently sends excitatory or inhibitory signals to one or more hidden (intermediate) layers. The input layer of the net can be thought of as a sensory interface and the units in this layer are 'clamped' to an activity value corresponding to the presented stimulus pattern. The hidden units, in turn, propagate their firing activity towards their synapses contacting units in the output layer of the net. Weight changes are essentially based on a multiplication of the activity of the presynaptic neuron and the error in the activity of an output neuron (calculated as the difference between desired and actual output [86,177]). The term back-propagation refers to the procedure in

which errors are retrogradely propagated in order to compute weight changes in synaptic contacts onto hidden units. Back-propagation of errors is an essential feature of this algorithm, since the sign and strength of any particular synapse, traversed in retrograde direction, is of crucial importance for computing the weight changes in the inputs onto the parent neuron of that synapse. Feedback from the output layer to the hidden layer by orthodromic signalling cannot be utilized, because the feedback pathways do not convey the required information about the weights of the forward synapses [86,177]. Furthermore, it is useful to mention that instead of networks having only one layer of modifiable synapses, multilayer networks are commonly used since they are capable of solving a wider range of computational problems, as pointed out by Hornik et al. [93] and Hertz et al. [86].

It has been recognized for several years that a back-propagation-type algorithm is most likely not implemented by interacting ensembles of neurons in the central nervous system (CNS). Two important arguments against this possibility (see also Crick [43]) are that: (1) the antidromic propagation of signals violates known principles of nerve impulse conduction along axons; (2) an independent teaching module with a priori knowledge of desired outputs of individual neurons is unrealistic. A related problem is the *inflexibility* in learning by backpropagation. Due to the fact that desired outputs are necessarily given beforehand, a network is only able to learn a prespecified task and cannot cope with global changes in output requirements — unless an ad hoc module would translate a global change into altered desired activities for the individual output neurons. Even if one attempts to envision a separate brain area (instead of the environment) as a teaching module, these problems remain as serious as they are.

For these reasons, many modellers of biological neural networks have resorted to a different class of training algorithms, viz. reinforcement learning [8,10,11,86,227]. In this class of models, a scalar reinforcement signal is broadcast to all modifiable synapses in the network. This global signal serves to orchestrate synaptic weight changes throughout the net, but cannot do so in a neuron-specific manner because only one such signal is available for the entire network. Although my review intends to discuss the biological plausibility of reinforcement algorithms in general, two classes of models will be highlighted: *associative reward-penalty* [8,11] and *temporal difference* learning [203,204]; closely related forms are referred to as 'reinforcement comparison' or 'learning with an adaptive heuristic critic' [10,86,204]. In biological network modelling, variations on both themes have been used by various workers, e.g., Dehaene and Changeux [49], Mazzoni et al. [130], Montague et al. [142], Friston et al. [72] and Dayan et al. [47]. Furthermore, the main emphasis of this evaluation will lie on algorithms and the various assumptions attached to them, not on biological or psychological theories of reinforcement.

3. Associative reward-penalty

Associative reward-penalty (ARP) is an important example of reinforcement learning. Like backpropagation, ARP is a general-purpose algorithm that can be used, for instance, to train oculomotor networks on saccadic eye movement tasks or to convert retinal and eye position information into head-centered coordinates [130]. The algorithm operates at the level of *trials*, i.e., it does not spell out how the temporal sequence of events and neural activities during a single learning trial are structured. When translated into terms of real-time behavioural training, one 'pass' of the model is equivalent to a sensory pattern presentation, the action subsequently undertaken by the animal and the ensuing evaluative feedback from the environment, telling the network how good or how bad its action was. In the present discussion, the output pattern of a net is regarded as a motor action or sequence of actions, resulting in environmental feedback. In the version of ARP presented here [11,130], this feedback signal is a scalar reward value r that is broadcast to all modifiable synapses in the network (Fig. 1). The neurons are binary stochastic units, i.e., their activity is either 0 or 1 according to a

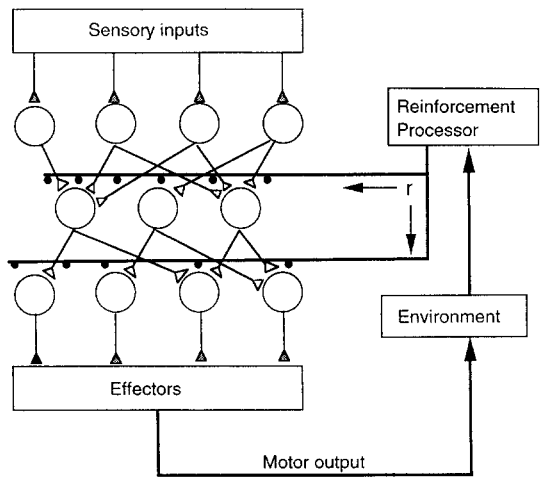

Fig. 1. Learning of sensorimotor tasks by a neural network guided by reinforcing feedback from the environment. A set of sensory inputs is fed into the input layer of the net (upper row of circles; each circle represents a neuron), which subsequently activates an intermediate (hidden) layer. In turn, the intermediate neurons alter the activation state of neurons in the output layer (lowest row), which directly regulates the activity of effectors (muscular contraction, neuroendocrine secretion, etc.). As a result of the effector action initiated by the output layer, alterations in the configuration of environmental variables may lead to a change in the availability of positive or negative reinforcers. Models of reinforcement learning assume that reinforcement is processed by one or more brain areas that broadcast a scalar reinforcement signal (r, or reward value, in the case of associative reward-penalty) throughout parts of the net containing modifiable synapses. Release sites from which r would start diffusing into the neuropil are represented by black dots. In this scheme only input-to-hidden and hidden-to-output synapses are assumed to be modifiable (as indicated by open triangular endings) and hence the reinforcement signal only needs to reach these connections. Grey shaded triangular endings represent non-modifiable synapses.

probability p that is defined by the integrated synaptic input:

$$p_i = \frac{1}{1 + e^{-\beta(\sum_{j=0}^{m} W_{ij} A_j - \theta)}} \quad \text{(Eqn. 1)}$$

in which p_i is the probability of firing for unit i, β the steepness of the sigmoidal activation function, W_{ij} the strength of the synapse from neuron j to neuron i, A_j the activity of neuron j, θ a threshold factor, and m the total number of neurons presynaptic to i. The product of each presynaptic activity value (A_j) and the synaptic weight of one efferent axonal branch of the unit (W_{ij}) determines the contribution of neuron j to the integrated synaptic input ($\Sigma W_{ij} A_j$) received by postsynaptic neuron i. This integrated synaptic input roughly corresponds to the net somatic depolarization caused by all active afferent fibers, whether inhibitory or excitatory. Its magnitude determines the probability of firing of each non-input neuron. According to this probability, a binary activity value is stochastically assigned to each neuron. The use of stochastic variations in neuronal activity in this algorithm is mandatory to guarantee sufficient exploration of the activity space of the network [11].

Following the assignment of binary activity values to the hidden and output units, the error and reward can be computed. The reinforcement signal r is based on a measure for the error (ϵ) in network output:

$$\epsilon = \left\{ \frac{1}{N} \sum_{i=1}^{N} |A_i^* - A_i| \right\}^{1/n} \quad \text{(Eqn. 2)}$$

and

$$r = 1 - \epsilon \quad \text{(Eqn. 3)}$$

where N denotes the number of output units, A_i the activity (0 or 1) of each output unit, A_i^* the desired activity and n a constant relating the sum of errors of individual units to ϵ. Usually r and ϵ vary between 0 and 1, although it is possible to use binary values [11]. Next, the reward value is transmitted to all modifiable synapses in the network in order to enable locally operating synaptic mechanisms to execute weight modifications according to the following equation ('learning rule'):

$$\Delta W_{ij} = \rho\left[r(A_i - p_i) + \lambda(1 - r)(1 - A_i - p_i) \right] A_j \quad \text{(Eqn. 4)}$$

where ΔW_{ij} is the change in synaptic weight from neuron j to neuron i, and ρ and λ are rate constants. For closely related learning rules, see [86].

Let me briefly point out why the learning rule (Eqn. 4) is suitable for training networks on sensorimotor tasks. Firstly, suppose that an untrained, naive network happens to produce a nearly optimal output, yielding a reward $r \approx 1$. This causes the right-hand term containing $(1 - r)$ in Eqn. 4 to be ≈ 0, so that the left-hand term predominates. Units that flipped upwards ($A_i = 1$) will now have their synaptic inputs potentiated (provided that the presynaptic side of the synapse was active and p_i was below 1), whereas those that flipped downwards will have their active inputs depressed. These are desirable modifications because the potentiated synapses will enhance the probability that the active units will be reactivated upon subsequent presentation of the same input pattern, whereas the depressed synapses will prevent the inactive neurons from becoming active. Secondly, suppose that the network performs poorly, resulting in $r \approx 0$. This condition causes the right-hand term of the learning rule to predominate, and this part of the rule is meant to reverse existing states in neuronal connectivity towards neutral conditions. If neuron i has a high probability of firing, its actual firing activity A_i will be mostly equal to 1. Thus, the term $(1 - A_i - p_i)$ will turn negative, causing the active synaptic inputs to that cell to become depressed. Conversely, if $p_i \approx 0$, leading to firing failure ($A_i = 0$) on most trials, the term $(1 - A_i - p_i)$ will approach 1, resulting in a potentiation of active synaptic inputs to neuron i. Thus, consistently low reward leads to depressed firing in previously active neurons and enhanced firing in previously silent neurons. The λ-parameter serves to regulate the relative influence of the state-reversing part of the learning rule with respect to the left-hand term, which is concerned with establishing new connectivity patterns under conditions of moderate to high reward. Although in practice λ assumes values between 0.01 and 0.05, the right-hand part is extremely important for efficient learning since it guards the network against slipping into local minima (non-optimal solutions).

To what extent does ARP embody an advancement towards biological realism as compared to backpropagation? In principle, the algorithm stands up to the objections raised above. Firstly, the assumption of antidromic propagation can be omitted (Fig. 1). Secondly, in ARP the teacher has been reduced to a 'critic' that does not need to have a priori knowledge about the desired activities of *individual* output units. The error can be computed by evaluating an *overall* measure for the motor performance of the network, without the necessity to take activities of single output units into account. Nevertheless, one may also use the sum of differences between desired and actual activity values of individual output units, as in Eqn. 2. The advantage in comparison to back-propagation is that, if one thinks of the 'critic' as an ecological environment providing reinforcing feedback to the organism, interactions of biological CNS networks with the environment do indeed become manifest as global actions, while single unit activities underlying these actions remain unknown to external agents. In other words, in a realistic biological context the state space of neuronal output activity of the CNS is not transparent to whatever entity acts as a critic. Thirdly, reinforcement algorithms are much less vulnerable to the

objection concerning inflexibility than backpropagation. With changing environmental conditions, the function that computes r on the basis of network output can be altered in a natural manner, and this modified output-reinforcement coupling will result in a repatterning of connectivity according to Eqn. 4.

These considerations place ARP within the realm of algorithms that are biologically possible. However, is it also justified to consider them biologically *plausible*? Although ARP has certain non-physiological traits such as the use of binary activity values in a learning process that is specified at the level of trials, our main concern is to address the question of what physiological mechanisms might account for the triadic structure of the learning rule, viz. its Hebbian components (pre- and postsynaptic activity; see below) and the third multiplicative factor (reinforcement signal). Before this identification problem is discussed in further detail, another class of reinforcement algorithms will be briefly reviewed.

4. Temporal difference learning

ARP is especially suited for learning tasks in which a single action directly leads to environmental feedback. Sometimes a learning task can only be completed after a complex sequence of actions aided by sensory processing and navigational computations. Algorithms relying on cumulative instead of instantaneous reinforcement values are able to cope with this class of problems. One of the successful approaches is to let an evaluative module (be it a single unit or an auxiliary network) compute a prediction of future reward, and to use the difference between actual and predicted reward as a signal that is broadcast to all the synapses in the network executing the task. This strategy is illustrated here by the Temporal Difference (TD) algorithm of Sutton and Barto [203,204]; see also [10]). Our first example of TD learning concerns their model of classical conditioning (for related models, see [74,208]; for a general account of the related approach of dynamic programming, see [223]). However, the application of TD learning is not limited to conditioning, as will be illustrated by a model on the formation of oculomotor maps in the developing brain.

An important advantage of the TD model is its operation in real time, avoiding the problem that, in a natural setting, trial boundaries are usually absent or hardly recognizable for the subject. The model of classical conditioning as presented by Sutton and Barto [203] consists of only one postsynaptic neuron receiving two types of input. One synapse is fixed and constant over time and relays information about an unconditioned stimulus (US; Fig. 2A). Other synapses onto the same neuron are modifiable and relay information about one or more conditioned stimuli (CSs). Let it be recalled that in classical conditioning the CS (e.g., a neutral tone) always precedes the US (e.g., tail

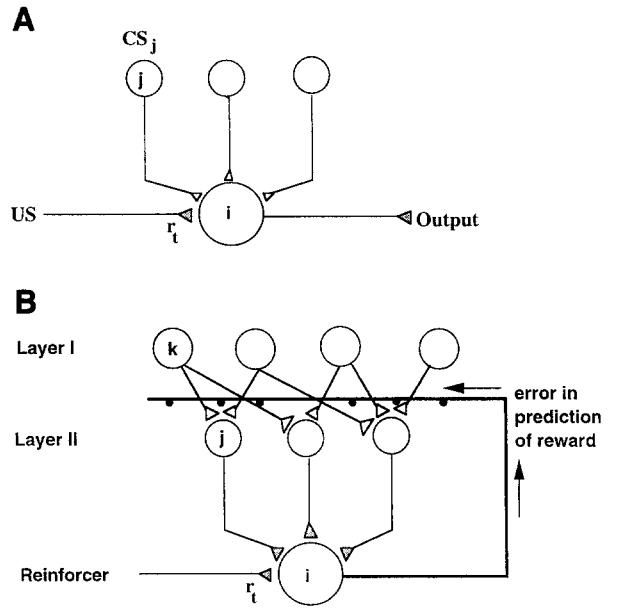

Fig. 2. A: temporal difference learning in a unicellular model for classical conditioning (after Sutton and Barto, [203]). Various conditioned stimuli (CSs, with index j) can be presented in advance of an unconditioned stimulus (US). The US invariably elicits a response (Output) in the postsynaptic neuron (i) by way of a non-modifiable synapse (grey shaded triangular ending) that transmits the reinforcement signal r_t to the postsynaptic neuron at time t. By consistent temporal association between a CS_j and the US, the weight from presynaptic neuron j to the postsynaptic neuron is modified according to the temporal difference learning rule (Eqn. 6, see text). Open triangular endings represent modifiable synapses. B: temporal difference learning in a multicellular model for the formation of oculomotor maps (after Montague et al. [142]). Visual inputs are processed by layer I and II and generate an oculomotor response (not shown). This motor output eventually leads to the delivery of a reinforcer that is processed by neuron i. Note that the converging layer II inputs to neuron i are non-modifiable (gray shaded) whereas those from layer I to layer II are modifiable (open triangular endings). Neuron i propagates its output (the error in the prediction of reward) to its target area, the synaptic interface between layer I and II. This output is assumed to reach all modifiable synapses by means of a diffuble chemical messenger. Its release sites are symbolized by black dots. Neurons in layer I and II are indexed by k and j, respectively.

pinch), which triggers an unconditioned response (e.g., escape). In the course of learning, the CS comes to elicit the unconditioned response. The activity ($A_{i,t}$) of the postsynaptic neuron at time t is computed as the sum of two terms:

$$A_{i,t} = r_t + \sum_{j=1}^{m} W_{ij,t} A_{j,t} \qquad \text{(Eqn. 5)}$$

with r_t, the reinforcement value at time t (in classical conditioning, r_t denotes the strength of the US such as a food reward), $\Sigma W_{ij,t} A_{j,t}$, the integrated CS-related synaptic input to the neuron at time t, and index j denoting one of m presynaptic neurons. $\Sigma W_{ij,t} A_{j,t}$ is clipped at a lower boundary of 0. Furthermore, the weight of the synapse transmitting r_t to the neuron is fixed. Activity values can

be continuous in this algorithm. The learning rule is given by:

$$\Delta W_{ij} = c\left(r_t + \gamma \sum_{j=1}^{m} W_{ij,t} A_{j,t} - \sum_{j=1}^{m} W_{ij,t} A_{j,t-1} \right) \bar{A}_{j,t} \quad \text{(Eqn. 6)}$$

where ΔW_{ij} is the weight change in the synapse from presynaptic neuron j to postsynaptic neuron i, c is a learning rate and γ a discount rate, with $0 < \gamma < 1$. $\bar{A}_{j,t}$ is the activity trace of presynaptic neuron j at time t and is computed according to:

$$\bar{A}_{j,t} = \beta \bar{A}_{j,t-1} + (1 - \beta) A_{j,t-1} \quad \text{(Eqn. 7)}$$

The activity trace represents a running average of the presynaptic activity and serves to retain this activity until the reward signal r_t has arrived at neuron i. Recall that the CSs are always initiated before the onset of the reward signal. The constant β regulates the retention of previous activity in presynaptic neurons.

