The Reward Signal of Midbrain Dopamine Neurons

Wolfram Schultz

Dopamine projections from the midbrain to striatum and frontal cortex play a major role in behavioral reactions controlled by rewards. Recent experiments have shown that dopamine neurons code the discrepancy between the prediction and occurrence of rewards and in this way signal a crucial learning term for approach behavior.

When multicellular organisms arose through the evolution of self-reproducing molecules, they developed endogenous, autoregulatory mechanisms that assured that their needs for welfare and survival were met. Subjects engage in various forms of approach behavior to obtain resources for maintaining homeostatic balance and to reproduce themselves. These biological resources are said to have "rewarding" functions because they elicit and reinforce approach behavior. Although initially related to biological needs, rewards developed further during the evolution of higher mammals to support more sophisticated forms of individual and social behavior. Higher forms of rewards are often based on cognitive representations, and they concern such objects and constructs as novelty, challenge, acclaim, power, money, territory, and security. Thus biological and cognitive needs define the nature of rewards, and the availability of rewards determines some of the basic parameters of the subject's life conditions.

Rewards have three basic functions (1). First, they elicit approach and consummatory behavior and serve as goals of voluntary behavior. In doing so, they interrupt ongoing behavior and change the priorities of behavioral actions. Second, rewards have positive reinforcing effects. They increase the frequency and intensity of behavior leading to such objects (learning) and maintain learned behavior by preventing extinction. This function constitutes the essence of "coming back for more" and relates to the notion of receiving rewards for having done something useful. Learning proceeds when rewards occur unpredictably and slows as rewards become more and more predicted (8). Thus, reward-driven learning depends on the discrepancy or "error" between the prediction of reward and its actual occurrence. In their third function, rewards induce subjective feelings of pleasure (hedonia) and positive emotional states. This function is difficult to investigate in animals.

Reduced dopamine neurotransmission in parkinsonian patients and experimentally lesioned animals is associated with severe deficits in movement, motivation, attention, and cognition. One consistent motivational deficit concerns the use of reward information for learning and maintaining approach and consummatory behavior (2, 9). The deficits occur mainly with destruction of projections from midbrain dopamine neurons to the nucleus accumbens and, to a lesser extent, to frontal cortex and striatum (caudate nucleus and putamen). These systems are also involved in the addictive properties of major drugs of abuse, such as cocaine, amphetamine, heroin, and nicotine.

Responses of dopamine neurons in behaving primates

Cell bodies of midbrain dopamine neurons are located in groups A8 (dorsal to lateral substantia nigra), A9 (pars compacta of substantia nigra),

"Rewards have three basic functions…."

W. Schultz is in the Institute of Physiology of the University of Fribourg, CH-1700 Fribourg, Switzerland.
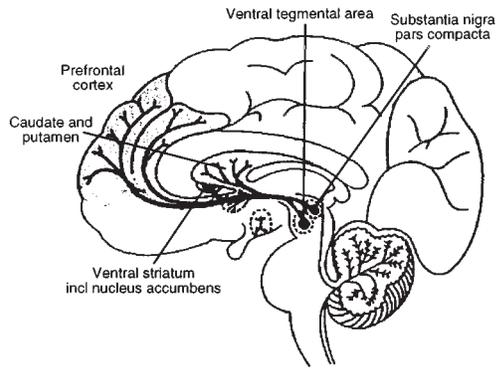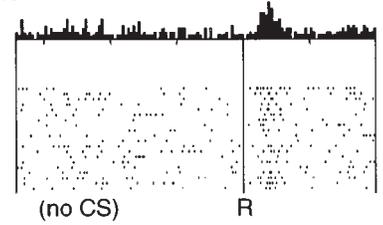
FIGURE 1. Overview of the projection territories of midbrain dopamine neurons. Cell bodies of dopamine neurons are located mainly in the pars compacta of substantia nigra and the medially adjoining ventral tegmental area. Their axons project mainly to the striatum (caudate nucleus, putamen), ventral striatum including nucleus accumbens, and frontal cortex (dorsolateral, ventrolateral and orbital prefrontal cortex). Dopamine is released from axon terminals with impulses and influences neurons in these structures. Our experiments investigate the impulse activity at the level of dopamine cell bodies.

