

BEHAVIORAL THEORIES AND THE NEUROPHYSIOLOGY OF REWARD

Wolfram Schultz

*Department of Anatomy, University of Cambridge, CB2 3DY United Kingdom;
email: ws234@cam.ac.uk*

Key Words learning theory, conditioning, microeconomics, utility theory, uncertainty

■ **Abstract** The functions of rewards are based primarily on their effects on behavior and are less directly governed by the physics and chemistry of input events as in sensory systems. Therefore, the investigation of neural mechanisms underlying reward functions requires behavioral theories that can conceptualize the different effects of rewards on behavior. The scientific investigation of behavioral processes by animal learning theory and economic utility theory has produced a theoretical framework that can help to elucidate the neural correlates for reward functions in learning, goal-directed approach behavior, and decision making under uncertainty. Individual neurons can be studied in the reward systems of the brain, including dopamine neurons, orbitofrontal cortex, and striatum. The neural activity can be related to basic theoretical terms of reward and uncertainty, such as contiguity, contingency, prediction error, magnitude, probability, expected value, and variance.

CONTENTS

INTRODUCTION	88
GENERAL IDEAS ON REWARD FUNCTION, AND A CALL FOR THEORY	89
A Call for Behavioral Theory	91
REWARD FUNCTIONS DEFINED BY ANIMAL LEARNING THEORY	91
Learning	92
Approach Behavior	94
Motivational Valence	94
NEUROPHYSIOLOGY OF REWARD BASED ON ANIMAL LEARNING THEORY	95
Primary Reward	95
Contiguity	95
Contingency	95
Prediction Error	99
Approach Behavior and Goal Directedness	102
REWARD FUNCTIONS DEFINED BY MICROECONOMIC UTILITY THEORY	105

NEUROPHYSIOLOGY OF REWARD BASED ON ECONOMIC THEORY 107

 Magnitude 107

 Probability 108

 Expected Value 109

 Uncertainty 109

CONCLUSIONS 110

INTRODUCTION

How can we understand the common denominator of Pavlov’s salivating dogs, an ale named Hobgoblin, a market in southern France, and the bargaining for lock access on the Mississippi River? Pavlov’s dogs were presented with pieces of delicious sausage that undoubtedly made them salivate. We know that the same animal will salivate also when it hears a bell that has repeatedly sounded a few seconds before the sausage appears, as if the bell induced the well-known, pleasant anticipation of the desired sausage. Changing slightly the scenery, imagine you are in Cambridge, walk down Mill Lane, and unfailingly end up in the Mill pub by the river Cam. The known attraction inducing the pleasant anticipation is a pint of Hobgoblin. Hobgoblin’s provocative ad reads something like “What’s the matter Lager boy, afraid you might taste something?” and refers to a full-bodied, dark ale whose taste alone is a reward. Changing the scenery again, you are in the middle of a Saturday morning market in a small town in southern France and run into a nicely arranged stand of rosé and red wines. Knowing the presumably delicious contents of the differently priced bottles to varying degrees, you need to make a decision about what to get for lunch. You can do a numerical calculation and weigh the price of each bottle by the probability that its contents will please your taste, but chances are that a more automatic decision mechanism kicks in that is based on anticipation and will tell you quite quickly what to choose. However, you cannot use the same simple emotional judgment when you are in the shoes of an economist trying to optimize the access to the locks on the Mississippi River. The task is to find a pricing structure that assures the most efficient and uninterrupted use of the infrastructure over a 24-hour day, by avoiding long queues during prime daytime hours and inactive periods during the wee hours of the night. A proper pricing structure known in advance to the captains of the barges will shape their decisions to enter the locks at a moment that is economically most appropriate for the whole journey. The common denominator in these tasks appears to relate to the anticipation of outcomes of behavior in situations with varying degrees of uncertainty: the merely automatic salivation of a dog without much alternative, the choice of sophisticated but partly unknown liquids, or the well-calculated decision of a barge captain on how to get the most out of his money and time.

The performance in these tasks is managed by the brain, which assesses the values and uncertainties of predictable outcomes (sausage, ale, wine, lock pricing, and access to resources) and directs the individuals’ decisions toward the current

optimum. This review describes some of the knowledge on brain mechanisms related to rewarding outcomes, without attempting to provide a complete account of all the studies done. We focus on the activity of single neurons studied by neurophysiological techniques in behaving animals, in particular monkeys, and emphasize the formative role of behavioral theories, such as animal learning theory and microeconomic utility theory, on the understanding of these brain mechanisms. Given the space limits and the only just beginning neurophysiological studies based on game theory (Barracough et al. 2004, Dorris & Glimcher 2004), the description of the neurophysiology of this promising field will have to wait until more data have been gathered. The review will not describe the neurobiology of artificial drug rewards, which constitutes a field of its own but does not require vastly different theoretical backgrounds of reward function for its understanding. Readers interested in the rapidly emerging and increasingly large field of human neuroimaging of reward and reward-directed decision making are referred to other reviews (O'Doherty 2004).