The activity of the postsynaptic neuron ($A_{i,t}$) is the actual output of the model and encodes both the conditioned and unconditioned response in this learning paradigm. Instead of pointing out how experimental findings on classical conditioning can be mimicked by this unicellular model [203], I will focus here on how the learning rule (Eqn. 6) functions in practice. The central concept in TD-learning is that the temporal difference between two predictions made by a neural system can be used for directing changes in synaptic efficacy. In the model of Fig. 2A, the integrated synaptic input from the neurons relaying CSs ($\Sigma W_{ij,t} A_{j,t}$) can be viewed as a prediction about the total amount of reward delivered in the future. The total amount of reward can be represented by the sum of all expected rewards at each time step in the future. Then $\Sigma W_{ij,t} A_{j,t-1}$ represents the system's prediction about all rewards after time $(t-1)$, including the reward r_t at the next time step. $\Sigma W_{ij,t} A_{j,t}$ represents the prediction about the rewards after t. If the system has been trained to perfection, then the early prediction $\Sigma W_{ij,t} A_{j,t-1}$ matches the sum $(r_t + \Sigma W_{ij,t} A_{j,t})$ exactly, so that the system correctly anticipates the upcoming delivery of r_t. The temporal difference is 0 and no learning occurs. In an untrained situation, the early prediction $\Sigma W_{ij,t} A_{j,t-1}$ usually differs from the sum $(r_t + \Sigma W_{ij,t} A_{j,t})$, because r_t is not an expected but an actual reward value, not yet known at time $(t-1)$. Thus, the quantity $(r_t + \gamma \Sigma W_{ij,t} A_{j,t} - \Sigma W_{ij,t} A_{j,t-1})$ can be understood as the *error in the prediction of reward* . For a full account of the discount rate γ, typically about 0.95, the reader is referred to [203]. Basically, γ weighs the importance of temporally near versus distant rewards. In classical conditioning, γ serves to attribute a stronger associative capacity to CS–US pairs that follow closely in time as opposed to temporally distant pairs.

Let us next consider an untrained model which is exposed to a close temporal CS–US pairing. As illustrated in Fig. 2, the CS excites presynaptic neuron j, thereby causing a slowly rising and decaying trace $\bar{A}_{j,t}$ in the efferent projection from j to postsynaptic neuron i. Once the US is presented, r_t will cause the postsynaptic activity $A_{i,t}$ to exceed 0 (Eqn. 5). The other components contributing to $A_{i,t}$ are still small ($\Sigma W_{ij,t} A_{j,t} \approx 0$). If $\bar{A}_{j,t} > 0$, the weight change given by Eqn. 6 will be positive, since $r_t > 0$ and both $\gamma \Sigma W_{ij,t} A_{j,t}$ and $\Sigma W_{ij,t} A_{j,t-1}$ are nearly 0. Thus, a close temporal CS–US relationship results in a positive weight change reflecting conditioned association. In the course of further training, the CS-induced prediction will approximate the upcoming US to an increasing degree and thus the error in the prediction of reward decreases, causing synaptic modification to slow down.

Although a major application of TD learning lies in modelling classical conditioning, it is worth mentioning a number of other tasks for which this algorithm has been proven useful. In a forerunner of TD learning, Barto and Sutton's model [9] of spatial navigation, a small associative network learned to move towards a central target by sensing the 'odors' diffusing from landmarks that surrounded a target. Barto et al. [10] applied TD-learning in a cart-pole balancing task, using a simple system consisting of a decoder (relaying cart-pole state variables such as pole angular velocity to the model), an adaptive critic element (the prediction-reward comparing unit) and an associative search element that directed movement of the cart-pole system. Training by TD-learning proved to be faster and final performance much better than in a model lacking a prediction-reward comparing element. This indicates the usefulness of this learning method in at least one important class of instrumental conditioning tasks. Dayan et al. [47] have recently used TD learning for simulating how bees associate flower colors with nectar reward in developing foraging strategies.

It is possible to apply TD-learning to multicellular models if a postsynaptic activity term is added to the learning rule (cf. [9]). In these multicellular models, synaptic modifications need not be restricted to the neuron that receives both US and CS inputs (Fig. 2A; Eqns. 5 and 6). Montague et al. [142] modified the Sutton–Barto approach for classical conditioning to simulate the development of oculomotor maps in the vertebrate brain. Briefly, the reward-processing neuron of Fig. 2A is preserved in this model, but this neuron is now envisaged to broadcast a diffusible signal throughout its target areas (e.g., cortex; Fig. 2B). These target areas may be multi-layered and produce a particular motor output (not represented in Fig. 2B). The lowermost layer has projections to the reward-processing neuron. The weights from this projection (denoted by indices j to i) are kept constant now, whereas the weights from layer I to II (indices k to j) are the ones to be modified. Thus, the modifiable CS-related inputs to the reward-processing neuron (Fig. 2A) have been replaced by non-modifiable inputs from layer II (Fig. 2B). Improve-

ment in reward prediction occurs by modifications of synapses from layer I to II. In addition to the pre- and postsynaptic activity at these synapses, the reinforcement-related signal broadcast by the reward-processing neuron is of paramount importance in mediating synaptic weight changes and, again, corresponds to the error in the prediction in reward:

$$\Delta W_{jk} = c\left(r_t + \gamma \sum_{j=1}^{m} W_{ij,t} A_{j,t} - \sum_{j=1}^{m} W_{ij,t} A_{j,t-1} \right) A_{k,t} A_{j,t} \qquad \text{(Eqn. 8)}$$

with ΔW_{jk}, the weight change in the synapse from neuron k (layer I) to neuron j (layer II), $A_{k,t}$ the presynaptic activity and $A_{j,t}$ the postsynaptic activity. Depending on the precise implementation of the model, one may use activity traces (Eqn. 7) instead of instantaneous values (cf. [10]).

At any time t, layer I receives a visual input and drives layer II to generate oculomotor output as well as a prediction about future rewards. The predictive value is passed on to the reward-processing unit i by the W_{ij} synapses, subjected to the differentiation-like subtraction of the previous prediction and summed with the current r_t value. This reward value is not necessarily derived from a consumptive object, but may correspond to proprioceptive feedback from the ocular muscles. If the motor output of an untrained network happens to produce a reward r_t of considerable magnitude, the reinforcement signal $(r_t + \gamma \Sigma W_{ij,t} A_{j,t} - \Sigma W_{ij,t} A_{j,t-1})$ will be large and positive since the early and late predictions $\gamma \Sigma W_{ij,t} A_{j,t}$ and $\Sigma W_{ij,t} A_{j,t-1}$ are still close to 0. Thus, the positive error in predicted reward will lead to potentiation of the active synapses onto layer II cells. Gradually, an efficient sensorimotor mapping between visual input and oculomotor output emerges, accompanied by smaller errors and less modifications. Note that the pre- and postsynaptic activities of layer I and II neurons are not directly affected by the reinforcement signal.

In summary, for multicellular systems the learning rules of the TD and ARP algorithms are both characterized by the multiplication of three factors. In addition to the Hebbian product of terms related to pre- and postsynaptic activity, the reward r guides synaptic weight changes in ARP, whereas $(r_t + \gamma \Sigma W_{ij,t} A_{j,t} - \Sigma W_{ij,t} A_{j,t-1})$ constitutes the third multiplicative factor in TD learning. This factor, like r, is a reinforcement signal that is assumed to be broadcast throughout systems engaged in learning. The essential properties of ARP and TD learning are compared in Table 1. Before we start searching for possible physiological correlates of both ARP- and TD-type of rules, it is important to delineate what functional requirements should be fulfilled by a neurochemical system emitting a reinforcement signal.

5. Gating, biasing and reinforcement signals

First, it is crucial to point out in what sense ARP- and TD-algorithms fundamentally differ from Hebbian learning rules. A Hebbian rule is commonly defined as having the following general form:

$$\Delta W_{ij} = A_j^* f(V_i) \qquad \text{(Eqn. 9)}$$

with ΔW_{ij}, synaptic weight change, A_j, firing activity of the presynaptic neuron and V_i, membrane potential of the postsynaptic neuron. The function $f(V_i)$ defines how the postsynaptic membrane potential influences weight changes and is referred to here as a plasticity curve (Fig. 3). In their model of the development of stimulus selectivity, Bienenstock, Cooper and Munro (BCM [22]) use a biphasic plasticity curve, with relatively negative membrane potentials giving rise to synaptic depression and strong depolarizations to potentiation. A threshold θ marks the membrane potential where weight changes switch from depression to potentiation. Instead of V_i, causally related parameters such as postsynaptic calcium concentration can be used as the independent variable [4,13,26,32,115]. Conversely, Hebbian learning can be cast into a more abstract framework, such as the statistical correlation rule of Kohonen [106].

It is important to note that this definition of Hebbian learning implies that the sign of weight change ΔW_{ij} is exclusively determined by $f(V_i)$, since $A_j \geq 0$. Furthermore, weight changes are determined by only two multiplicative variables. These two features distinguish Hebbian learning from the reinforcement algorithms described

Table 1
Comparative overview of two main reinforcement learning algorithms

	Associative reward-penalty	Temporal difference learning
Feedback signal	reward	error in reward prediction
Activity of units	binary stochastic (0 or 1)	continuous range, from 0 to 1
Learning rule	multiplicative:	multiplicative:
	– presynaptic activity	– presynaptic activity
	– postsynaptic factor	– postsynaptic activity
	– reward	– error in reward prediction
Specification of time	trial-by-trial	real time

Features of temporal difference learning refer to the multicellular version explained in the text.

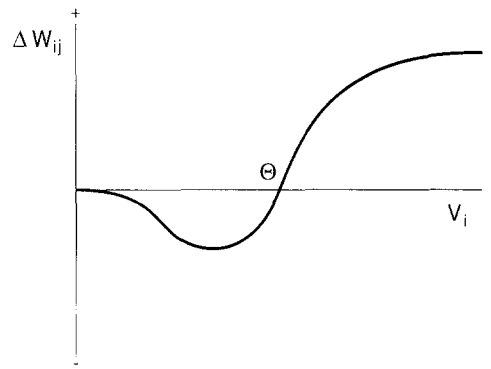

Fig. 3. Hypothesized relationship between the postsynaptic membrane potential (V_i) and the amount and sign of synaptic weight changes. Presynaptic terminals are presumed to be fully active (see Eqn. 9; after Bienenstock et al. [22]). This relationship, also referred to as a plasticity curve ($f(V_i)$), exemplifies the general form of Hebbian learning rules. ΔW_{ij}, change of synaptic weight between presynaptic neuron j and postsynaptic neuron i; Θ, transition point between depression (−) and potentiation (+) of synaptic weights.

above. In ARP, the sign of weight change is determined by two out of three multiplicative factors, viz. the postsynaptic factor (depending on A_i and p_i in Eqn. 4) and the reward r. In the multicellular version of TD-learning, weight changes are also computed by multiplying three factors, and the sign of weight change is exclusively determined by the error in prediction of reward, $(r_t + \gamma \Sigma W_{ij,t} A_{j,t} - \Sigma W_{ij,t} A_{j,t-1})$, if at least all activity-values are taken to have a lower bound of zero. In both algorithms, the reinforcement-related factor assumes the same value for all synapses of the network.

In the context of Hebbian learning, two types of modulatory signal can be defined. A *gating* or permissive signal ν is defined as a variable scalar between 0 and 1 that allows synaptic weights to change without altering presynaptic activity A_j, postsynaptic membrane potential V_i or plasticity curve $f(V_i)$:

$$\Delta W_{ij} = \nu^* A_j^* f(V_i) \qquad \text{(Eqn. 10)}$$

with $0 < \nu < 1$. It may not be immediately obvious why a gating signal differs from a reinforcement signal. What they have in common is that they are not synapse-specific, since their value is identical throughout the network. However, the constraint that a gating signal assumes values between 0 and 1 prevents it from influencing the sign of weight change, and thus distinguishes it from a reinforcement signal as defined above. This may seem only a subtlety, but the essential point is that reinforcement signals in the sense of Eqns. 4 and 8 should be able to steer synapses towards depression if performance is bad (or worse than previously) and towards potentiation in the opposite case. In contrast, a gating signal merely *enables* mechanisms for changing synaptic weights, without discriminating between depression or potentiation.

The second subclass of modulatory signals is termed *biasing* signal here and comprises all modulators that do not act as a signed or unsigned multiplicative factor in the learning rule, but affect weight changes by altering the presynaptic activity A_j, postsynaptic membrane potential V_i, firing activity or the threshold setting and shape of the plasticity curve $f(V_i)$. Some biasing signals may be capable of affecting the sign of weight change without acting as a separate, multiplicative factor in the learning rule. For instance, a neuromodulator that closes a leaky postsynaptic K^+ conductance may lift the membrane potential above the depression–potentiation transition point θ within a BCM-type of framework. Even though such signals affect the sign of weight changes, they are distinct from reinforcement signals as the learning rule cannot be written as a product of three independent variables (i.e., reinforcement signal, pre- and postsynaptic activity). The concept of gating and biasing signals can be applied to many types of network, e.g., associative memories [83], feature-extracting feed-forward networks [113] and Kohonen maps [106].

To recapitulate, a reinforcement signal is a scalar that multiplies factors relating to pre- and postsynaptic activities in the learning rule and strongly affects the sign of weight changes. Gating and biasing signals are far less constrained in that they constitute or affect one or more variables in the learning rule, but are not be represented by a multiplicative factor that modulates weight changes bidirectionally. Now that these classes of putative feedback signals have been defined, expected physiological and anatomic properties of reinforcement signals will be deduced from these theoretical considerations.

6. Predicted physiological and anatomic properties of reinforcing systems

In biological network models based on ARP or TD-learning, it has been hypothesized that a distinct neuroactive agent, various candidates of which are treated below, fulfils the role of a reinforcement signal. This signal is assumed to reach each modifiable synapse in a neural system engaged in learning. Glutamatergic synapses, which are nearly ubiquitous throughout the vertebrate CNS, would not be suited to implement this interaction, since their activity is expressed mainly as AMPA/kainate receptor mediated EPSPs in the postsynaptic cell, supplemented with a voltage-dependent NMDA receptor mediated component [87,102,141]. The main effect of AMPA/kainate receptor activation is a fast, transient depolarization of the postsynaptic membrane, which has two major implications. (1) Glutamatergic synapses affect the postsynaptic parameter in learning rules to a major extent and thus violate the assumption that the postsynaptic activity is set independently of the amplitude of the reinforcement-related signal. This is a general problem for candidate reinforcement signals and is referred to as *interference with pre- or postsynaptic activity*. (2) Due to dendritic cable filtering

effects, the depolarization caused by glutamatergic synapses relaying a reinforcement signal may not be felt equally by all modifiable synapses on the same postsynaptic cell. This exemplifies the general problem of *unequal access*, which arises whenever a reinforcement signal does not reach all modifiable synapses in a network to the same extent. The TD and ARP learning rules assume that synapses having equal pre- and postsynaptic activities will be equally potentiated or depressed given a single reinforcement value. However, an inhomogeneous distribution of glutamatergic synapses signalling reinforcement across dendritic trees would cause some modifiable synapses to undergo larger changes than others with equal activity.

Modellers have therefore sought to confirm that diffusible modulatory substances (presumably acetylcholine, norepinephrine and/or dopamine) act as a reinforcement signal [32,130,137,142,143,226]. The problem of unequal access would be remedied if these substances diffuse freely throughout the extracellular space and reach homogeneous concentrations at distributed synaptic sites. It should be noted that the assumption of a homogeneous distribution (and thereby equal access) is a demanding one, since neuromodulator release sites, be it unconnected axonal varicosities or terminals on postsynaptic elements, act as point-like sources, while uptake and breakdown mechanisms surrounding them can be very powerful. Diffusibility does not guarantee a solution to the first mentioned problem, viz. interference with pre- or postsynaptic activity. Furthermore, it is important to point out that a reinforcement system should be able to direct plastic changes in multiple target areas cooperating in the execution of a given task. Our current understanding of behavioural adaptation is that multiple brain structures, connected serially and/or in parallel, are engaged even in relatively simple learning processes such as adaptation of the vestibulo-ocular reflex [114,191], eyeblink conditioning [191,210], fear conditioning [112] and incentive motivational learning [17,158,179]. Moreover, many complex problems (e.g., not linearly separable tasks) require synaptic changes in multiple interacting brain areas or layers [86]. Therefore, it seems reasonable to exclude systems that can release diffusible substances only in a restricted target area (e.g., enkephalins and dynorphin [2]). Furthermore, the locally diffusible messenger NO is excluded from this discussion because it is presumably emitted from postsynaptic cells instead of distinct afferent synapses, thus essentially operating in a Hebbian framework of plasticity. Henceforth we will focus on the ascending cholinergic and catecholaminergic systems, which do indeed reach numerous distributed brain sites.