FIGURE 2. Dopamine neurons report rewards as far as they deviate from prediction (error in prediction of reward). *Top*: a drop of liquid reward (R) occurs, although no reward is predicted at this time. The occurrence of reward thus constitutes a positive error in the prediction of reward. The dopamine neuron is activated by the unpredicted occurrence of the liquid. *Middle*: a learned stimulus (conditioned stimulus, CS) predicts a reward, and the reward occurs according to the prediction, hence no error in the prediction of reward. The dopamine neuron fails to be activated by the predicted reward (*right*). It also shows an activation following the reward-predicting stimulus (*left*). *Bottom*: a conditioned stimulus predicts a reward, but the reward is omitted. The activity of the dopamine neuron is depressed exactly at the time when the reward would have occurred. Note the depression occurring >1 s after the conditioned stimulus without any intervening stimuli, revealing an internal process of expectation of reward at a precise moment. The dots in the individual panels denote neuronal impulses. Each line of dots shows one trial, the original sequence in each panel being from *top* to *bottom*. Reprinted with permission from Ref. 11. Copyright (1997) American Association for the Advancement of Science.
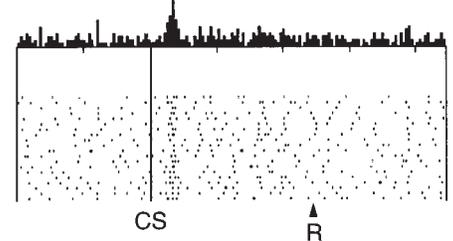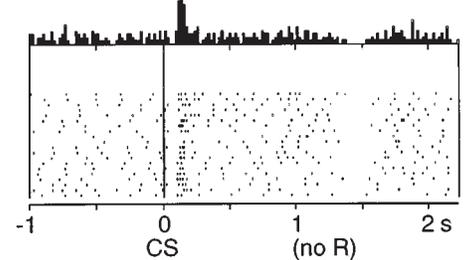
and A10 (ventral tegmental area medial to substantia nigra). These neurons release dopamine with nerve impulses from axonal varicosities in the striatum, nucleus accumbens, and frontal cortex, to name the most important sites (Fig. 1). We record the impulse activity from cell bodies of single dopamine neurons during periods of 20-60 min with movable microelectrodes from extracellular positions while monkeys learn and perform behavioral tasks. The neurons are easily distinguishable from other midbrain neurons by their characteristic polyphasic, relatively long impulses discharged at low frequencies.

We consistently fail to find clear covariations with movements. By contrast, dopamine neurons show phasic activations after reward-related events and certain attention-inducing stimuli of the somatosensory, visual, and auditory modality (5, 10, 12). These responses occur in a very similar manner in 60-80% of neurons in groups A8, A9, and A10 in a range of behavioral situations, whereas the remaining dopamine neurons do not respond at all. Tested situations include classical conditioning, various simple and choice reaction time tasks, direct and delayed go-no go tasks, spatial delayed-response task, spatial delayed alternation, visual discrimination, and self-initiated movements. Neurons respond slightly more in medial midbrain regions, such as the ventral tegmental area and medial substantia nigra, compared with more lateral regions, a difference that occasionally reaches statistical significance. The activations occur with similar laten-

cies (50–110 ms) and durations (<200 ms) after food and fluid rewards, conditioned stimuli, and attention-inducing stimuli. Thus the dopamine response constitutes a relatively homogeneous, scalar population signal that is graded by the response magnitude of individual neurons and by the fractions of neurons responding.

Phasic activations occur when animals touch a morsel of hidden food or when drops of liquid are delivered to their mouth outside of behavioral tasks or during learning (Fig. 2, *top*). Dopamine neurons distinguish rewards from nonreward objects but do not appear to discriminate between different food objects or liquid rewards. Only a few show the phasic activations after primary aversive stimuli, such as nonnoxious air puffs to the hand or hypertonic saline to the mouth (7). These stimuli are aversive because they disrupt behavior and induce active avoidance reactions.

Most dopamine neurons are also activated by conditioned visual and auditory stimuli that have become valid reward predictors through repeated and contingent pairing with rewards in operant or classical conditioning procedures (Fig. 2, *middle*). In contrast, only a few dopamine neurons are phasically activated by learned visual or auditory stimuli in active avoidance tasks in which animals release a key to avoid an air puff or a drop of hypertonic saline.