GENERAL IDEAS ON REWARD FUNCTION, AND A CALL FOR THEORY

Homer's Odysseus proclaims, "Whatever my distress may be, I would ask you now to let me eat. There is nothing more devoid of shame than the accursed belly; it thrusts itself upon a man's mind in spite of his afflictions. . . my heart is sad but my belly keeps urging me to have food and drink. . . it says imperiously: 'eat and be filled'." (*The Odyssey*, Book VII, 800 BC). Despite these suggestive words, Homer's description hardly fits the common-sensical perceptions of reward, which largely belong to one of two categories. People often consider a reward as a particular object or event that one receives for having done something well. You succeed in an endeavor, and you receive your reward. This reward function could be most easily accommodated within the framework of instrumental conditioning, according to which the reward serves as a positive reinforcer of a behavioral act. The second common perception of reward relates to subjective feelings of liking and pleasure. You do something again because it produced a pleasant outcome before. We refer to this as the hedonic function of rewards. The following descriptions will show that both of these perceptions of reward fall well short of providing a complete and coherent description of reward functions.

One of the earliest scientifically driven definitions of reward function comes from Pavlov (1927), who defined it as an object that produces a change in behavior, also called learning. The dog salivates to a bell only after the sound has been paired with a sausage, but not to a different, nonpaired sound, suggesting that its behavioral response (salivation) has changed after food conditioning. It is noteworthy that this definition bypasses both common-sensical reward notions, as the dog does not need to do anything in particular for the reward to occur (notion 1) nor is it

relevant what the dog feels (notion 2). Yet we will see that this definition is a key to neurobiological studies.

Around this time, Thorndike's (1911) Law of Effect postulated that a reward increases the frequency and intensity of a specific behavioral act that has resulted in a reward before or, as a common interpretation has it, "rewards make you come back for more." This definition comes close to the idea of instrumental conditioning, in that you get a reward for having done something well, and not automatically as with Pavlovian conditioning. It resembles Pavlov's definition of learning function, as it suggests that you will do more of the same behavior that has led previously to the rewarding outcome (positive reinforcement). Skinner pushed the definition of instrumental, or operant, conditioning further by defining rewards as reinforcers of stimulus-response links that do not require mental processes such as intention, representation of goal, or consciousness. Although the explicit antimental stance reduced the impact of his concept, the purely behaviorist approach to studying reward function allowed scientists to acquire a huge body of knowledge by studying the behavior of animals, and it paved the way to neurobiological investigations without the confounds of subjective feelings.

Reward objects for animals are primarily vegetative in nature, such as different foodstuffs and liquids with various tastes. These rewards are necessary for survival, their motivational value can be determined by controlled access, and they can be delivered in quantifiable amounts in laboratory situations. The other main vegetative reward, sex, is impossible to deliver in neurophysiological laboratory situations requiring hundreds of daily trials. Animals are also sensitive to other, nonvegetative rewards, such as touch to the skin or fur and presentation of novel objects and situations eliciting exploratory responses, but these again are difficult to parameterize for laboratory situations. Humans use a wide range of nonvegetative rewards, such as money, challenge, acclaim, visual and acoustic beauty, power, security, and many others, but these are not considered as this review considers neural mechanisms in animals.

An issue with vegetative rewards is the precise definition of the rewarding effect. Is it the seeing of an apple, its taste on the tongue, the swallowing of a bite of it, the feeling of its going down the throat, or the rise in blood sugar subsequent to its digestion that makes it a reward and has one come back for more? Which of these events constitutes the primary rewarding effect, and do different objects draw their rewarding effects from different events (Wise 2002)? In some cases, the reward may be the taste experienced when an object activates the gustatory receptors, as with saccharin, which has no nutritional effects but increases behavioral reactions. The ultimate rewarding effect of many nutrient objects may be the specific influence on vegetative parameters, such as electrolyte, glucose, and amino acid concentrations in plasma and brain. This would explain why animals avoid foods that lack such nutrients as essential amino acids (Delaney & Gelperin 1986, Hrupka et al. 1997, Rogers & Harper 1970, Wang et al. 1996). The behavioral function of some reward objects may be determined by innate mechanisms, whereas a much larger variety might be learned through experience.