I should briefly note why the histaminergic and serotoninergic systems are not considered here as candidate reinforcement signalling substances, even though they fan out extensively across cortical and subcortical areas. Histaminergic fibers, originating in the hypothalamic tuberomammilary nucleus, regulate autonomous functions such as hormone secretion, cardiovascular function, emesis, motion sickness, glycogenolysis and body temperature [184,216]. In addition, histamine functions in controlling arousal and responsiveness to noxious stimuli have been suggested, but there are no convincing indications for a specific role in learning. As for serotonin, the firing activity of dorsal raphe neurons in behaving cats was found to correlate to the level of arousal and tonic motor activity, in agreement with the widely recognized role of serotonin in regulating sleep-waking cycles [99,100]. Although the functional ramifications of serotonin extend into the domain of nociception, mood and affective states, research in this area has not raised sufficient evidence to suspect a specific function in learning by reinforcement.

At least three other requirements must be met in order for a neuroactive substance to be classified as a candidate reinforcement signal. In vivo unit recordings from neurons that synthesize the putative reinforcing substance should reveal a *one-to-one relationship* between a reinforcement-related parameter and firing activity. If a candidate group of brain neurons fires in relation to many sensory stimuli, including generally arousing, appetitive as well as stressfull stimuli, it is safe to conclude that this firing activity does not encode an evaluative signal in the sense outlined above. Indeed, it would have disastrous consequences if a neuron transmitting a positively reinforcing signal to a target network would fire in response to non-reinforcing stimuli, e.g., the sight of a novel or distressing object. Further evidence may originate from neurochemical studies assessing how release of the substance changes as a function of environmental contingencies. For ARP-like algorithms, unit firing is expected to reflect the reinforcing value of environmental feedback, whereas TD-learning requires a correlation to the error in the prediction of reward. Since it is quite conceivable that other, as yet undefined algorithms would rely on different reinforcement-related parameters, it stands to reason that discrepancies between experimentally observed neuronal activity and the behaviour predicted by a particular algorithm should not be interpreted as conclusive evidence against the general concept of reinforcement-related signals. However, it is essential to realize that one should be able to reduce the observed correlations to a one-to-one relationship, not many-to-one.

A further requirement can be formulated from a behavioural-pharmacological viewpoint. Blockers of receptors that bind reinforcement signals should be able to disrupt reward-dependent learning tasks. Similarly, depletion and lesion studies of the neurochemical system under consideration are expected to reveal deficits in learning. For some systems, uptake blockers and substances that potentiate release of the signalling substance are available, allowing complementary studies of how task acquisition changes when the concentration of the substance at postsynaptic sites is enhanced.

Reinforcing systems must also meet the demand of temporally asymmetric action, as motor actions preceding reward should be reinforced but not those following reward. Similarly, only conditioned stimuli preceding a US are useful for learning to predict, not those following it [143]. Unfortunately, this requirement has not yet been experimentally tested as far as the plasticity-modulating capabilities of NA, ACh and DA is concerned, precluding evaluation of these systems on this point. However, the difficulty of fulfilling this predicted property can be appreciated by noting that most of the neuromodulatory systems considered below operate by activating or inhibiting second messenger systems (cf. [226]). These changes may decay on a time scale in the order of seconds and could easily entail undesirable effects on post-reward, motor-related neural activity.

Finally, all of the above hinges on the fundamental premise that learning processes are substantiated by changes in synaptic efficacy, not by less specific alterations in postsynaptic parameters. At this point it is still too early to decide which physiologically identified forms of synaptic plasticity are truly relevant for learning. Nevertheless, it will be instructive to review briefly how various neuromodulators affect LTP and LTD. These two forms of synaptic plasticity have been found in multiple brain structures [4,28,89,104,122,157], have some reproducible properties [26,104,147] and there is at least some evidence for their involvement in learning and memory [75,111,144,145,189]. It would be premature, however, to refute or accept any model of reinforcement learning solely by arguments drawn from LTP/LTD studies.

In summary, the ARP and TD algorithms outline a general class of models for reinforcement learning. These models predict that a corresponding neural system should have a set of specific physiological and anatomic characteristics. However, some of the predicted characteristics are more critical for assessing the plausibility of these models than others. If a candidate reinforcing system in the brain fails to meet a certain requirement, it may be possible to evade such a requirement by ad hoc adjustment of the model. Although it is impossible to foresee precisely which characteristics may be amenable to ad hoc modification, I can tentatively indicate how critical each prediction is:

1. The reinforcement signal must be highly diffusible (equal access; very critical);
2. The brain structures emitting the reinforcement signal must have wide ramifications throughout large areas of the CNS so that learning in multiple cooperating structures can be coordinated (moderately critical);
3. The reinforcement signal should not heavily interfere with pre- and postsynaptic activity (moderately critical);
4. The reinforcement signal modulates synaptic weight changes bidirectionally (moderately critical);
5. The firing activity of neurons emitting reinforcement signals, and the ensuing release of neuromodulator, must be tightly correlated to the reinforcement-related parameter figuring in the learning rule (very critical);
6. Pharmacological manipulation of the reinforcing system must specifically disrupt learning as assessed in behavioural studies (very critical);
7. The effect of reinforcement signals on learning sensorimotor systems must be asymmetric in time to ensure that only preceding motor actions are subject to evaluation and modification (very critical).

A full-length review of all of the above mentioned functional properties of the ascending modulatory systems is clearly outside the scope of this discussion. The reader will be referred to other reviews specialized in a particular subject that can be only briefly discussed here.

7. Noradrenaline

The hypothesis that noradrenaline (NA) acts as a reinforcement signal will be evaluated according to the predicted functional properties outlined in the previous section.

7.1. Equal access

It has been well established that many CNS structures, including neocortex, hippocampal formation, striatum and thalamus are innervated by NA fibers originating in the locus coeruleus (LC), located in the pontine brainstem. Early findings by Beaudet and Descarries [15] indicated that only about 5% of the axonal NA-positive varicosities in cortex are clearly associated with postsynaptic neurons, fuelling suggestions that NA may diffuse from its release sites to remote receptors. The widespread fan-out of NA fibers throughout both superficial and deep layers of the neocortex, combined with a 'paracrine' mode of operation, would allow NA to have good access to pre- and postsynaptic target sites [53]. It is therefore not surprising that network modellers have considered NA to be a candidate evaluative signal that can be broadcast throughout an entire network. However, more recently it has been recognized that NA fibers form more conventional synaptic contacts than originally thought [152,154]. Thus, at this point it is still uncertain whether NA diffusion from release sites is functionally significant and would allow NA to have equal access to all modifiable synapses within networks engaged in learning.

7.2. Interference with pre- or postsynaptic activity

System physiologists have suggested a function of NA in enhancing the signal-to-noise ratio of cortical neurons, taking the background activity of a neocortical cell as a measure for 'noise' and a stimulus-evoked response as the signal [68,218]. In in vitro slice studies, various effects of NA on postsynaptic excitability have been reported, a

prominent example being the β_1-receptor mediated reduction of the Ca^{2+}-dependent K^+ conductance that regulates spike adaptation in neocortical, thalamic and hippocampal neurons [78,120,149,155]. Without going into further detail, let it suffice here to stress that further examples of NA-mediated effects on postsynaptic conductances are amply available for many telencephalic structures (e.g., [76,149]). Although these modulatory effects by itself do not disprove an additional function of NA as some type of reinforcement signal, they do indicate that NA fails to fulfil the prediction that it acts as a multiplicative factor in learning processes that leaves pre- or postsynaptic parameters untouched.

7.3. Influence on synaptic plasticity

There is no shortage of studies reporting NA effects on long-lasting forms of synaptic weight modifications. Bliss et al. [25] and Stanton and Sarvey [192] reported impairment of LTP induction in the dentate gyrus after NA depletion. Furthermore, Stanton and Sarvey [193] reported a NA-induced LTP effect in the medial perforant path in the absence of high frequency stimulation. In contrast, lateral perforant path responses in the dentate gyrus can be depressed by NA for prolonged periods of time [45]. In hippocampal field CA3, NA has been reported to facilitate LTP induction [91]. Finally, Heginbotham and Dunwiddie [84] observed a β-receptor induced long-lasting enhancement of the population spike in field CA1 of hippocampus without any change in the field EPSP. While these and other findings certainly argue for a modulatory role for NA in synaptic plasticity in general, the relevant point for our purposes is that in fact none of these experiments raises support for NA as a reinforcement signal in the sense of Eqns. 4 and 8. Both algorithms for reinforcement learning predict that the magnitude of the reinforcement signal (presumably corresponding to the local concentration of NA at the modifiable synapses) is of paramount importance in determining the sign of weight change. Without an experimental confirmation of this critical prediction, the current body of evidence on LTP and NA is more compatible with a less constrained function of NA in synaptic plasticity such as a biasing or gating signal.

7.4. In vivo unit recordings

This field of research poses the most serious problem for a NA-reinforcement hypothesis. It is commonly agreed that locus coeruleus neurons in unanaesthetized mammals can respond to a range of external events, including novel, aversive as well as rewarding stimuli [5,6,67,68,98]. A common denominator in discharge-enhancing stimuli is their arousing or alerting character. A single, phasic stimulus presentation, whether noxious or not, may elicit vigorous discharges in unanaesthetized rats, while stimulus repetition leads to response decrement [5,6]. In freely moving animals, firing of LC neurons correlates to orienting responses, adaptive transitions in behaviour and the initiation of novel response strategies [5,67,68]. Novelty and unexpectedness are stimulus features particularly powerful in triggering strong excitatory responses in LC neurons of cats and rhesus monkeys. In addition, stimuli that are challenging to the animal and elicit a behavioural response generally evoke strong firing responses. During consumption of food rewards, a decrease or no change of LC neuronal activity was observed [98]. In conditioned emotional response and conditioned food reward training, Rasmussen and Jacobs [164] found increased LC responses to previously neutral stimuli that were associated with a noxious air puff, whereas pairing to a reward did not increase LC activity. In a recent microdialysis study [40], NA release in the medial frontal cortex of rats increased in response to aversive as well as rewarding conditions, suggesting a general role for NA in emotional arousal. This study agrees with the results obtained with unit recordings, which is relevant to note because microdialysis techniques can directly assess neuromodulator release at a fairly macroscopic level and take into account release-regulating processes at the axon terminal level. In contrast, a NA-reinforcement hypothesis would have predicted a response of opposite sign to these conditions. Instead of supporting a function of NA in reinforcement signalling, these findings argue in favour of a saliency-signalling function of NA, conjoint with a role in attentive mechanisms and in reorganizing and initiating reactive behaviour.

7.5. Behavioural pharmacology

In interpreting findings from this field, caution is due because a disruption of learning by experimental manipulation of a neuromodulatory system does not allow to draw inferences about the underlying mechanistic functions of the neuromodulator. If lesioning of a neuromodulatory system is observed to impair acquisition of an arbitrary task, this can be due to impairment of a large variety of physiological functions. In spite of this limitation, some behavioural experiments have incorporated methods to control for pharmacological effects not related to learning per se, or support some neuromodulatory function in a specific phase of learning by varying the timing of drug administration. In the following paragraphs on the behavioural pharmacology of NA, DA and ACh, caution is also warranted when considering lesioning studies. With extended post-lesion survival times, adaptations may occur that obscure the normal functioning of the neuromodulatory system.

Two decades of intensive investigation have provided a somewhat varied picture of how NA functions in adaptive behaviour. A number of acquisition deficits, induced by LC lesions or 6-hydroxy-dopamine (6-OHDA) lesions of the dorsal noradrenergic bundle, have been reported in

tasks such as auditory and visual discrimination [59], appetitive conditional discrimination [168], fear conditioning [185] and spatial delayed conditional discrimination [110]. The studies by Everitt et al. [59] and Carli et al. [38] have argued against non-specific performance impairments, caused by sensory or motor deficits or a lack of motivation. Watson and McElligott [219] found impairments in a learning task where rats had to walk over a rod-runway after 6-OHDA lesioning of the coeruleocerebellar pathway. Performance of a locomotor task that was acquired prior to lesion of the coeruleocerebellar pathway remained intact in post-lesion tests. In contrast to this positive evidence for NA functions in learning, spatial learning in a Morris water maze was not affected by systemic administration of the β-receptor blocker propranolol [48]. There is both positive and negative evidence for the hypothesis that noradrenaline receptors contribute to retention of inhibitory avoidance [48,132]. One of the unresolved problems relating to this learning task is whether β-adrenergic receptor antagonists act against effects induced by peripherally released adrenaline or by noradrenaline released from LC axon terminals. Other studies have emphasized changes in extinction and/or reversal learning by lesions of NA systems [82,126,127] and, sometimes in parallel, an absence of acquisition deficits [80,127,178] (cf. Squire [191]). In general, the behavioural studies reviewed above justify an implication of central (nor)adrenaline receptors in some forms of learning and memory (especially retention of emotionally arousing events), but they offer few clues as to what type of signals are emitted by locus coeruleus neurons.

Some other studies have attempted to dissociate specific memory-related functions of NA from other functions in behaviour and cognition. Along the lines of investigation outlined above, attention has been paid in particular to functions of NA in: (1) facilitation of learning and memory consolidation, possibly by acting as a reinforcement signal [44,103]; (2) mediation of anxiogenic signals [165]; and (3) regulation of attention [126,128,129]. Although the available evidence does not allow a definite rejection of any of these hypotheses, findings by Mason and Fibiger [125], Carli et al. [38], Sara [178] and Selden et al. [185] argue against the reinforcement and anxiety hypotheses of NA function, while corroborating the attentional hypothesis (see, however, [161,162]). In a fear conditioning paradigm set in a distinctive environment, Selden et al. [185] exposed rats with cortical NA depletions to auditory clicks (conditioned stimulus) and footshocks (unconditioned stimulus) with variable time intervals. When compared to control rats, their findings indicated a shift in NA-depleted animals from fear conditioning to the CS (the explicit cue) towards an enhanced fear of contextual cues in the environment. The observation that NA depletion apparently has different consequences for explicit versus environmental cues argues against a role of NA as a general reinforcement signal and against a punishment- or anxiety-related signal, while being consistent with the attentional/saliency hypothesis of NA.

In contrast to adult rats, some evidence suggests a reinforcement-related function of NA in infant rats. In an olfactory conditioning task, Wilson and Sullivan [230] paired odor presentations to electrical stimulation of the medial forebrain bundle-lateral hypothalamus and found the β-receptor antagonist propranolol to block the acquisition of conditioned responses to the odor, a process that depends on the integrity of the olfactory bulb. Further experiments on the role of NA in mechanisms of reinforcement in infant rats were reported by Sullivan et al. [201], but whether or not NA specifically acts as a reinforcement signal in this process awaits further investigation.

In conclusion, in vivo electrophysiological evidence argues most decisively against a specific function of NA as a reinforcement signal, being supported by behavioural evidence. Our knowledge of catecholamine diffusion is too limited at present to assess how well NA would be capable of equally reaching modifiable synapses in e.g., neocortex. Although NA appears to have multiple effects on synaptic plasticity, none of these effects supports a specific function of NA as a reinforcement signal. Furthermore, well-characterized pre- and postsynaptic effects of NA compromise a role for NA as a simple multiplicative factor in learning rules. Based on this accumulative evidence, it seems warranted to meet NA-based hypotheses on reinforcement with the utmost scepticism. However, olfactory conditioning in infant rats may form an exception to this.

8. Acetylcholine

8.1. Equal access

Throughout the mammalian brain, acetylcholine (ACh) is supplied mainly by six anatomically segregated cholinergic cell groups, four of which are located in the basal forebrain (medial septum, group Ch1; vertical nucleus of the diagonal band of Broca, Ch2; horizontal nucleus of the diagonal band, Ch3; nucleus basalis of Meynert/substantia innominata (NBM/SI), Ch4 [79,133,188]. The remaining two groups are found in the pedunculopontine region (Ch5) and laterodorsal tegmental nucleus (Ch6) located in the mesopontine hindbrain, and these groups project to several nuclei of the thalamus, substantia nigra and other parts of the basal ganglia [73,79,118,195,196]. The cholinergic innervation of neocortex is mainly derived from Ch4 in NBM/SI [133], while the hippocampal formation and hypothalamus are supplied by medial septal group Ch1 and Ch2 in the vertical diagonal band of Broca. In the neocortex, choline acetyltransferase (ChAT)-positive fibers are more or less homogeneously distributed across many areas of the cortical mantle and reach all cortical layers, although the relative ChAT density shows some variation depending on the area [94,217]. At first sight, these find-

ings seem to favour an equal access of ACh to modifiable cortical synapses. However, most ChAT synapses have been reported to be symmetrical and are located on clearly specified postsynaptic targets (dendritic shafts and spines [94,217]). Thus, it is unlikely that ACh exerts its major effects by diffusing freely into the extracellular milieu, without any specific effect on the postsynaptic sites opposing cholinergic terminals. This possibility becomes even more remote when considering the potent ACh breakdown by extracellular acetylcholinesterase (AChE). Even in slices superfused with normal physiological medium, the activity of AChE can be pronounced, as illustrated by the finding that EPSPs in dorsal and ventral striatal tissue are attenuated by bath application of AChE inhibitors [54,159]. Layer V pyramidal cells and fibers in the visual cortex show dense AChE staining [12]. Although the currently available observations do not allow to draw definite conclusions, they do cast severe doubt on free ACh diffusion throughout the extracellular space.