Concurrently with the development of the dopamine response to reward-predicting stimuli during learning, the response to the predicted reward itself is lost, as if the response is transferred from the reward to the reward-predicting stimulus (Fig. 2, *top* vs. *middle*). This is observed when free rewards are delivered outside of behavioral tasks and become predicted by conditioned stimuli through learning or when rewards occur surprisingly during individual learning phases and become predicted when a phase is fully acquired. Thus rewards are only effective in activating dopamine neurons when they are not predicted by phasic stimuli.

Dopamine neurons have a limited capacity to discriminate between appetitive and neutral or aversive stimuli. Only stimuli that are physically sufficiently dissimilar are well discriminated. Stimuli that do not explicitly predict rewards but physically resemble reward-predicting stimuli induce small activations followed by depressions in a limited fraction of neurons.

Dopamine neurons are depressed at the habitual time of reward when a predicted reward fails to occur after an error of the animal, withholding by the experimenter, or delayed delivery (Fig. 2, *bottom*). The depression occurs in the absence of a stimulus immediately preceding the omitted reward. This reflects an expectation process based on an internal clock that concerns the precise time of the predicted reward. On the other hand, an activation follows the reward when this is presented at a different time than predicted (Fig. 3). These data suggest that the prediction influenc-
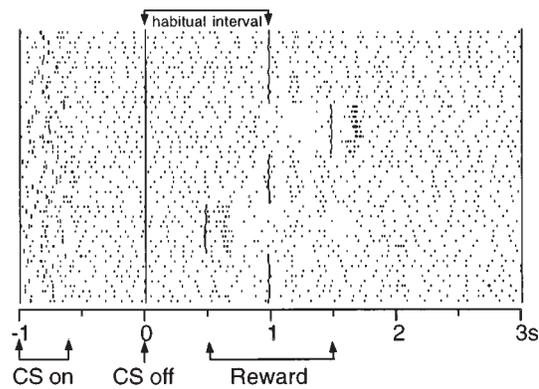

FIGURE 3. The prediction error coded by dopamine neurons includes temporal aspects. In the behavioral task, a well-learned color picture appears on a computer monitor (CS on), the animal presses a lever below the picture, upon which action the CS switches off and a drop of juice reward is delivered shortly afterwards at the animal's mouth. Activity is recorded from a single dopamine neuron while the lever touch-reward interval is being varied. Although the reward occurring at the predicted, habitual time does not elicit a dopamine response (top trial block), a sudden delay of reward by 0.5 s depresses the activity of the dopamine neuron at the original time of reward and induces an activation at the new time (subsequent trial block). An activation is only observed when reward occurs earlier than predicted (second trial block from bottom). Original trial sequence is from *top* to *bottom*. Reprinted from Ref. 4 with permission of Nature America.

ing dopamine neurons concerns both the occurrence and the time of reward.

Attention-inducing stimuli, such as novel or physically intense stimuli not necessarily related to rewards, elicit activations in dopamine neurons that are often followed by depressions. Novelty responses subside together with behavioral orienting reactions after several stimulus repetitions, the duration being longer with physically more salient stimuli. Intense stimuli, such as loud clicks or large pictures immediately in front of an animal, elicit strong responses that still induce measurable activations after >1,000 trials. However, responses to novel or intense stimuli subside rapidly during conditioning of active avoidance behavior. These data suggest that dopamine neurons are not exclusively driven by reward-related stimuli but are also influenced by attention-inducing stimuli.

Taken together, most dopamine neurons show phasic activations after food and fluid rewards, and after conditioned, reward-predicting stimuli. They show biphasic activation-depression responses after stimuli that resemble reward-predicting stimuli or are novel or particularly salient. However, only few a phasic activations follow aversive stimuli. Thus dopamine neurons label environmental stimuli with an appetitive

"…activations occur when animals touch a morsel of hidden food…."

"tag," predict and detect rewards, and signal alerting and motivating events.

All responses to rewards and reward-predicting stimuli depend on event predictability that concerns the precise time of reward. The more tonic reward-predicting environmental context in which a reward occurs does not appear to influence dopamine neurons. The dopamine reward response appears to indicate to what extent a reward occurs differently than predicted, termed an "error" in the prediction of reward. Thus dopamine neurons report rewards relative to their prediction, rather than signaling rewards unconditionally. They appear to be feature detectors for the goodness of environmental events relative to prediction, being activated by rewarding events that are better than predicted, remaining uninfluenced by events that are as good as predicted, and being depressed by events that are worse than predicted (Fig. 2). However, they fail to discriminate between different rewards and thus appear to emit an alerting message about the surprising presence or absence of rewards without indicating the particular nature of each reward. They process the time and prediction of rewards but not the nature of the particular reward.