8.2. Interference with pre- and postsynaptic activity

Numerous electrophysiological effects of acetylcholine (primarily muscarinic) receptor activation have been demonstrated in structures scattered across all major brain regions. The powerful muscarinic atttenuation of glutamate release has been reported for many telencephalic structures, e.g., hippocampus [213], amygdala [199], dorsal and ventral striatum [54,159] and prefrontal cortex [214]. The most pronounced postsynaptic effects include membrane depolarization [54,81,134,149] and reduction of the Ca^{2+} activated K^+ current that underlies spike frequency adaptation in cortical pyramidal cells [149,155]. Part of the depolarizing effects is explained by closure of the M-current [81,149]. In in vivo unit recordings throughout cortex, both the membrane depolarization and reduction of frequency adaptation may be expressed in an enhancement of firing frequency [68,188]. As for NA, a cholinergic function in enhancing signal-to-noise ratio has been proposed based on unit recordings in visual cortex [188] and dorsolateral prefrontal cortex [96].

Another line of research has focussed on the role of NBM/SI cholinergic neurons in transitions from sleep to wakefulness. Metherate et al. [134] found several effects of in vivo stimulation of the nucleus basalis region on electrical activity in the neocortex, including EEG activation, depolarization of cortical neurons and a transition in subthreshold membrane potential oscillations from 1–5 Hz to 20–40 Hz. In addition, the role of medial septal ACh in the generation of theta activity in hippocampus and adjacent structures should be recalled [24]. These and other findings argue for a major role of NBM/SI ACh in cortical arousal, and in mediating transitions between different dynamic neural network states. Finally, the mesopontine cholinergic cell groups are also strongly implicated in behavioural arousal (see below), an effect that may in part be accomplished by promoting burst discharges via activation of a K^+ conductance in thalamic nucleus reticularis neurons [131]. As for NA, it is clear that there are ample opportunities for ACh to interfere with both pre- and postsynaptic activity in a multitude of brain structures.

8.3. Influence on synaptic plasticity

Muscarinic modulation of LTP and/or LTD has been extensively investigated. Among the great variety of reported effects are an augmentation or facilitation of LTP by cholinergic agonists in the dentate gyrus and areas CA1 and CA3 [27,35,88,205]. Metherate et al. [135] revealed a permissive or facilitatory effect of iontophoretically applied ACh on enhancement of neural responses in cat somatosensory cortex that were evoked by cutaneous stimulation. In the mossy fiber-CA3 synapse, Williams and Johnston [228] observed an attenuation of LTP induction by muscarine. In contrast, Katsuki et al. [101] found an attenuation of LTP in this synapse by the muscarinic antagonist atropine, suggesting a facilitating action of endogenous ACh. Further indications for a complex, possibly bidirectional muscarinic control of plasticity in these synapses were obtained by Maeda et al. [121]. These varied effects point to a complex function of ACh in regulating synaptic plasticity. In particular the area-dependence of the effects suggests that ACh has a more varied function than to emit a uniform reinforcement signal across distributed brain sites.

8.4. In vivo unit recordings

The question as to whether the firing activity of cholinergic neurons correlates to delivery or anticipation of reinforcing stimuli has hitherto remained unanswered. Although recordings from some neurons in NBM/SI and the diagonal band of Broca sometimes do indeed reveal enhanced firing specifically in response to the sight and taste of food rewards [50,167,174,229], it is unknown whether these responses are generated by cholinergic or non-cholinergic neurons. In the absence of evidence as crucial as this, it is impossible to draw support from these findings for a reinforcement hypothesis of cholinergic function. The point in case is that many neurons located in the NBM/SI area are not cholinergic and may belong to the ventral striatopallidal system, which is known to be involved in brain stimulation reward and drug addiction (see below). As regards the mesopontine cholinergic cell groups, their firing behaviour appears to be linked to transitions between global brain states associated with waking and rapid-eye-movement sleep [195,196], generation of hippocampal theta rhythm [24] and locomotor activity [73].

8.5. Behavioural pharmacology

Substances interfering with cholinergic function affect learning in a great variety of behavioural tasks (for re-

views, see [56,63,79,132]. Among these learning tasks are active and inhibitory avoidance, appetitive motivational learning, several discrimination paradigms, spontaneous and rewarded alternation, spatial learning and delayed (non-)matching to sample. Investigators have made considerable efforts in trying to decide whether ACh-related behavioural impairments can be specifically attributed to deficits in learning and memory or to other functions, e.g., attention, behavioural planning, perception and sensorimotor gating. Indeed, there is a substantial body of literature supporting such additional functions of ACh [63,79]. A major difficulty lies in the fact that in most studies cholinergic drugs were applied by systemic administration. Considering the widespread distribution of muscarinic receptors throughout the peripheral and central nervous system, Fibiger et al. [63] have remarked that cholinergic agents are likely to interfere with an entire gamut of brain functions. Consequently, a search for a unitary function of ACh in behaviour may well prove to be in vain. Given this pessimistic, but justified outlook, we will briefly attempt to extract from the literature evidence for a reinforcement signalling function of ACh.

Inhibitory avoidance in rodents is impaired when cholinergic antagonists are systemically injected before or during training. However, a specific involvement of ACh in learning this task has not been demonstrated, as some of the effects could be attributed to hyperactivity [190]. Muscarinic antagonist effects on active avoidance are confounded by similar side effects. Moreover, this paradigm has produced mixed results in that antagonists attenuate, e.g., one-way, but not two-way avoidance of electric shock in a shuttle box [79,200]. In a spatial alternation task (in which rats were given a sugar water reward when they alternately pressed a right or left lever depending on a prior, memorized stimulus [85]), scopolamine was found to impair accuracy of discrimination. In a simpler left-right alternation task, in which reward application was the same as in the first task of Heise et al. [85], scopolamine had no effect, making it unlikely that reward sensitivity was affected in these tasks (see also [108] for similar results with scopolamine). In line with these findings, Pontecorvo and Evans [163] observed a delay-dependent decrease in choice accuracy of monkeys treated with scopolamine in a delayed matching to sample task. Performance appeared to be intact when no stimulus-response delay was included, suggesting a short-term memory impairment, whereas no evidence was obtained for a diminished impact or appreciation of reward. Similar results were recently obtained in rats by Broersen et al. [30] using locally applied scopolamine in a delayed matching to position task. Evidence for a specific role of ACh in retention and information storage was supplied by Flood et al. [65] using post-training intracerebroventricular injections of cholinergic antagonists in an active avoidance task in mice. They found a dose-dependent deficit in retention. In a related study on active avoidance, Flood et al. [66] found enhanced memory retention following oral administration of the cholinergic agonist arecoline in the post-training phase, whereas the learning phase was unaffected. The fact that arecoline improved retention when applied in a post-training phase argues for a role of ACh in memory consolidation. If, however, ACh acted as a reinforcement signal one would have expected a change during the learning phase as well.

Finally, studies using electrolytic or excitotoxic lesions of NBM/SI will be briefly considered. Two groups have recently argued that gross lesioning of this area by electrolytic methods or by injecting ibotenic acid or kainate are invalid methods for assessing the role of cholinergic basalis neurons in learning [56,63]. Both lesioning methods induce concomitant damage to mainly GABAergic neurons belonging to the ventral striatopallidal system which projects to the caudal parts of the basal ganglia, the mediodorsal thalamic nucleus and hindbrain. This system is strongly implied in incentive motivational learning and forms a major target for addictive psychostimulants [56,107,158,179,235]. Using milder excitotoxins such as quisqualic acid, Dunnett and colleagues found a relative sparing of non-cholinergic cells in NBM/SI and, correspondingly, many behavioural deficits reported for ibotenic acid lesions could not be reproduced for these toxins. The deficits that could be reproduced comprised impairments of visual attention, discriminative accuracy and retention in passive avoidance tasks and may thus be attributable to damage of cholinergic cells (see also [124,146,171,221]).

In summary, the broad spectrum of electrophysiological ACh effects throughout almost the entire brain is paralleled by a variety of behavioural functions. A few behavioural studies suggest specific functions of basal forebrain ACh in attention and memory retention. Electrophysiological studies indicate varied and region-specific muscarinic effects on LTP. None of the in vivo or in vitro electrophysiological studies, however, specifically supports a function of ACh as a reinforcement signal. Given the likely involvement of ACh in arousal, attention, precision of performance, perceptual acuity and memory consolidation, it would be surprising if an hitherto undiscovered reinforcing function of ACh proved to be central. In fact, many behavioural studies indicate an absence of ACh antagonist effects on reward processing in learning.

9. Dopamine

9.1. Equal access

The dopaminergic (DA) innervation of the telencephalon is mainly derived from the substantia nigra pars compacta (SNpc) and ventral tegmental area (VTA), located in the ventral mesencephalon. First, it is important to recall the enormous differences in innervation density and DA content between the caudate nucleus, putamen, nu-

cleus accumbens and olfactory tubercle versus other parts of the telencephalon. In rhesus monkey, for instance, the ratio of DA concentrations in caudate/amygdala/premotor cortex/hippocampus was estimated to be 411:9.4:2.7:1 [31]. Within the neocortex of both lower and higher mammals, a rostrocaudal gradient of DA-fiber density is present, with the prefrontal cortex at the high end and occipital and parietal cortex having only a very sparse innervation [20,21,31,68]. The overall DA input to the neocortex is less extensive in rodents than primates [20,21]. Another source of inhomogeneity in DA innervation are laminar differences in fiber density. In rat neocortex, [^{3}H]DA labeled varicosities are predominantly found in layers V and VI of the prefrontal, cingulate, parietal, motor and supplementary motor areas [21,52]. Thus, the overall DA innervation of telencephalic structures shows great variability in density, while there can also be large differences within a single structure.

Early ultrastructural studies in this field [15,207] led researchers to propose a diffuse, paracrine mode of DA action along with NA [53] (cf. [137]). More recent studies, however, have drawn attention to the widespread occurrence of DA-containing terminals that make conventional synaptic contacts with dendritic spines and shafts in the striatum [77,187,215] and neocortex [20,153,154]. Conjoint with the presence of powerful DA uptake mechanisms [92,97,212], these findings raise doubts about the free diffusibility of DA in extracellular space and suggest that this amine may well exert modulatory effects restricted to neuronal microdomains. Of further relevance is the common pharmacological distinction between D1 and D2-groups of DA receptors. Since the D2-group has a much higher affinity for DA than the D1-group, equal access might be less of a problem if reinforcement would be mediated by D2 receptors. As pointed out below, however, D2 receptors do not seem to exclusively mediate dopaminergic effects on learning.

9.2. *Interference with pre- and postsynaptic activity*

Various DA actions on dorsal striatal neurons in vitro have been reported, including a D1 receptor mediated depression of EPSPs and of a slow inward conductance [36], a D2 receptor mediated opening of potassium channels [69] and a mixed D1–D2 receptor mediated reduction of sodium current [202]. In the ventral striatum in vivo and in vitro, the most reproducible effect demonstrated thus far is a presynaptic attenuation of EPSPs elicited by stimulation of amygdalar or limbic cortical afferents [156,233,234]. Primarily inhibitory effects have been reported for various neocortical areas, in line with the findings in ventral striatum. Ferron et al. [62] observed an attenuation of the excitatory response of neocortical neurons to thalamic stimulation when concurrently stimulating the ventromedial mesencephalon. With iontophoretic DA application in prefrontal cortex, Bunney and Aghajanian [34] found a decrease of spontaneous activity in layer V–VI cells, and Sesack and Bunney [186] and Pirot et al. [160] attributed this effect to D2-receptors. While for some structures it is not yet clear which DA receptor subtypes are exactly involved, the conclusion is warranted that DA is capable of interfering considerably with the excitability of both pre- and postsynaptic elements in structures known to participate in learning processes.

9.3. *Influence on synaptic plasticity*

No consistent picture has emerged as to how DA modulates synaptic plasticity in various brain structures. In area CA1 of the hippocampus in vitro, Frey et al. [70] reported no significant effect of the DA antagonist domperidone on LTP in the first 4 hours after tetanization, but a suppression of LTP maintenance at longer intervals. With the D1 antagonist Sch 23390, Frey et al. [71] found a similar effect, which reached statistical significance already after 30 min. following tetanization but remained modest. In the dorsal corticostriatal pathway, Calabresi et al. [37] observed a blockade of LTD by administration of D1 and D2 antagonists, indicating that coactivation of D1 and D2 receptors may be necessary for LTD induction in this projection. Finally, absence of dopaminergic modulation on LTP was reported for the prefrontal-ventral striatal pathway in vitro, as indicated by a lack of effect of both DA and of a mixture of D1 and D2 antagonists in intra- and extracellular recordings [157]. In the subicular projection to the ventral striatum in vivo, LTP induction remained intact after 6-OHDA lesions of this structure (Mulder and Lopes da Silva, unpublished data). If these variable results permit any conclusion to be drawn at all at this point, it is that DA is unlikely to exert a generalized, robust effect on the plasticity of multiple glutamatergic systems in the telencephalon under a broad range of conditions. Furthermore, even the most dramatic effect claimed until now [37] is not the type of modulation expected if a reinforcement algorithm were implemented in corticostriatal systems. This effect suggests a permissive action of the D1 and D2 type of receptors on LTD induction, not a bidirectional modulation. Cepeda et al. [41] recently reported a D1 receptor mediated reversible enhancement of NMDA responses in the dorsal striatum, which is likewise suggestive of a permissive action of DA on NMDA receptor-mediated effects.

9.4. *Unit recordings in awake animals*

Investigators have been able to gain insight into the functional correlates of the firing activity of DA neurons due to the fortunate circumstance that these cells have typical spike characteristics [136,182,194,198]. The firing behaviour of SNpc and VTA dopaminergic neurons will be discussed separately. As concerns the SNpc, research has

been focussed on how sensory stimuli affect responsiveness of DA neurons, since their discharge appears to be uncorrelated to limb or whole body movements of the animal [51,194]. Chiodo et al. [42] found stimuli capable of behaviourally activating the animal to be most effective in altering DA neuron discharge rates in SNpc; most of the stimuli used in this study were neither noxious nor rewarding. Strecker and Jacobs [198] exposed awake cats to a variety of experimental conditions, some of which were arousing and/or stressful, and could not detect significant effects on the tonic discharge rate of DA neurons in SNpc. Remarkably, short-lasting auditory or visual stimuli did elicit phasic responses of DA neurons in quiet awake cats, but not in cats pre-exposed to arousing or stressful conditions. These results suggest that long-lasting stimuli of behavioural significance may, by themselves, not be sufficient to alter discharge rate, but may modulate neuronal responsiveness to other stimuli, which can be brief and neutral. If DA neurons in SNpc play a general role in reward signalling, food presentation and feeding would be expected to elicit changes in firing activity. However, Strecker et al. [197] failed to observe any change during feeding in freely moving cats, as well as after glucose injections. In the analysis of Steinfels et al. [194], stimuli that attracted the attention of the animal and evoked orienting responses were most effective in altering (i.e., decreasing) the discharge of DA neurons in SNpc. In the studies on SNpc, VTA and dopaminergic cell group A8 by Schultz et al. [182,183], monkeys were trained to reach for a food reward in response to a trigger stimulus (i.e., door opening of a food box conjoint with a sound). The most vigorous responses were elicited by the trigger stimulus, while a minority of responses were associated with forelimb movement or food reward. Altogether, these observations on SNpc DA neurons indicate functions in attention, orientation and the formation of behavioural responses to significant and salient stimuli, but not in reinforcement signalling per se.

As regards VTA, recent work by Schultz et al. [117,180] in monkeys has shed new light on the motivational conditions that are capable of evoking responses in mesocorticolimbic neurons. In a spatial delayed response task, significantly more neurons in VTA as compared to SNpc and area A8 responded to reward delivery. Importantly, phasic reward-evoked responses were most often seen during learning (25% out of 75 neurons) and much less often once acquisition had been completed (9% out of 163 neurons [180]). Similar observations were made by Ljungberg et al. [117] for VTA, SNpc and A8 in an operant conditioning task, where the animal was instructed to perform an arm reaching movement into a food box following onset of a cue light. In the acquisition phase, DA responses were most effectively elicited by novel stimuli, primary reward and conditioned incentive stimuli. Again, DA neurons were more responsive to reward than to the conditioned light during learning, whereas the reverse was true directly after learning was completed. In the course of overtraining, responses to reward and light tended to weaken altogether.

These recordings suggest that discharges of DA neurons in VTA may indeed be related to reward, but in a way that depends on its predictability (cf. Mirenowicz and Schultz [139]). The finding that most of the reward responses disappear after learning indicates that VTA or SNpc do not signal the availability of reward per se, since this condition is constant across conditioning, postconditioning and overtraining trials. In this regard, DA neurons fail to show a direct correlation to reward delivery as much as seems to be the case for NA and ACh neurons. However, as Rescorla and Wagner [166] have pointed out, conditioning is very much a process of learning to predict upcoming significant events. The findings of Schultz et al. suggest that DA neurons may take part in this process of learning to anticipate or to expect because responses to reward tend to fade whenever its occurrence has become predictable, while responses to predictive stimuli gradually take over [139]. The suggestion has been raised [142] that the mesencephalic DA system, and the systems feeding afferent fibers into it, may therefore implement a form of TD learning (see Eqns. 6 and 8). We will next examine this intriguing idea in more detail.