Potential use of the reward prediction error signal

The moderately bursting, short-duration, nearly synchronous response of the majority of dopamine neurons leads to optimal, simultaneous dopamine release from the majority of closely spaced varicosities in the striatum and frontal cortex. The short puff of dopamine quickly reaches regionally homogeneous concentrations likely to influence the dendrites of probably all striatal and many cortical neurons. In this way, the reward prediction error message in 60-80% of dopamine neurons is broadcast as a divergent, rather global reinforcement signal to the striatum, nucleus accumbens, and frontal cortex, phasically influencing a maximum number of synapses involved in the processing of stimuli and actions leading to reward. The reduction of dopamine release induced by depressions with omitted rewards would reduce the tonic stimulation of dopamine receptors by ambient dopamine.

The basic arrangement of synaptic influences of dopamine neurons on striatal and frontal cortex neurons consists of a triad comprising dendritic spines, excitatory cortical terminals at the tip of dendritic spines, and dopamine varicosities contacting the same dendritic spines (Fig. 4). Every medium-sized striatal spiny neuron receives ~1,000 dopaminergic synapses at its
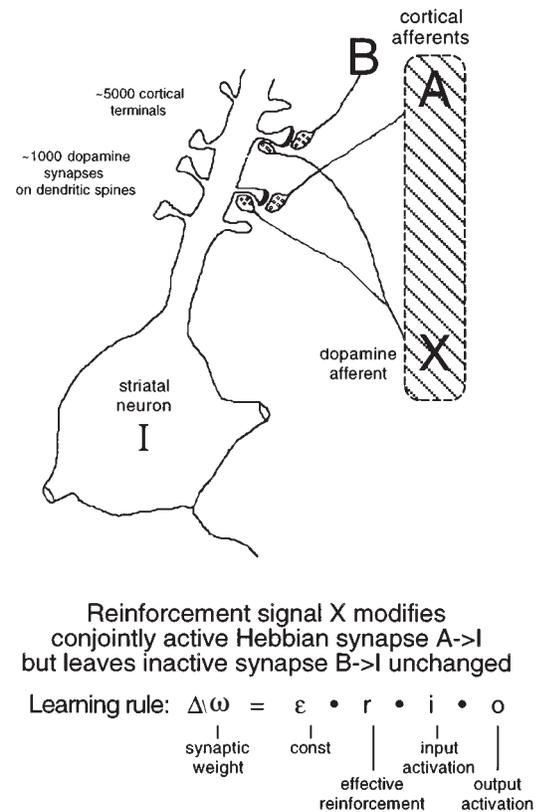
"The released dopamine may act on the striatal and cortical neurons…."

Reinforcement signal X modifies conjointly active Hebbian synapse A->I but leaves inactive synapse B->I unchanged

Learning rule: $\Delta\backslash\omega = \varepsilon \bullet r \bullet i \bullet o$

synaptic weight — const — input activation — output activation

effective reinforcement

FIGURE 4. Basic design of possible influence of dopamine prediction error signal on neurotransmission in the striatum. Synaptic inputs from a single dopamine *axon X* and 2 cortical *axons A* and *B* contact a typical medium-sized spiny striatal *neuron I*. Corticostriatal transmission can be modified by dopamine *input X* contacting indiscriminately the stems of dendritic spines that are also contacted by specific cortical *inputs A* and *B*. In the present example, cortical *input A*, but not *B*, is active at the same time as dopamine *neuron X* (shaded area), e.g., after a reward-related event. This could lead to a modification of the $A \rightarrow I$ transmission, but leave the $B \rightarrow I$ transmission unaltered. Anatomical drawing is based on anatomical data (3) and modified from a previous figure (14).

dendritic spines and ~5,000 cortical synapses. This arrangement would allow dopamine neurons to influence the synaptic effects of cortical inputs to striatal neurons. The released dopamine may act on the striatal and cortical neurons in several possible ways. *1*) The immediate effect may consist in a change of corticostriatal neurotransmission. This would modify information circulating in cortico-basal ganglia loops and influence neurons in cortical structures involved in structuring behavioral output. *2*) The relatively slow time course of dopamine membrane action may leave a short trace of the reward event and influence all subsequent activity for a short while. *3*) The potential dopamine-dependent plasticity in the striatum and the observed forms of dopamine responses may

induce plastic changes in striatal and cortical synapses concurrently activated by the events leading to reward.

In a model of dopamine influences on striatal neurotransmission, A and B are inputs that separately contact dendritic spines of a striatal *neuron I* (Fig. 4). The synaptic weights $A \rightarrow I$ and $B \rightarrow I$ are short-term or long-term Hebbian modifiable. The same spines are indiscriminately contacted by the global reward prediction error signal from dopamine *input X*. Both *neuron X* and *neuron A*, but not *neuron B*, are activated when a reward-related signal is encountered. *Neuron X* transmits the message that a rewarding event has occurred without giving specific details, whereas *neuron A* sends a message about one of several detailed aspects of the reward-related event, such as color, texture, position, surroundings, etc. of the stimulus or may code a movement leading to obtaining the reward. The weights of striatal synapses could be modified according to the learning rule $\Delta\omega = \varepsilon \cdot r \cdot i \cdot o$, where ω is synaptic weight, ε is learning constant, r is dopamine prediction error signal, i is input activation, and o is activation of striatal neuron. Thus, through the simultaneity or near simultaneity of activity in A and X, the activity of *neuron* X may induce a change in neurotransmission at the active $A \rightarrow I$ synapse, but leave the inactive $B \rightarrow I$ neurotransmission unchanged. In the case of lasting changes in synaptic transmission, subsequent input from *neuron A* would lead to an increased response in *neuron I*, whereas input from *neuron B* leads to an unchanged response in *neuron* I. Thus, the synaptic changes of $A \rightarrow I$ and $B \rightarrow I$ neurotransmission are conditional on dopamine *neuron X* being conjointly active with A or B.

The dopamine response coding an error in the prediction of reward resembles in all major aspects the reinforcement signal of a particularly effective class of reinforcement models that incorporate temporal difference algorithms (6, 13, 15). They are based on behavioral learning theories that assume that learning depends crucially on the discrepancy or error between the prediction of reinforcement and its actual occurrence (1, 8). In these models, a critic module generates a global reinforcement signal and sends it to the actor module that learns and executes behavioral output. The critic-actor architecture closely resembles the connectivity of the basal ganglia, including the dopamine projection to the striatum and the reciprocal striatonigral projection. Models using temporal difference algorithms learn a wide variety of behavioral tasks, reaching from balancing a pole on a cart wheel to playing world-class backgammon (for references, see Ref. 11). Robots using temporal difference algorithms learn to move about two-dimensional space and avoid obstacles, reach and grasp, or insert a peg into a hole. Neurobiologically inspired temporal difference models replicate foraging behavior of honeybees, simulate human decision making, and learn orienting reactions, eye movements, sequential movements, and spatial delayed-response tasks. It is particularly interesting to see that teaching signals using prediction errors result in faster and more complete learning, compared with unconditional reinforcement signals.

Conclusions and extensions

The investigation of the activity of dopamine neurons resulted in the surprising finding that these neurons are not modulated in relation to movements, although movements are deficient in parkinsonian patients. Rather, dopamine neurons code in a very special form the rewarding aspects of environmental stimuli, together with certain attention-inducing characteristics. The responses are elicited by primary rewards ("unconditioned stimuli"), conditioned reward-predicting stimuli, stimuli resembling reward-related stimuli, and novel or intense stimuli. However, reward-related stimuli are only reported when they occur differently than predicted, the prediction concerning both the occurrence and the time of the event. The prediction error message is a very powerful signal for directing behavior and inducing learning, according to animal learning theories and reinforcement models. However, the dopamine signal does not specify exactly which reward it is that occurs differently than predicted or whether it is really a reward or, rather, a reward-predicting stimulus. Stimuli resembling rewards and novel or particularly salient stimuli elicit activation-depression sequences that resemble the monophasic activations elicited by unpredicted reward-related stimuli. The dopamine signal thus appears to be a predominantly reward-alerting signal, and other brain systems must process additional information for learning correct behavioral reactions to motivating environmental stimuli.