The first multiplicative factor on the right-hand side of Eqn. 8 represents the error in reward prediction. If DA neurons are hypothesized to release this factor, their firing rate should go up if a reward exceeds an animal's prediction (or expectation, which is equivalent in this context) and should remain constant if the animal correctly anticipated its occurrence. Thus far, the hypothesis is compatible with the findings of Schultz and coworkers. However, a reinforcement system should also be able to cope with aversive conditions and cases where expected rewards are omitted. The TD algorithm prescribes that firing of DA neurons should go down if the rewarding value of the unconditioned stimulus is lower than expected (Eqns. 6 and 8). Current studies by the group of Schultz (pers. comm.), however, suggest that DA neurons are relatively insensitive to aversive stimuli. In an active avoidance task, they failed to respond clearly to primary aversive stimuli like a mild air puff or hypertonic saline. Likewise, they did not respond to conditioned aversive stimuli, except if the stimulus features were similar to a conditioned appetitive stimulus applied alternately during the same experimental session (Mirenowicz and Schultz [138]). On account of these observations, it would appear that the DA system preferentially responds to stimuli with positive, but not negative reward value. In contrast, a long series of neurochemical studies showed an enhancement of DA release in prefrontal cortex and dorsal and ventral striatum in response to stressors such as electric tail or foot shock, handling and noxious tail pinch [1,40,95,123,176,209]. Microdialysis studies have shown that DA release can be enhanced by both stressful conditions and by events of positive evolutionary significance, such as mating be-

haviour [23,46], intake of palatable liquids [23] and food rewards [40]. While these electrophysiological and neurochemical analyses support a role of DA neurons in signalling salient (and especially unpredicted) environmental cues and primary reinforcers, the most important inference to be drawn here is that DA neuron discharges and release do not simply encode how 'good' or how 'bad' the unconditioned stimulus is. Microdialysis and unit recording studies agree that the activation of the DA system does not clearly decrease in response to negative reinforcers. Thus, at least one other neurochemical system would be needed to encode negative reward values. Another problem faced by the DA-reinforcement hypothesis is that DA neurons tend to respond to novel stimuli [116,117]. By itself, novelty need not be rewarding or predictive of upcoming rewarding events. Therefore a dopamine neuron's response may signal unexpected events in a more general sense, not only unpredicted reward-related stimuli. Of course, these considerations should not be taken as objections against the possibility that dopamine may have a more general modulatory function in learning and plasticity.

9.5. Behavioural pharmacology

An extensive body of evidence supports some role for DA in learning processes that depend on the dorsal and ventral striatum as well as prefrontal cortex. The mesocorticolimbic DA system has been implicated in various learning tasks including conditioned place preference, conditioned avoidance, conditioned reinforcement and maze learning (reviewed by [23,64,158,169,179,225]. Furthermore, electrical self-stimulation of the medial forebrain bundle also appears to depend significantly on DA fibers [225,231]. The psychostimulant drugs amphetamine and cocaine exert their addictive effects primarily through dopaminergic synapses [107,225,231]. DA in neocortical-dorsal striatal pathways presumably contributes more strongly to learning of motor skills and habits than to motivational learning [64,169,179]. We will therefore focus on the mesolimbic and mesocortical DA systems.

Beninger and colleagues examined effects of the DA antagonist pimozide on incentive motivational learning, in which initially neutral stimuli come to elicit conditioned responses by being consistently presented in association with reinforcement. In cocaine- or amphetamine-induced environmental conditioning, pimozide was observed to block conditioning but left behavioural expression of conditioning intact after acquisition had been completed [18,19,39]. Wise et al. [232] observed a pimozide-induced decrease of lever-pressing and running for food reward in rats, and hypothesized that DA mediates the rewarding effect of hedonic stimuli on adaptive behaviour. Although there is no full agreement as to which DA receptor subtypes mediate these effects, some studies support a role for both dopamine D1 and D2 receptors in place conditioning [224] and operant conditioning for, e.g., water, saccharin and electrical VTA stimulation [17,109,137,148]. In contrast to incentive motivational tasks, learning to associate two sensory stimuli, presented in close succession, is apparently less sensitive to DA antagonists [16,57,64,172].

Two objections can be raised against the claim that the above mentioned findings would support a role of DA as a reinforcement signal. Firstly, pimozide and some other neuroleptics are potent calmodulin inhibitors [220] and inhibit LTP in hippocampal slices [140]. Inasmuch as LTP and other forms of calmodulin-dependent plasticity may be involved in incentive motivational learning, it cannot be decided whether pimozide exerts its behavioural effects through DA receptor or calmodulin block. Secondly, impairments of incentive motivational learning by DA antagonists do not permit to specify whether DA unequivocally signals reward (or errors in reward prediction) or mediates different functions leading to similar behavioural effects. While motor-related side effects of DA antagonists could in some studies be isolated from learning-related effects, it has proven difficult to dissociate a putative reinforcement-signalling function of DA from other components of incentive motivational learning [23,57,169]. In particular, the mesocorticolimbic DA system has been hypothesized to attribute incentive salience to stimuli preceding emotionally significant events [60,61,172]. 'Incentive salience' refers to the attractiveness of stimuli and their attention-grabbing and approach-eliciting capacities, and should be held apart from their capacity to trigger subjective emotional states (e.g., euphoria) or behavioural adaptation (cf. [172]).

Some further indications against a general reward-signalling or reward-predicting function of DA can be extracted from the behavioural literature. If DA release would signal reward, DA receptor antagonists or amphetamine would be expected to influence the response strategy of animals in, e.g., operant conditioning. Evenden and Robbins [58] trained rats in operant chambers containing two levers, one of which led to food pellet delivery with 0.33 probability when pressed down. It was found that amphetamine, chlordiazepoxide and α-flupentixol did not alter the win-stay or lose-stay percentages of responding except for very high doses. The lack of effect on win-stay behaviour suggests that associations between previous motor responses and rewards were correctly made and remembered across subsequent trials and thus argues against a gross disruption of reward signalling by these drugs. In a different experiment [170], rats were classically conditioned with water as the unconditioned stimulus and light plus noise as a conditioned stimulus. Secondly, rats were trained to choose between two levers, one of which led to delivery of the conditioned reinforcer (i.e., light plus noise). It was found that amphetamine enhanced responding for the light, whereas 6-OHDA lesions of the ventral striatum blocked this effect. However, the lesions did not affect the choice between two levers. These results are difficult to

reconcile with reward-signalling or reward-predicting theories of DA function, because lever choice is very likely to depend on an animal's capacity to appreciate reward or stimuli associated with or predictive of reward. Finally, it is worthwhile to draw attention to some studies indicating an overall lack of specific DA influence on other forms of learning than those mentioned, e.g. place navigation learning (6-OHDA lesions of rat nucleus accumbens, dorsal striatum and medial frontal cortex) [80] and delayed matching to position (local infusion of *cis*-flupentixol in rat medial frontal cortex) [30].

Thus, much of the earlier evidence regarding acquisition of incentive motivational tasks relied on neuroleptics with demonstrated effects on calmodulin. However, the observation that both amphetamine and 6-OHDA lesions of ventral striatum can affect motivational learning [60,64,170,179,206,225] argues in favour of some function of DA in this process. We may tentatively conclude that DA does have an important function in reward-dependent learning in addition to its regulatory action on motor activity. Taking the findings of Robbins, Everitt and colleagues into account, it seems appropriate to define DA function in terms of how salient stimuli become linked to appetitive and instrumental behaviour. In contrast to a reinforcement-signalling theory of DA, there is reasonable experimental evidence for the hypothesis that dopamine functions in this coupling process by potentiating the incentive properties of conditioned reinforcers.

In summary, in vivo electrophysiological findings can be brought in line with behavioural pharmacology inasmuch as these two disciplines currently acknowledge salient and behaviourally significant (especially: reinforcement-predictive) stimuli as important triggers for activating midbrain DA systems. Unit recordings and neurochemical studies in awake animals agree in that DA neurons can be activated by positive reinforcers, but discrepancies remain as to their responsiveness to negative reinforcers. Further investigations will be necessary to examine whether these differences may be attributed to regulation of DA release at the axon terminal. Meanwhile, both types of study are incompatible with a general purpose function of dopamine in reinforcement signalling, since negative reward does not appear to be encoded by a decrease in dicharge rate of DA neurons or in DA release. Furthermore, a reinforcement signalling function of DA fails to draw support from anatomical and in vitro electrophysiological evidence. Ultrastructural studies suggest a less diffuse, more site-specific action of dopaminergic fibers in striatum and neocortex than previously purported. The present picture of how DA modulates synaptic plasticity is too diverse to allow firm conclusions. The most reproducible cellular DA effect reported so far is a reversible attenuation of excitatory transmission in the ventral striatum and frontal cortex — an effect that would hinder implementation of a reinforcement algorithm by this system. Instead of functioning in reinforcement-signalling, dopaminergic neurons are likely to transmit information about novel or unexpected events of behavioural significance and may modulate behavioural reactivity to incentive stimuli.

10. Alternatives

Having noted that the NA, ACh and DA systems fall short of fulfilling some critical predictions based on reinforcement algorithms, it is important to consider possible alternatives. One of the possibilities not yet explored is that a combination of neuromodulators, released simultaneously, could signal reinforcement and accomplish appropriate synaptic weight changes. In rat visual cortex, shifts in ocular dominance were reduced after 6-OHDA lesions of the dorsal noradrenergic bundle and *N*-methylaspartate lesions of the cholinergic innervation from the basal forebrain, but not after lesioning either system alone [14]. Findings by Bröcher et al. [29] suggested a synergistic interaction between NA and ACh in LTP-induction in rat visual cortex in vitro. Taken together, these findings open a new avenue for exploring the combined effects of neuromodulators in synaptic plasticity, but as yet they cannot offer a sufficient basis for a reinforcement hypothesis based on combined modulatory actions. The problem is that in adult animals the combined information transmitted by the NA and ACh systems lacks an evaluative component as much as either of the two systems acting in isolation.

Besides the neurochemical systems already discussed, several brain areas are known to contain at least a fraction of reward-sensitive neurons. An important feature of some of these neurons is their response preference for either negatively or positively reinforcing stimuli. As indicated by in vivo electrophysiological studies, subregions of the amygdala and frontal and cingulate cortex contain neurons that may specifically fire in response to reward or punishment, although there also appear to be neurons with less specific firing behaviour. Nishijo et al. [151] described visually and auditory responsive units in the monkey amygdala that discriminated between positive or negative familiar stimuli or palatable versus non-palatable foods. All units responding to oral-sensory and visual stimuli altered their firing when the affective significance of food was changed. Wilson and Rolls (see [173]) found amygdala neurons that were more responsive to reward-related than to punishment-related stimuli. In the monkey orbitofrontal cortex, similar selective responses to reinforcing or aversive objects were reported by Thorpe et al. [211,173] (see also [175]). Like the amygdala units of Nishijo and colleagues, these units proved to be plastic in their firing behaviour since their responses were rapidly and flexibly modified when the reinforcement value of the stimulus was altered. Of special relevance are the findings of Niki and Watanabe [150] in monkey prefrontal and

cingulate cortex. In a differential reinforcement task, they not only found unit firing in relation to reward delivery, but also units showing a selective increase in firing following behavioural errors of the monkey. In addition to these cortical and amygdalar regions, Rolls and coworkers [174,229] described neurons in the lateral hypothalamus, diagonal band of Broca and NBM/SI-region which selectively responded to either food objects or aversive visual stimuli. Finally, Schultz et al. [181,3] identified reward-related responses in the ventral and (to a lesser extent) dorsal striatum of macaque monkeys.

Altogether, these studies reveal that subpopulations of neurons in primarily limbic areas can be selectively responsive to aversive versus rewarding stimuli (or vice versa). Moreover, the finding that some of these neurons are adaptive in their firing behaviour when reward contingencies are changing suggests that these neurons may actively participate in learning by reinforcement. Although the adaptive and discriminative properties of these neurons are useful in building a neural system for learning sensorimotor tasks, it should be added that the transmitter utilized by these neurons is unlikely to meet the requirements posed by reinforcement learning rules (e.g., Eqns. 4 and 8). Projection neurons of the amygdala, frontal and cingulate cortex are most probably glutamatergic. Those of the striatum, hypothalamus and NBM/SI (partially including ventral pallidum) are predominantly GABAergic, although neurons in NBM/SI may also have been cholinergic. If one adopts the hypothesis that evaluative information is relayed by some of these areas instead of NA, DA or ACh neurons, the assumption of a diffusible and multiplicative reinforcement signal has to be abandoned. Recently, a new algorithm has been constructed to overcome this problem (Pennartz, in preparation). In this model, reward-related information is relayed by linearly interacting, glutamatergic synapses with adaptive thresholds for modification. On account of computational efficiency this algorithm is comparable to previously published reinforcement models.

11. Conclusions and future prospects

In this commentary the biological plausibility of computational models for reinforcement learning was evaluated considering the properties of widespread neuromodulatory systems in the vertebrate brain. If one demands a complete fulfilment of all computational predictions based on either ARP or TD learning, supportive evidence is lacking for the NA system on account of anatomical, electrophysiological and behavioural findings. We have seen that a one-to-one relationship between reward and firing of locus coeruleus neurons is lacking, as these cells can respond to a broad range of arousing and alerting stimuli, especially when they elicit behavioural reactions. The behavioural evidence is in line with attentional and saliency-signalling functions of NA, as well as a function in the retention of emotionally arousing events. Apart from a general lack of behavioural and electrophysiological evidence for a NA-reinforcement hypothesis, the pronounced postsynaptic effects of NA compromise a biological implementation of existing reinforcement algorithms by this system. However, it remains possible that NA exerts a reinforcement-related function under specific conditions, as may be the case in conditioning in the olfactory bulb of infant rats [201].

As concerns ACh, the behavioural evidence fails to support a specific reinforcement-signalling function, but instead supports a role in attention, short-term memory and memory consolidation. Furthermore, the multitude of pre- and postsynaptic effects of muscarinic agonists throughout the CNS is at odds with the reinforcement algorithms. An important avenue for future research concerns the point that cholinergic neurons in NBM/SI have not yet been recorded in behaving animals with certainty. It would be valuable to examine whether the reward-related units reported by DeLong [50,167] and Rolls et al. [174,229] in this area are indeed cholinergic or belong to the ventral striatopallidal system. A significant step towards this would be to combine recordings in awake animals with intracortical stimulation and spike collision tests in order to identify cortically projecting neurons. In behavioural pharmacology, a site-directed approach using intracranial drug administration may help to avoid an accumulation of multiple drug effects and thus to define the memory-related functions of neuromodulatory subsystems more precisely.

Of all three candidate reinforcement systems, DA remains the most controversial one. Whereas some previous behavioural experiments have been taken to indicate a reinforcing function for this catecholamine in incentive motivational learning, this view was criticized here in terms of interpretational difficulties and pharmacological side effects of the drugs used. At first glance, electrophysiological findings in awake monkeys seem to make a case for the plausibility of temporal difference learning, but upon closer examination the discrepancies between neurochemical studies (suggesting enhanced DA release in response to negative reinforcement) and unit recordings (suggesting no or only weak responses) pose a problem that calls for further investigation. The DA hypothesis of reinforcement fails to draw convincing support from electrophysiological findings in vitro. Nevertheless, unexpected support may emerge when in vitro and in vivo tests will be conducted in brain areas other than those examined so far and under a broader range of experimental conditions.

Despite these cautionary remarks, the conclusion is inevitable that none of the three modulators appears to function in reinforcement signalling *in general*, i.e., in many brain areas of young as well as adult animals and in a variety of sensorimotor learning tasks. It should be emphasized that even the least unplausible candidate for reinforcement signalling — dopamine — has been clearly

implied in only one type of learning (i.e., incentive motivational learning) mediated by one subsystem (i.e., the mesocorticolimbic DA projections from VTA). This conclusion should, however, not be understood as implying that NA, Ach or DA would not serve any function in behavioural reinforcement at all. As much as attentional, anticipatory, consolidating and short-term memory processes form an inextricable part of learning in general, these modulators should be considered integral parts of the brain machinery mediating learning by reinforcement. At present, the central tenet of reinforcement algorithms, viz. the existence of a diffusely acting reinforcement system, fails to draw support from the known properties of the systems discussed. Especially for dopamine, however, it cannot be ruled out that support may still emerge pending further research. Rather than searching for a unique modulator signalling reinforcement and orchestrating synaptic weight changes throughout the brain, it seems more appropriate to ask: how do neuromodulators *cooperate* with signals from other brain areas (most notably the limbic cortex and amygdala) in accomplishing learning by reinforcement?

Thus far, reinforcement hypotheses have been evaluated specifically in relation to the learning rules of existing algorithms. Perhaps, one might reason, it would be fair to be less stringent and maintain a neuromodulatory reinforcement hypothesis in a general sense. As long as this approach relies on intuitive arguments, it cannot guarantee computational efficiency in a variety of sensorimotor learning tasks. Moreover, a more loosely specified attitude would be hard to maintain in the face of the overall lack of experimental evidence for the hypothesis. It is on these grounds that a search for alternative models should be encouraged. Broadly speaking, this search may follow two pathways. One pathway is to maintain the overall concept of multiplicative reinforcement signals and to test which of the assumptions embedded in the existing algorithms are critical and which can be omitted. The second pathway, already alluded to above, is to search for other neurobiological substrates of reinforcement signals than diffusible modulators (e.g., information carried by glutamatergic or GABAergic fibers) and adopt those as building blocks for computational models. Certainly, if this search is accompanied by experiments designed to test forthcoming computational predictions, modelling approaches to the problem of learning by reinforcement will yield exciting new insights in the future.

Acknowledgements

I would like to thank J.J. Hopfield for his stimulating advice during the course of this research project and F.H. Lopes da Silva, J. van Pelt, A. van Ooijen for their critical reading of the manuscript. This project was supported by a Talent Fellowship of the Netherlands Organization for Scientific Research.

References

[1] Abercrombie, E.D., Keefe, K.A., DiFrischia, D.S. and Zigmond, M.J., Differential effect of stress on in vivo dopamine release in striatum, nucleus accumbens, and medial frontal cortex, *J. Neurochem.*, 52 (1989) 1655–1658.

[2] Angevine, J.B. and Cotman, C.W., *Principles of Neuroanatomy*, Oxford University Press, New York, 1981.

[3] Apicella, P., Ljungberg, T., Scarnati, E. and Schultz, W., Responses to reward in monkey dorsal and ventral striatum, *Exp. Brain Res.*, 85 (1991) 491–500.

[4] Artola, A. and Singer, W., Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation, *Trends Neurosci.*, 16 (1993) 480–487.

[5] Aston-Jones, G., The locus coeruleus. Behavioral functions of locus coeruleus derived from cellular attributes, *Physiol. Psychol.*, 13 (1985) 118–126.

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

[7] Barto, A.G., Adaptive critics and the basal ganglia. In J.C. Houk, J.L. Davis and D.G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge, MA, 1995, pp. 215–232.

[8] Barto, A.G. and Anandan, P., Pattern-recognizing stochastic learning automata, *IEEE Transact. Syst. Man Cybern.*, 15 (1985) 360–375.

[9] Barto, A.G. and Sutton, R.S., Landmark learning: an illustration of associative search, *Biol. Cybern.*, 42 (1981) 1–8.

[10] Barto, A.G., Sutton, R.S. and Anderson, C.W., Neuronlike adaptive elements that can solve difficult learning control problems, *IEEE Transact. Syst. Man Cybern.*, 13 (1983) 834–846.

[11] Barto, A.G. and Jordan, M.I., Gradient following without backpropagation in layered networks, *IEEE First Int. Conf. Neural Networks, Vol. II*, 1987, pp. 629–636.

[12] Bear, M.F., Carnes, K.M. and Ebner, F.F., An investigation of cholinergic circuitry in cat striate cortex using acetylcholinesterase histochemistry, *J. Comp. Neurol.*, 234 (1985) 411–430.

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

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

[15] Beaudet, A. and Descarries, L., The monoamine innervation of rat cerebral cortex: synaptic and nonsynaptic axon terminals, *Neuroscience*, 3 (1978) 851–860.

[16] Beninger, R.J., The role of dopamine in locomotor activity and learning, *Brain Res. Rev.*, 6 (1983) 173–196.

[17] Beninger, R.J., Receptor subtype-specific dopamine agonists and antagonists and conditioned behaviour. In P. Willner and J. Scheel-Krüger (Eds.), *The Mesolimbic Dopamine System: from Motivation to Action*, Wiley, Chichester UK, 1991, pp. 273–299.

[18] Beninger, R.J. and Hahn, B.L., Pimozide blocks establishment but not expression of amphetamine-produced environment-specific conditioning, *Science*, 220 (1983) 1304–1306.

[19] Beninger, R.J. and Herz, R.S., Pimozide blocks establishment but not expression of cocaine-produced environment-specific conditioning, *Life Sci.*, 38 (1986) 1425–1431.

[20] Berger, B., Trottier, S., Verney, C., Gaspar, P. and Alvarez, C., Regional and laminar distribution of the dopamine and serotonin innervation in the macaque cerebral cortex: a radioautographic study, *J. Comp. Neurol.*, 273 (1988) 99–119.

[21] Berger, B., Gaspar, P. and Verney, C., Dopaminergic innervation of the cerebral cortex: unexpected differences between rodents and primates, *Trends Neurosci.*, 14 (1991) 21–27.

[22] Bienenstock, E.L., Cooper, L.N. and Munro, P.W., Theory for the

development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, *J. Neurosci.*, 2 (1982) 32–48.
[23] Blackburn, J.R., Pfaus, J.G. and Phillips, A.G., Dopamine functions in appetitive and defensive behaviours, *Progr. Neurobiol.*, 39 (1992) 247–279.
[24] Bland, B.H. and Colom, L.V., Extrinsic and intrinsic properties underlying oscillation and synchrony in limbic cortex, *Progr. Neurobiol.*, 41 (1993) 157–208.
[25] Bliss, T.V.P., Goddard, G.V. and Riives, M., Reduction of long-term potentiation in the dentate gyrus of the rat following selective depletion of monoamines, *J. Physiol.*, 334 (1983) 475–491.
[26] Bliss, T.V.P. and Collingridge, G.L., A synaptic model of memory: long-term potentiation in the hippocampus, *Nature*, 361 (1993) 31–39.
[27] Blitzer, R.D., Gil, O. and Landau, E.M., Cholinergic stimulation enhances long-term potentiation in the CA1 region of rat hippocampus, *Neurosci. Lett.*, 119 (1990) 207–210.
[28] Boeijinga, P.H., Mulder, A.B., Pennartz, C.M.A., Manshanden, I. and Lopes da Silva, F.H., Responses of the nucleus accumbens following fornix/fimbria stimulation in the rat. Identification and long-term potentiation of mono- and polysynaptic pathways, *Neuroscience*, 53 (1993) 1049–1058.
[29] Bröcher, S., Artola, A. and Singer, W., Agonists of cholinergic and noradrenergic receptors facilitate synergistically the induction of long-term potentiation in slices of rat visual cortex, *Brain Res.*, 573 (1992) 27–36.
[30] Broersen, L.M., Heinsbroek, R.P.W., de Bruin, J.P.C., Joosten, R.N.J.M.A., van Hest, A. and Olivier, B., Effects of local application of dopaminergic drugs into the dorsal part of the medial prefrontal cortex of rats in a delayed matching to position task: comparison to local cholinergic blockade, *Brain Res.*, 645 (1994) 113–122.
[31] Brown, R.M., Crane, A.M. and Goldman, P.S., Regional distribution of monoamines in the cerebral cortex and subcortical structures of the rhesus monkey: concentrations and in vivo synthesis rates, *Brain Res.*, 168 (1979) 133–150.
[32] Brown, T.H., Kairiss, E.W. and Keenan, C.L., Hebbian synapses: biophysical mechanisms and algorithms, *Ann. Rev. Neurosci.*, 13 (1990) 475–511.
[33] Bryson, A.E. and Ho, Y.-C., *Applied Optimal Control*, Blaisdell, New York, 1969.
[34] Bunney, B.S. and Aghajanian, G.K., Dopamine and norepinephrine innervated cells in the rat prefrontal cortex: pharmacological differentiation using microiontophoretic techniques, *Life Sci.*, 19 (1976) 1783–1792.
[35] Burgard, E.C. and Sarvey, J.M., Muscarinic receptor activation facilitates the induction of long-term potentiation (LTP) in the rat dentate gyrus, *Neurosci. Lett.*, 116 (1990) 34–39.
[36] Calabresi, P., Mercuri, N., Stanzione, P., Stefani, A. and Bernardi, G., Intracellular studies on the dopamine-induced firing inhibition of neostriatal neurons in vitro: evidence for D1 receptor involvement, *Neuroscience*, 20 (1987) 757–771.
[37] Calabresi, P., Maj, R., Mercuri, N. and Bernardi, G., Coactivation of D1 and D2 dopamine receptors is required for long-term synaptic depression in the striatum, *Neurosci. Lett.*, 142 (1992) 95–99.
[38] Carli, M., Robbins, T.W., Evenden, J.L. and Everitt, B.J., Effects of lesions to ascending noradrenergic neurones on performance of a 5-choice serial reaction task in rats; implications for theories of dorsal noradrenergic bundle function based on selective attention and arousal, *Behav. Brain Res.*, 9 (1983) 361–380.
[39] Carr, G.D. and White, N.M., Conditioned place preference from intra-accumbens but not intra-caudate amphetamine injections, *Life Sci.*, 33 (1983) 2551–2557.
[40] Cenci, M.A., Kalén, P., Mandel, R.J. and Björklund, A., Regional differences in the regulation of dopamine and noradrenaline release in medial frontal cortex, nucleus accumbens and caudate-putamen: a microdialysis study in the rat, *Brain Res.*, 581 (1992) 217–228.
[41] Cepeda, C., Buchwald, N.A. and Levine, M.S., Neuromodulatory actions of dopamine in the neostriatum are dependent upon the excitatory amino acid receptor subtypes activated, *Proc. Natl. Acad. Sci. USA*, 90 (1993) 9576–9580.
[42] Chiodo, L.A., Antelman, S.M., Caggiula, A.R. and Lineberry, C.G., Sensory stimuli alter the discharge rate of dopamine (DA) neurons: evidence for two functional types of DA cells in the substantia nigra, *Brain Res.*, 189 (1980) 544–549.
[43] Crick, F., The recent excitement about neural networks, *Nature*, 337 (1989) 129–132.
[44] Crow, T.J., Cortical synapses and reinforcement: a hypothesis, *Nature*, 219 (1968) 736–737.
[45] Dahl, D. and Sarvey, J.M., Norepinephrine induces pathway-specific long-lasting potentiation and depression in the hippocampal dentate gyrus, *Proc. Natl. Acad. Sci. USA*, 86 (1989) 4776–4780.
[46] Damsma, G., Pfaus, J.G., Wenkstern, D., Phillips, A.G. and Fibiger, H.C., Sexual behavior increases dopamine transmission in the nucleus accumbens and striatum of male rats: comparison with novelty and locomotion, *Behav. Neurosci.*, 106 (1992) 181–191.
[47] Dayan, P., Montague, P.R. and Sejnowski, T.J., Foraging in an uncertain environment using predictive Hebbian learning. In J.D. Cowan, G. Tesauro and J. Alspector (Eds.), *Advances in Neural Information Processing Systems, Vol. 6*, Morgan Kaufmann, San Mateo, CA, 1994.
[48] Decker, M.W., Gill, T.M. and McGaugh, J.L., Concurrent muscarinic and beta-adrenergic blockade in rats impairs place-learning in a water maze and retention of inhibitory avoidance, *Brain Res.*, 513 (1990) 81–85.
[49] Dehaene, S. and Changeux, J.P., The Wisconsin card sorting test: theoretical analysis and modeling in a neuronal network, *Cerebral Cortex*, 1 (1991) 62–79.
[50] DeLong, M.R., Activity of pallidal neurons during movement, *J. Neurophysiol.*, 34 (1971) 414–427.
[51] DeLong, M.R., Crutcher, M.D. and Georgopoulos, A.P., Relations between movement and single cells discharge in the substantia nigra of the behaving monkey, *J. Neurosci.*, 3 (1983) 1599–1606.
[52] Descarries, L., Lemay, B., Doucet, G. and Berger, B., Regional and laminar density of the dopamine innervation in adult rat cerebral cortex, *Neuroscience*, 21 (1987) 807–824.
[53] Dismukes, K., New look at the aminergic nervous system, *Nature*, 269 (1977) 557–558.
[54] Dodt, H.U. and Misgeld, U., Muscarinic slow excitation and muscarinic inhibition of synaptic transmission in the rat neostriatum, *J. Physiol.*, 380 (1986) 593–608.
[55] Dudek, S.M. and Bear, M.F., Homosynaptic long-term depression in area CA1 of hippocampus and effects of *N*-methyl-D-aspartate receptor blockade, *Proc. Natl. Acad. Sci. USA*, 89 (1992) 4363–4367.
[56] Dunnett, S.B., Everitt, B.J. and Robbins, T.W., The basal forebrain-cortical cholinergic system: interpreting the functional consequences of excitotoxic lesions, *Trends Neurosci.*, 14 (1991) 494–501.
[57] Ettenberg, A., Dopamine, neuroleptics and reinforced behavior, *Neurosci. Biobehav. Rev.*, 13 (1989) 105–111.
[58] Evenden, J.L. and Robbins, T.W., Dissociable effects of d-amphetamine, chlordiazepoxide and alpha-flupentixol on choice and rate measures of reinforcement in the rat, *Psychopharmacology*, 79 (1983) 180–186.
[59] Everitt, B.J., Robbins, T.W., Gaskin, M. and Fray, P.J., The effects of lesions to ascending noradrenergic neurones on discrimination learning and performance in the rat, *Neuroscience*, 10 (1983) 397–410.
[60] Everitt, B.J., Cador, M. and Robbins, T.W., Interactions between the amygdala and ventral striatum in stimulus-reward associations: studies using a second-order schedule of sexual reinforcement, *Neuroscience*, 30 (1989) 63–75.

[61] Everitt, B.J., Morris, K.A., O'Brien, A. and Robbins, T.W., The basolateral amygdala-ventral striatal system and conditioned place preference: further evidence of limbic-striatal interactions underlying reward-related processes, *Neuroscience*, 42 (1991) 1–18.

[62] Ferron, A., Thierry, A.M., Le Douarin, C., Glowinski, J., Inhibitory influence of the mesocortical dopaminergic system on spontaneous activity or excitatory response induced from the thalamic mediodorsal nucleus in the rat medial prefrontal cortex, *Brain Res.*, 302 (1984) 257–265.

[63] Fibiger, H.C., Damsma, G. and Day, J.C., Behavioral pharmacology and biochemistry of central cholinergic neurotransmission, *Adv. Exp. Med. Biol.*, 295 (1991) 399–414.

[64] Fibiger, H.C. and Phillips, A.G., Reward, motivation, cognition: psychobiology of mesotelencephalic dopamine systems. In *Handbook of Physiology IV: Intrinsic Regulatory Systems of the Brain*, American Physiological Society, Bethesda, MD, 1986, pp. 647–675.

[65] Flood, J.F., Landry, D.W. and Jarvik, M.E., Cholinergic receptor interactions and their effects on long-term memory processing, *Brain Res.*, 215 (1981) 177–185.

[66] Flood, J.F., Smith, G.E. and Cherkin, A., Memory retention test performance in mice: improvement by chronic oral cholinergic drug treatment, *Pharmacol. Biochem. Behav.*, 21 (1984) 169–173.

[67] Foote, S.L., Aston-Jones, G. and Bloom, F.E., Impulse activity of locus coeruleus neurons in awake rats and monkeys is a function of sensory stimulation and arousal, *Proc. Natl. Acad. Sci. USA*, 77 (1980) 3033–3037.

[68] Foote, S.L. and Morrison, J.H., Extrathalamic modulation of cortical function, *Ann. Rev. Neurosci.*, 10 (1987) 67–95.

[69] Freedman, J.E. and Weight, F.F., Single K^+ channels activated by D2 dopamine receptors in acutely dissociated neurons from rat corpus striatum, *Proc. Natl. Acad. Sci. USA*, 85 (1988) 3618–3622.

[70] Frey, U., Schroeder, H. and Matthies, H., Dopaminergic antagonists prevent long-term maintenance of posttetanic LTP in the CA1 region of rat hippocampal slices, *Brain Res.*, 522 (1990) 69–75.

[71] Frey, U., Matthies, H., Reymann, K.G. and Matthies, H., The effect of dopaminergic D1 receptor blockade during tetanization on the expression of long-term potentiation in the rat CA1 region in vitro, *Neurosci. Lett.*, 129 (1991) 111–114.

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

[73] Garcia-Rill, E., The pedunculopontine nucleus, *Progr. Neurobiol.*, 36 (1991) 363–389.

[74] Gelperin, A., Tank, D.W. and Tesauro, G., Olfactory processing and associative memory: cellular and modeling studies. In J.H. Byrne and W.O. Berry (Eds.), *Neural Models of Plasticity*, Academic Press, San Diego, 1989, pp. 133–159.

[75] Grant, S.G.N., O'Dell, T.J., Karl, K.A., Stein, P.L., Soriano, P. and Kandel, E.R., Impaired long-term potentiation, spatial learning, and hippocampal development in fyn mutant mice, *Science*, 258 (1992) 1903–1910.

[76] Gray, R. and Johnston, D., Noradrenaline and β-adrenoceptor agonists increase activity of voltage-dependent calcium channels in hippocampal neurons, *Nature*, 327 (1987) 620–622.

[77] Groves, P.M., Linder, J.C. and Young, S.J., 5-Hydroxydopamine-labeled dopaminergic axons: three-dimensional reconstructions of axons, synapses and postsynaptic targets in rat neostriatum, *Neuroscience*, 58 (1994) 593–604.