Information concerning food and fluid rewards is also processed in brain structures other than dopamine neurons, such as dorsal and ventral striatum, subthalamic nucleus, amygdala, dorsolateral prefrontal cortex, orbitofrontal cortex, and anterior cingulate cortex. However, these structures do not appear to emit a global reward prediction error signal similar to dopamine neurons. These structures show 1) transient responses after

"…temporal difference models replicate foraging behavior…."

the delivery of rewards, 2) transient responses to reward-predicting cues, 3) sustained activations during the expectation of rewards, and 4) modulations of behavior-related activity by predicted rewards (for references, see Ref. 11). Many of these neurons differentiate well between different food or fluid rewards. Thus they may process the specific nature of the rewarding event. Some reward responses depend on reward unpredictability in being reduced or absent when the reward is predicted by a conditioned stimulus, although it is unclear whether they signal prediction errors similar to dopamine neurons. It thus appears that the processing of specific rewards for learning and maintaining approach behavior would strongly profit from a cooperation between dopamine neurons signaling the unpredicted occurrence or omission of reward and neurons in the other structures simultaneously indicating the specific nature of the reward.

Impaired dopamine neurotransmission with Parkinson's disease, experimental lesions, or neuroleptic treatment is associated with many behavioral deficits in movement (akinesia, tremor, rigidity), cognition (attention, bradyphrenia, planning, learning), and motivation (reduced emotional responses, depression). Most deficits are considerably ameliorated by systemic dopamine precursor or receptor agonist therapy, which cannot in a simple manner restitute the phasic information transmission by neuronal impulses. It appears that dopamine neurotransmission plays two separate functions in the brain, the phasic processing of appetitive and alerting information and the tonic enabling of a large variety of motor, cognitive, and motivational processes without temporal coding (11). The tonic dopamine function is based on low, sustained extracellular dopamine concentrations in the striatum (5-10 nM) and other dopamine-innervated areas. The ambient dopamine concentration is regulated locally within a narrow range by spontaneous impulses, synaptic overflow, reuptake transport, metabolism, autoreceptor-controlled release and synthesis, and presynaptic transmitter interaction. The tonic stimulation of dopamine receptors should be neither too low nor too high for an optimal function of a given brain region. Other neurotransmitters exist in similarly low ambient concentrations, such as glutamate in striatum, cerebral cortex, hippocampus, and cerebellum, aspartate and GABA in striatum and frontal cortex, and adenosine in hippocampus. Neurons in many brain structures are apparently bathed in a "soup" of neurotransmitters that have powerful, specific physiological effects on neuronal excitability. Given the general importance of tonic extracellular concentrations of neurotransmitters, it appears that the wide range of parkin-

sonian symptoms would not be caused by deficient transmission of reward information by dopamine neurons but would reflect a malfunction of striatal and cortical neurons caused by impaired enabling by reduced ambient dopamine. Dopamine neurons would not be actively involved in the wide range of processes deficient in parkinsonism but would provide the important background concentration of dopamine necessary to maintain proper functioning of striatal and cortical neurons involved in these processes.

References

1. Dickinson, A. *Contemporary Animal Learning Theory*. Cambridge, UK: Cambridge Univ. Press, 1980.
2. Fibiger, H. C., and A. G. Phillips. Reward, motivation, cognition: psychobiology of mesotelencephalic dopamine systems. In: *Handbook of Physiology. The Nervous System. Intrinsic Regulatory Systems of the Brain*. Bethesda, MD: Am. Physiol. Soc., 1986, sect. 1, vol. IV, chapt. 12, p. 647–675.
3. Freund, T. F., J. F. Powell, and A. D. Smith. Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. *Neuroscience* 13: 1189–1215, 1984.
4. Hollerman, J. R., and W. Schultz. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* 1: 304–309, 1998.
5. Ljungberg, T., P. Apicella, and W. Schultz. Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 67: 145–163, 1992.
6. Montague, P. R., P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16: 1936–1947, 1996.
7. Mirenowicz, J., and W. Schultz. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379: 449–451, 1996.
8. Rescorla, R. A., and A. R. Wagner. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: *Classical Conditioning II: Current Research and Theory*, edited by A. H. Black and W. F. Prokasy. New York: Appleton Century Crofts, 1972, p. 64–99.
9. Robbins, T. W., and B. J. Everitt. Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* 6: 228–236, 1996.
10. Romo, R., and W. Schultz. Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* 63: 592–606, 1990.
11. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80: 1–27, 1998.

"Neurons in many brain structures are apparently bathed in a 'soup'…."

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

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

14. Smith, A. D., and J. P. Bolam. The neural network of the basal ganglia as revealed by the study of synaptic connections of identified neurones. *Trends Neurosci.* 13: 259–265, 1990.

15. Sutton, R. S., and A. G. Barto. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* 88: 135–170, 1981.