[78] Haas, H.L. and Konnerth, A., Histamine and noradrenaline decrease calcium-activated potassium conductance in hippocampal pyramidal cells, *Nature*, 302 (1983) 432–434.

[79] Hagan, J.J. and Morris, R.G.M., The cholinergic hypothesis of memory: a review of animal experiments. In L.L. Iversen (Ed.), *Psychopharmacology of the Aging Nervous System*, Plenum Press, New York, 1989, pp. 237–324.

[80] Hagan, J.J., Alpert, J.E., Morris, R.G.M. and Iversen, S.D., The effects of central catecholamine depletions on spatial learning in rats, *Behav. Brain Res.*, 9 (1983) 83–104.

[81] Halliwell, J.V. and Adams, P.R., Voltage-clamp analysis of muscarinic excitation in hippocampal neurons, *Brain Res.*, 250 (1982) 71–92.

[82] Harrell, L.E., Barlow, T.S., Miller, M., Haring, J.H. and Davis, J.N., Facilitated reversal learning of a spatial-memory task by medial septal injections of 6-hydroxydopamine, *Exp. Neurol.*, 85 (1984) 69–77.

[83] Hasselmo, M.E., Anderson, B.P. and Bower, J.M., Cholinergic modulation of cortical associative memory function, *J. Neurophysiol.*, 67 (1992) 1230–1246.

[84] Heginbotham, L.R. and Dunwiddie, T.V., Long-term increases in the evoked population spike in the CA1 region of rat hippocampus induced by β-adrenergic receptor activation, *J. Neurosci.*, 11 (1991) 2519–2527.

[85] Heise, G.A., Conner, R. and Martin, R.A., Effects of scopolamine on variable intertrial interval spatial alternation and memory in the rat, *Psychopharmacology*, 49 (1976) 131–137.

[86] Hertz, J., Krogh, A. and Palmer, R.G., *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, 1991.

[87] Hestrin, S., Nicoll, R.A., Perkel, D.J. and Sah, P., Analysis of excitatory synaptic action in pyramidal cells using whole-cell recording from rat hippocampal slices, *J. Physiol.*, 422 (1990) 203–225.

[88] Hirotsu, I., Hori, N., Katsuda, N. and Ishihara, T., Effect of anticholinergic drug on long-term potentiation in rat hippocampal slices, *Brain Res.*, 482 (1989) 194–197.

[89] Hirsch, J.C. and Crepel, F., Use-dependent changes in synaptic efficacy in rat prefrontal neurons in vitro, *J. Physiol.*, 427 (1990) 31–49.

[90] Hopfield, J.J., Neurons with graded response have collective computational properties like those of two-state neurons, *Proc. Natl. Acad. Sci. USA*, 81 (1984) 3088–3092.

[91] Hopkins, W.F. and Johnston, D., Noradrenergic enhancement of long-term potentiation at mossy fiber synapses in the hippocampus, *J. Neurophysiol.*, 59 (1988) 667–687.

[92] Horn, A.S., Dopamine uptake: a review of progress in the last decade, *Progr. Neurobiol.*, 34 (1990) 387–400.

[93] Hornik, K., Stinchcombe, M. and White, H., Multilayer feedforward networks are universal approximators, *Neural Networks*, 2 (1989) 359–366.

[94] Houser, C.R., Crawford, G.D., Salvaterra, P.M. and Vaughn, J.E., Immunocytochemical localization of choline acetyltransferase in rat cerebral cortex: a study of cholinergic neurons and synapses, *J. Comp. Neurol.*, 234 (1985) 17–34.

[95] Imperato, A., Puglisi-Allegra, S., Casolini, P. and Angelucci, L., Changes in brain dopamine and acetylcholine release during and following stress are independent of the pituitary-adrenal axis, *Brain Res.*, 538 (1991) 111–117.

[96] Inoue, M., Oomura, Y., Nishino, H., Aou, S., Sikdar, S.K., Hynes, M., Mizuno, Y. and Katabuchi, T., Cholinergic role in monkey dorsolateral prefrontal cortex during bar-press feeding behaviour, *Brain Res.*, 278 (1983) 185–194.

[97] Izenwasser, S., Werling, L.L. and Cox, B.M., Comparison of the effects of cocaine and other inhibitors of dopamine uptake in rat striatum, nucleus accumbens, olfactory tubercle, and medial prefrontal cortex, *Brain Res.*, 520 (1990) 303–309.

[98] Jacobs, B.L., Single unit activity of locus coeruleus neurons in behaving animals, *Progr. Neurobiol.*, 27 (1986) 183–194.

[99] Jacobs, B.L. and Azmitia, E.C., Structure and function of the brain serotonin system, *Physiol. Rev.*, 72 (1992) 165–229.

[100] Jacobs, B.L., Heym, J. and Steinfels, G.F., Physiological and behavioral analysis of raphe unit activity. In L.L. Iversen, S.D. Iversen and S.H. Snyder (Eds.), *Handbook of Psychopharmacology, Vol. 18*, Plenum Press, New York, 1984, pp. 343–395.

[101] Katsuki, H., Saito, H. and Satoh, M., The involvement of muscarinic, β-adrenergic and metabotropic glutamate receptors in long-term potentiation in the fimbria-CA3 pathway of the hippocampus, *Neurosci. Lett.*, 142 (1992) 249–252.

[102] Keinänen, K., Wisden, W., Sommer, B., Werner, P., Herb, A., Verdoorn, T.A., Sakmann, B. and Seeburg, P.H., A family of AMPA-selective glutamate receptors, *Science*, 249 (1990) 556–560.

[103] Kety, S.S., The biogenic amines in the central nervous system: their possible roles in arousal, emotion, and learning. In F.O. Schmitt (Ed.), *The Neurosciences, Second Study Program*, Rockefeller University Press, New York, 1970, pp. 324–336.

[104] Kirkwood, A., Dudek, S.M., Gold, J.T., Aizenman, C.D. and Bear, M.F., Common forms of synaptic plasticity in the hippocampus and neocortex in vitro, *Science*, 260 (1993) 1518–1521.

[105] Knudsen, E.I., Supervised learning in the brain, *J. Neurosci.*, 14 (1994) 3985–3997.

[106] Kohonen, T., *Self-organization and Associative Memory*, Springer-Verlag, Berlin, 1989.

[107] Koob, G.F., Drugs of abuse: anatomy, pharmacology and function of reward pathways, *Trends Pharmacol. Sci.*, 13 (1992) 177–184.

[108] Ksir, C.J., Scopolamine effects on two-trial delayed-response performance in the rat, *Psychopharmacologia* (Berlin), 34 (1974) 127–134.

[109] Kurumiya, S. and Nakajima, S., Dopamine D1 receptors in the nucleus accumbens: involvement in the reinforcing effect of tegmental stimulation, *Brain Res.*, 448 (1988) 1–6.

[110] Langlais, P.J., Connor, D.J. and Thal, L., Comparison of the effects of single and combined neurotoxic lesions of the nucleus basalis magnocellularis and dorsal noradrenergic bundle on learning and memory in the rat, *Behav. Brain Res.*, 54 (1993) 81–90.

[111] LaRoche, S., Doyere, V. and Bloch, V., Linear relation between the magnitude of long-term potentiation in the dentate gyrus and associative learning in the rat. A demonstration using commissural inhibition and local infusion of an *N*-methyl-D-aspartate receptor antagonist, *Neuroscience*, 28 (1989) 375–386.

[112] LeDoux, J.E., Emotion and the amygdala. In J.P. Aggleton (Ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, Wiley-Liss, New York, 1992, pp. 339–351.

[113] Linsker, R., Perceptual neural organization: some approaches based on network models and information theory, *Ann. Rev. Neurosci.*, 13 (1990) 257–281.

[114] Lisberger, S.G. and Sejnowski, T.J., Motor learning in a recurrent network model based on the vestibulo-ocular reflex, *Nature*, 360 (1992) 159–161.

[115] Lisman, J., A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory, *Proc. Natl. Acad. Sci. USA*, 86 (1989) 9574–9578.

[116] Ljungberg, T., Apicella, P. and Schultz, W., Responses of monkey dopamine neurons to external stimuli: changes with learning. In G. Bernardi, M.B. Carpenter, G. Di Chiara, M. Morelli and P. Stanzione (Eds.), *The Basal Ganglia III*, Plenum Press, New York, 1991, pp. 487–494.

[117] Ljungberg, T., Apicella, P. and Schultz, W., Responses of monkey dopamine neurons during learning of behavioral reactions, *J. Neurophysiol.*, 67 (1992) 145–163.

[118] Losier, B.J. and Semba, K., Dual projections of single cholinergic and aminergic brainstem neurons to the thalamus and basal forebrain in the rat, *Brain Res.*, 604 (1993) 41–52.

[119] Lubow, R., *Latent Inhibition and Conditioned Attention Theory*, Cambridge University Press, Cambridge UK, 1989.

[120] Madison, D.V. and Nicoll, R.A., Noradrenaline blocks accomodation of pyramidal cell discharge in the hippocampus, *Nature*, 299 (1982) 636–638.

[121] Maeda, T., Kaneko, S. and Satoh, M., Bidirectional modulation of long-term potentiation by carbachol via M1 and M2 muscarinic receptors in guinea pig hippocampal mossy fiber-CA3 synapses, *Brain Res.*, 619 (1993) 324–330.

[122] Malenka, R.C. and Nicoll, R.A., NMDA-receptor-dependent synaptic plasticity: multiple forms and mechanisms, *Trends. Neurosci.*, 16 (1993) 521–527.

[123] Mantz, J., Thierry, A.M. and Glowinski, J., Effect of noxious tail pinch on the discharge rate of mesocortical and mesolimbic dopamine neurons: selective activation of the mesocortical system, *Brain Res.*, 476 (1989) 377–381.

[124] Marston, H.M., West, H.L., Wilkinson, L.S., Everitt, B.J. and Robbins, T.W., Effects of excitotoxic lesions of the septum and vertical limb nucleus of the diagonal band of Broca on conditional visual discrimination: relationship between performance and choline acetyltransferase activity in the cingulate cortex, *J. Neurosci.*, 14 (1994) 2009–2019.

[125] Mason, S.T. and Fibiger, H.C., Noradrenaline, fear and extinction, *Brain Res.*, 165 (1979) 47–56.

[126] Mason, S.T. and Iversen, S.D., Learning in the absence of forebrain noradrenaline, *Nature*, 258 (1975) 422–424.

[127] Mason, S.T. and Iversen, S.D., An investigation of the role of cortical and cerebellar noradrenaline in associative motor learning in the rat, *Brain Res.*, 134 (1977) 513–527.

[128] Mason, S.T. and Iversen, S.D., Theories of the dorsal bundle extinction effect, *Brain Res. Rev.*, 1 (1979) 107–137.

[129] Mason, S.T. and Lin, D., Dorsal noradrenergic bundle and selective attention in the rat, *J. Comp. Physiol. Psychol.*, 94 (1980) 819–832.

[130] Mazzoni, P., Andersen, R.A. and Jordan, M.I., A more biologically plausible learning rule for neural networks, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 4433–4437.

[131] McCormick, D.A. and Prince, D.A., Acetylcholine induces burst firing in thalamic reticular neurones by activating a potassium conductance, *Nature*, 319 (1986) 402–405.

[132] McGaugh, J.L., Involvement of hormonal and neuromodulatory systems in the regulation of memory storage, *Ann. Rev. Neurosci.*, 12 (1989) 255–287.

[133] Mesulam, M.M., Mufson, E.J., Levey, A.I. and Wainer, B.H., Cholinergic innervation of cortex by the basal forebrain: cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (substantia innominata), and hypothalamus in the rhesus monkey, *J. Comp. Neurol.*, 214 (1983) 170–197.

[134] Metherate, R., Cox, C.L. and Ashe, J.H., Cellular bases of neocortical activation: modulation of neural oscillations by the nucleus basalis and endogenous acetylcholine, *J. Neurosci.*, 12 (1992) 4701–4711.

[135] Metherate, R., Tremblay, N. and Dykes, R.W., Acetylcholine permits long-term enhancement of neuronal responsiveness in cat primary somatosensory cortex, *Neuroscience*, 22 (1987) 75–81.

[136] Miller, J.D., Sanghera, M.K. and German, D.C., Mesencephalic dopaminergic unit activity in the behaviorally conditioned rat, *Life Sci.*, 29 (1981) 1255–1263.

[137] Miller, R., Wickens, J.R. and Beninger, R.J., Dopamine D-1 and D-2 receptors in relation to reward and performance: a case for the D-1 receptor as a primary site of therapeutic action of neuroleptic drugs, *Progr. Neurobiol.*, 34 (1990) 143–183.

[138] Mirenowicz, J. and Schultz, W., Activity of dopamine neurons in monkey performing an active avoidance task, *Soc. Neurosci. Abstr.*, 19 (1993) 1585.

[139] Mirenowicz, J. and Schultz, W., Importance of unpredictability for reward responses in primate dopamine neurons, *J. Neurophysiol.*, 72 (1994) 1024–1027.

[140] Mody, I., Baimbridge, K.G. and Miller, J.J., Blockade of tetanic- and calcium-induced long-term potentiation in the hippocampal slice preparation by neuroleptics, *Neuropharmacology*, 23 (1984) 625–631.

[141] Monaghan, D.T., Bridges, R.J. and Cotman, C.W., The excitatory amino acid receptors: their classes, pharmacology, and distinct properties in the function of the central nervous system, *Ann. Rev. Pharmacol. Toxicol.*, 29 (1989) 365–402.

[142] Montague, P.R., Dayan, P., Nowlan, S.J., Pouget, A. and Sej-

nowski, T.J., Using aperiodic reinforcement for directed self-organization during development. In S.J. Hanson, J.D. Cowan and C.L. Giles (Eds.), *Advances in Neural Information Processing Systems, Vol. 5*, Morgan Kaufmann, San Mateo CA, 1993.

[143] Montague, P.R. and Sejnowski, T.J., The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms, *Learning and Memory*, 1 (1994) 1–33.

[144] Morris, R.G.M., Does synaptic plasticity play a role in information storage in the vertebrate brain? In R.G.M. Morris (Ed.), *Parallel Distributed Processing. Implications for Psychology and Neurobiology*, Clarendon, Oxford, 1989, pp. 248–285.

[145] Moser, E., Moser, M.-B. and Andersen, P.A., Synaptic potentiation in the rat dentate gyrus during exploratory learning, *Neuroreport*, 5 (1993) 317–320.

[146] Muir, J.L., Everitt, B.J. and Robbins, T.W., AMPA-induced excitotoxic lesions of the basal forebrain: a significant role for the cortical cholinergic system in attentional function, *J. Neurosci.*, 14 (1994) 2313–2326.

[147] Mulkey, R.M. and Malenka, R.C., Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus, *Neuron*, 9 (1992) 967–975.

[148] Nakajima, S., Suppression of operant responding in the rat by dopamine D1 receptor blockade with Sch 23390, *Physiol. Psychol.*, 14 (1986) 111–114.

[149] Nicoll, R.A., Malenka, R.C. and Kauer, J.A., Functional comparison of neurotransmitter receptor subtypes in mammalian central nervous system, *Physiol. Rev.*, 70 (1990) 513–565.

[150] Niki, H. and Watanabe, M., Prefrontal and cingulate unit activity during timing behavior in the monkey, *Brain Res.*, 171 (1979) 213–224.

[151] Nishijo, H., Ono, T. and Nishino, H., Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance, *J. Neurosci.*, 8 (1988) 3570–3583.

[152] Papadopoulos, G.C., Parnavelas, J.G. and Buijs, R.M., Light and electron microscopic immunocytochemical analysis of the noradrenaline innervation of the rat visual cortex, *J. Neurocytol.*, 18 (1989a) 1–10.

[153] Papadopoulos, G.C., Parnavelas, J.G. and Buijs, R.M., Light and electron microscopic immunocytochemical analysis of the dopamine innervation of the rat visual cortex, *J. Neurocytol.*, 18 (1989b) 303–310.

[154] Parnavelas, J.G. and Papadopoulos, G.C., The monoaminergic innervation of the cerebral cortex is not diffuse and nonspecific, *Trends Neurosci.*, 12 (1989) 315–319.

[155] Pedarzani, P. and Storm, J.F., PKA mediates the effects of monoamine transmitters on the K+ current underlying the slow spike frequency adaptation in hippocampal neurons, *Neuron*, 11 (1993) 1023–1035.

[156] Pennartz, C.M.A., Dolleman-van der Weel, M.J., Kitai, S.T. and Lopes da Silva, F.H., Presynaptic dopamine D1 receptors attenuate excitatory and inhibitory limbic inputs to the shell region of the rat nucleus accumbens studied in vitro, *J. Neurophysiol.*, 67 (1992) 1325–1334.

[157] Pennartz, C.M.A., Ameerun, R.F., Groenewegen, H.J. and Lopes da Silva, F.H., Synaptic plasticity in an in vitro slice preparation of the rat nucleus accumbens, *Eur. J. Neurosci.*, 5 (1993) 107–117.

[158] Pennartz, C.M.A., Groenewegen, H.J. and Lopes da Silva, F.H., The nucleus accumbens as a complex of functionally distinct neuronal ensembles: an integration of behavioural, electrophysiological and anatomical data, *Progr. Neurobiol.*, 42 (1994) 719–761.

[159] Pennartz, C.M.A. and Lopes da Silva, F.H., Muscarinic modulation of synaptic transmission in slices of the rat ventral striatum is dependent on the frequency of afferent stimulation, *Brain Res.*, 645 (1994) 231–239.

[160] Pirot, S., Godbout, R., Mantz, J., Tassin, J.P., Glowinski, J. and Thierry, A.M., Inhibitory effects of ventral tegmental area stimulation on the activity of prefrontal cortical neurons: evidence for the involvement of both dopaminergic and GABAergic components, *Neuroscience*, 49 (1992) 857–865.

[161] Pisa, M. and Fibiger, H.C., Evidence against a role of the rat's dorsal noradrenergic bundle in selective attention and place memory, *Brain Res.*, 272 (1983) 319–329.

[162] Pisa, M., Martin-Iverson, M.T. and Fibiger, H.C., On the role of the dorsal noradrenergic bundle in learning and habituation to novelty, *Pharmacol. Biochem. Behav.*, 30 (1988) 835–845.

[163] Pontecorvo, M.J. and Evans, H.L., Effects of aniracetam on delayed matching-to-sample performance of monkeys and pigeons, *Pharmacol. Biochem. Behav.*, 22 (1985) 745–752.

[164] Rasmussen, K. and Jacobs, B.L., Single unit activity of locus coeruleus neurons in the freely moving cat. II. Conditioning and pharmacologic studies, *Brain Res.*, 371 (1986) 335–344.

[165] Redmond, D.E., New and old evidence for the involvement of a brain norepinephrine system in anxiety. In W.E. Fann, I. Karacan, A.D. Pokorny and R.L. Williams (Eds.), *Phenomenology and Treatment of Anxiety*, Spectrum, New York, 1979, pp. 153–203.

[166] Rescorla, R.A. and Wagner, A.R., A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black and W.F. Prokasy (Eds.), *Classical Conditioning. II: Current Research and Theory*, Appleton Century Crofts, New York, 1972, pp. 64–99.

[167] Richardson, R.T. and DeLong, M.R., Nucleus basalis of Meynert neuronal activity during a delayed response task in monkey, *Brain Res.*, 399 (1986) 364–368.

[168] Robbins, T.W. and Everitt, B.J., Comparative functions of the central noradrenergic, dopaminergic and cholinergic systems, *Neuropharmacology*, 26 (1987) 893–901.

[169] Robbins T.W., and Everitt, B.J., Functions of dopamine in the dorsal and ventral striatum, *Seminars Neurosci.*, 4 (1992) 119–127.

[170] Robbins, T.W., Cador, M., Taylor, J.R. and Everitt, B.J., Limbic-striatal interactions in reward-related processes, *Neurosci. Biobehav. Rev.*, 13 (1989a) 155–162.

[171] Robbins, T.W., Everitt, B.J., Marston, H.M., Wilkinson, J., Jones, G.H. and Page, K.J., Comparative effects of ibotenic acid- and quisqualic acid-induced lesions of the substantia innominata on attentional function in the rat: further implications for the role of the cholinergic neurons of the nucleus basalis in cognitive processes, *Behav. Brain Res.*, 35 (1989b) 221–240.

[172] Robinson, T.E. and Berridge, K.C., The neural basis of drug craving: an incentive-sensitization theory of addiction, *Brain Res. Rev.*, 18 (1993) 247–291.

[173] Rolls, E.T., Neurophysiology and functions of the primate amygdala. In J.P. Aggleton (Ed.), *The Amygdala: Neurobiological Aspects of Emotion, Memory and Mental Dysfunction*, Wiley-Liss, New York, 1992, pp. 143–165.

[174] Rolls, E.T., Sanghera, M.K. and Roper-Hall, A., The latency of activation of neurones in the lateral hypothalamus and substantia innominata during feeding in the monkey, *Brain Res.*, 164 (1979) 121–135.

[175] Rosenkilde, C.E., Bauer, R.H. and Fuster, J.M., Single cell activity in ventral prefrontal cortex of behaving monkeys, *Brain Res.*, 209 (1981) 375–394.

[176] Roth, R.H., Tam, S.Y., Ida, Y., Yang, J.X. and Deutch, A.Y., Stress and the mesocorticolimbic dopamine systems, *Ann. N.Y. Acad. Sci.*, 537 (1988) 138–147.

[177] Rumelhart, D.E., Hinton, G.E. and Williams, R.J., Learning internal representations by error propagation. In D.E. Rumelhart, J.L. McClelland and the PDP research group (Eds.), *Parallel Distributed Processing, Vol. 1*, Ch. 8, MIT press, Cambridge Mass., 1986.

[178] Sara, S.J., The locus coeruleus and cognitive function: attempts to relate noradrenergic enhancement of signal/noise in the brain to behavior, *Physiol. Psychol.*, 13 (1985) 151–162.

[179] Scheel-Krüger, J. and Willner, P., The mesolimbic system: principles of operation. In P. Willner and J. Scheel-Krüger (Eds.), *The*

Mesolimbic Dopamine System: from Motivation to Action, Wiley, Chichester UK, 1991, pp. 559–597.

[180] Schultz, W., Apicella, P. and Ljungberg, T., Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task, *J. Neurosci.*, 13 (1993) 900–913.

[181] Schultz, W., Apicella, P., Scarnati, E. and Ljungberg, T., Neuronal activity in monkey ventral striatum related to the expectation of reward, *J. Neurosci.*, 12 (1992) 4595–4610.

[182] Schultz, W., Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey, *J. Neurophysiol.*, 56 (1986) 1439–1461.

[183] Schultz, W. and Romo, R., Dopamine neurons of the monkey midbrain: contingencies of responses to stimuli eliciting immediate behavioural reactions, *J. Neurophysiol.*, 63 (1990) 607–624.

[184] Schwartz, J.C., Arrang, J.M., Garbarg, M., Pollard, H. and Ruat, M., Histaminergic transmission in the mammalian brain, *Physiol. Rev.*, 71 (1991) 1–51.

[185] Selden, N.R.W., Robbins, T.W. and Everitt, B.J., Enhanced behavioral conditioning to context and impaired behavioral and neuroendocrine responses to conditioned stimuli following ceruleocortical noradrenergic lesions: support for an attentional hypothesis of central noradrenergic function, *J. Neurosci.*,10 (1990) 531–539.

[186] Sesack, S.R. and Bunney, B.S., Pharmacological characterization of the receptor mediating electrophysiological responses to dopamine in the rat medial prefrontal cortex: a microiontophoretic study, *J. Pharmacol. Exp. Ther.*, 248 (1989) 1323–1333.

[187] Sesack, S.R. and Pickel, V.M., In the rat medial nucleus accumbens, hippocampal and catecholaminergic terminals converge on spiny neurons and are in apposition to each other, *Brain Res.*, 527 (1990) 266–279.

[188] Sillito, A.M. and Murphy, P.C., The cholinergic modulation of cortical function. In E.G. Jones and A. Peters (Eds.), *Cerebral Cortex, Vol. 6*, Plenum Press, New York, 1987, pp. 161–185.

[189] Silva, A.J., Paylor, R., Wehner, J.M. and Tonegawa, S., Impaired spatial learning in α-calcium-calmodulin kinase II mutant mice, *Science*, 257 (1992) 206–211.

[190] Smith, A.A. and Calhoun, W.H., Scopolamine: effects on conditioned suppression, *Neuropharmacology*, 11 (1972) 347–350.

[191] Squire, L.R., *Memory and Brain*, Oxford University Press, New York, 1987.

[192] Stanton, P.K. and Sarvey, J.M., Depletion of norepineprine, but not serotonin, reduces long-term potentiation in the dentate gyrus of rat hippocampal slices, *J. Neurosci.*, 5 (1985a) 2169–2176.

[193] Stanton, P.K. and Sarvey, J.M., Blockade of norepinephrine-induced long-lasting potentiation in the hippocampal dentate gyrus by an inhibitor of protein synthesis, *Brain Res.*, 361 (1985b) 276–283.

[194] Steinfels, G.F., Heym, J., Strecker, R.E. and Jacobs, B.L., Behavioral correlates of dopaminergic unit activity in freely moving cats, *Brain Res.*, 258 (1983) 217–228.

[195] Steriade, M., Paré, D., Datta, S., Oakson, G. and Curró Dossi, R., Different cellular types in mesopontine cholinergic nuclei related to ponto-geniculo-occipital waves, *J. Neurosci.*, 10 (1990a) 2560–2579.

[196] Steriade, M., Datta, S., Paré, D., Oakson, G. and Curró Dossi, R., Neuronal activities in brain-stem cholinergic nuclei related to tonic activation processes in thalamocortical systems, *J. Neurosci.*, 10 (1990b) 2541–2559.

[197] Strecker, R.E., Steinfels, G.F. and Jacobs, B.L., Dopaminergic unit activity in freely moving cats: lack of relationship to feeding, satiety, and glucose injections, *Brain Res.*, 260 (1983) 317–321.

[198] Strecker, R.E. and Jacobs, B.L., Substantia nigra dopaminergic unit activity in behaving cats: effects of arousal on spontaneous discharge and sensory evoked activity, *Brain Res.*, 361 (1985) 339–350.

[199] Sugita, S., Uchimura, N., Jiang, Z.G. and North, R.A., Distinct muscarinic receptors inhibit release of γ-aminobutyric acid and excitatory amino acids in mammalian brain, *Proc. Natl. Acad. Sci. USA*, 88 (1991) 2608–2611.

[200] Suits, E. and Isaacson, R.L., The effects of scopolamine hydrobromide on one-way and two-way avoidance learning in rats, *Int. J. Neuropharmacol.*, 7 (1968) 441–446.

[201] Sullivan, R.M., McGaugh, J.L. and Leon, M., Norepinephrine-induced plasticity and one-trial olfactory learning in neonatal rats, *Dev. Brain Res.*, 60 (1991) 219–228.

[202] Surmeier, D.J., Eberwine, J., Wilson, C.J., Cao, Y., Stefani, A. and Kitai, S.T., Dopamine receptor subtypes colocalize in rat striatonigral neurons, *Proc. Natl. Acad. Sci. USA*, 89 (1992) 10178–10182.

[203] Sutton, R.S. and Barto, A.G., A temporal-difference model of classical conditioning. In *Ninth Annual Conf. Cogn. Society*, Lawrence Erlbaum Assoc., 1987, pp. 355–378.

[204] Sutton, R.S., Learning to predict by the methods of temporal differences, *Machine Learning*, 3 (1988) 9–44.

[205] Tanaka, Y., Sakurai, M. and Hayashi, S., Effect of scopolamine and HP 029, a cholinesterase inhibitor, on long-term potentiation in hippocampal slices of the guinea pig, *Neurosci. Lett.*, 98 (1989) 179–183.

[206] Taylor, J.R. and Robbins, T.W., 6-Hydroxydopamine lesions of the nucleus accumbens, but not of the caudate nucleus, attenuate enhanced responding with reward-related stimuli produced by intra-accumbens d-amphetamine, *Psychopharmacology*, 90 (1986) 390–397.

[207] Tennyson, V.M., Heikkila, R., Mytilineou, C., Coté, L. and Cohen, G., 5-Hydroxydopamine 'tagged' neuronal boutons in rabbit neostriatum: interrelationship between vesicles and axonal membrane, *Brain Res.*, 82 (1974) 341–348.

[208] Tesauro, G., Simple neural models of classical conditioning, *Biol. Cybern.*, 55 (1986) 187–200.

[209] Thierry, A.M., Tassin, J.P., Blanc, G. and Glowinski, J., Selective activation of the mesocortical DA system by stress, *Nature*, 263 (1976) 242–244.

[210] Thompson, R.F. and Krupa, D.J., Organization of memory traces in the mammalian brain, *Ann. Rev. Neurosci.*, 17 (1994) 519–549.

[211] Thorpe, S.J., Rolls, E.T. and Maddison, S., The orbitofrontal cortex: neuronal activity in the behaving monkey, *Exp. Brain Res.*, 49 (1983) 93–115.

[212] Uchimura, N. and North, R.A., Actions of cocaine on rat nucleus accumbens neurones in vitro, *Br. J. Pharmacol.*, 99 (1990) 736–740.

[213] Valentino, R.J. and Dingledine, R., Presynaptic inhibitory effect of acetylcholine in the hippocampus, *J. Neurosci.*, 1 (1981) 784–792.

[214] Vidal, C. and Changeux, J.P., Nicotinic and muscarinic modulations of excitatory synaptic transmission in the rat prefrontal cortex in vitro, *Neuroscience*, 56 (1993) 23–32.

[215] Voorn, P., Jorritsma-Byham, B., Van Dijk, C. and Buijs, R.M., The dopaminergic innervation of the ventral striatum in the rat: a light- and electron-microscopical study with antibodies against dopamine, *J. Comp. Neurol.*, 251 (1986) 84–99.

[216] Wada, H., Inagaki, N., Yamatodani, A. and Watanabe, T., Is the histaminergic neuron system a regulatory center for whole-brain activity?, *Trends Neurosci.*, 14 (1991) 415–418.

[217] Wainer, B.H., Levey, A.I., Mufson, E.J. and Mesulam, M.M., Cholinergic systems in mammalian brain identified with antibodies against choline acetyltransferase, *Neurochem. Intl.*, 6 (1984) 163–182.

[218] Waterhouse, B.D., Moises, H.C. and Woodward, D.J., Noradrenergic modulation of somatosensory cortical neuronal responses to iontophoretically applied putative neurotransmitters, *Exp. Neurol.*, 69 (1980) 30–49.

[219] Watson, M. and McElligott, J.G., Cerebellar norepinephrine depletion and impaired acquisition of specific locomotor tasks in rats, *Brain Res.*, 296 (1984) 129–138.

[220] Weiss, B. and Wallace, T.L., Mechanisms and pharmacological implications of altering calmodulin activity. In W.Y. Cheung (Ed.),

Calcium and Cell Function, Vol. 1, Academic Press, New York, 1980, pp. 329–379.

[221] Wenk, G.L., Markowska, A.L. and Olton, D.S., Basal forebrain lesions and memory: alterations in neurotensin, not acetylcholine, may cause amnesia, *Behav. Neurosci.*, 103 (1989) 765–769.

[222] Werbos, P., *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D thesis, Harvard University, 1974.

[223] Werbos, P.J., Building and understanding adaptive systems: a statistical/numerical approach to factory automation and brain research, *IEEE Transact. Syst. Man Cybern.*, 17 (1987) 7–20.

[224] White, N.M., Packard, M.G. and Hiroi, N., Place conditioning with dopamine D1 and D2 agonists injected peripherally or into nucleus accumbens, *Psychopharmacology*, 103 (1991) 271–276.

[225] White, N.M. and Milner, P.M., The psychobiology of reinforcers, *Ann. Rev. Psychol.*, 43 (1992) 443–471.

[226] Wickens, J. and Kötter, R., Cellular models of reinforcement. In J.C. Houk., J.L. Davis and D.G. Beiser (Eds.), *Models of Information Processing in the Basal Ganglia*, MIT Press, Cambridge Mass., 1995, pp. 187–214.

[227] Widrow, B., Gupta, N.K. and Maitra, S., Punish/reward: learning with a critic in adaptive threshold systems, *IEEE Transact. Syst. Man Cybern.*, 3 (1973) 455–465.

[228] Williams, S. and Johnston, D., Muscarinic depression of long-term potentiation in CA3 hippocampal neurons, *Science*, 242 (1988) 84–87.

[229] Wilson, F.A.W. and Rolls, E.T., Neuronal responses related to reinforcement in the primate basal forebrain, *Brain Res.*, 509 (1990) 213–231.

[230] Wilson, D.A. and Sullivan, R.M., Olfactory associative conditioning in infant rats with brain stimulation as reward: II. Norepinephrine mediates a specific component of the bulb response to reward, *Behav. Neurosci.*, 105 (1991) 843–849.

[231] Wise, R.A. and Hoffman, D.C., Localization of drug reward mechanisms by intracranial injections, *Synapse*, 10 (1992) 247–263.

[232] Wise, R.A., Spindler, J., DeWit, H. and Gerber, G.J., Neuroleptic-induced 'anhedonia' in rats: pimozide blocks reward quality of food, *Science*, 201 (1978) 262–264.

[233] Yang, C.R. and Mogenson, G.J., Dopamine enhances terminal excitability of hippocampal–accumbens neurons via D2 receptor: role of dopamine in presynaptic inhibition, *J. Neurosci.*, 6 (1986) 2470–2478.

[234] Yim, C.Y. and Mogenson, G.J., Neuromodulatory action of dopamine in the nucleus accumbens: an in vivo intracellular study, *Neuroscience*, 26 (1988) 403–415.

[235] Zito, K.A., Vickers, G. and Roberts, D.C.S., Disruption of cocaine and heroin self-administration following kainic acid lesions of the nucleus accumbens, *Pharmacol. Biochem. Behav.*, 23 (1985) 1029–1036.