Although these theories provide important insights into reward function, they tend to neglect the fact that individuals usually operate in a world with limited nutritional and mating resources, and that most resources occur with different degrees of uncertainty. The animal in the wild is not certain whether it will encounter a particular fruit or prey object at a particular moment, nor is the restaurant goer certain that her preferred chef will cook that night. To make the uncertainty of outcomes tractable was the main motive that led Blaise Pascal to develop probability theory around 1650 (see Glimcher 2003 for details). He soon realized that humans make decisions by weighing the potential outcomes by their associated probabilities and then go for the largest result. Or, mathematically speaking, they sum the products of magnitude and probability of all potential outcomes of each option and then choose the option with the highest expected value. Nearly one hundred years later, Bernoulli (1738) discovered that the utility of outcomes for decision making does not increase linearly but frequently follows a concave function, which marks the beginning of microeconomic decision theory. The theory provides quantifiable assessments of outcomes under uncertainty and has gone a long way to explain human and animal decision making, even though more recent data cast doubt on the logic in some decision situations (Kahneman & Tversky 1984).

A Call for Behavioral Theory

Primary sensory systems have dedicated physical and chemical receptors that translate environmental energy and information into neural language. Thus, the functions of primary sensory systems are governed by the laws of mechanics, optics, acoustics, and receptor binding. By contrast, there are no dedicated receptors for reward, and the information enters the brain through mechanical, gustatory, visual, and auditory receptors of the sensory systems. The functions of rewards cannot be derived entirely from the physics and chemistry of input events but are based primarily on behavioral effects, and the investigation of reward functions requires behavioral theories that can conceptualize the different effects of rewards on behavior. Thus, the exploration of neural reward mechanisms should not be based primarily on the physics and chemistry of reward objects but on specific behavioral theories that define reward functions. Animal learning theory and microeconomics are two prominent examples of such behavioral theories and constitute the basis for this review.

REWARD FUNCTIONS DEFINED BY ANIMAL LEARNING THEORY

This section will combine some of the central tenets of animal learning theories in an attempt to define a coherent framework for the investigation of neural reward mechanisms. The framework is based on the description of observable behavior and superficially resembles the behaviorist approach, although mental states

of representation and prediction are essential. Dropping the issues of subjective feelings of pleasure will allow us to do objective behavioral measurements in controlled neurophysiological experiments on animals. To induce subjective feelings of pleasure and positive emotion is a key function of rewards, although it is unclear whether the pleasure itself has a reinforcing, causal effect for behavior (i.e., I feel good because of the outcome I got and therefore will do again what produced the pleasant outcome) or is simply an epiphenomenon (i.e., my behavior gets reinforced and, in addition, I feel good because of the outcome).

Learning

Rewards induce changes in observable behavior and serve as positive reinforcers by increasing the frequency of the behavior that results in reward. In Pavlovian, or classical, conditioning, the outcome follows the conditioned stimulus (CS) irrespective of any behavioral reaction, and repeated pairing of stimuli with outcomes leads to a representation of the outcome that is evoked by the stimulus and elicits the behavioral reaction (Figure 1a). By contrast, instrumental, or operant, conditioning requires the subject to execute a behavioral response; without such response there will be no reward. Instrumental conditioning increases the frequency of those behaviors that are followed by reward by reinforcing stimulus-response links. Instrumental conditioning allows subjects to influence their environment and determine their rate of reward.

The behavioral reactions studied classically by Pavlov are vegetative responses governed by smooth muscle contraction and gland discharge, whereas more recent Pavlovian tasks also involve reactions of striated muscles. In the latter case, the final reward usually needs to be collected by an instrumental contraction of striated muscle, but the behavioral reaction to the CS itself, for example, anticipatory licking, is not required for the reward to occur and thus is classically conditioned. As a further emphasis on Pavlovian mechanisms, the individual stimuli in instrumental tasks that predict rewards are considered to be Pavlovian conditioned. These distinctions are helpful when trying to understand why the neural mechanisms of reward prediction reveal strong influences of Pavlovian conditioning.

Three factors govern conditioning, namely contiguity, contingency, and prediction error. Contiguity refers to the requirement of near simultaneity (Figure 1a). Specifically, a reward needs to follow a CS or response by an optimal interval of a few seconds, whereas rewards occurring before a stimulus or response do not contribute to learning (backward conditioning). The contingency requirement postulates that a reward needs to occur more frequently in the presence of a stimulus as compared with its absence in order to induce “excitatory” conditioning of the stimulus (Figure 1b); the occurrence of the CS predicts a higher incidence of reward compared with no stimulus, and the stimulus becomes a reward predictor. By contrast, if a reward occurs less frequently in the absence of a stimulus, compared with its presence, the occurrence of the stimulus predicts a lower incidence of reward, and the stimulus becomes a conditioned inhibitor, even though the contiguity

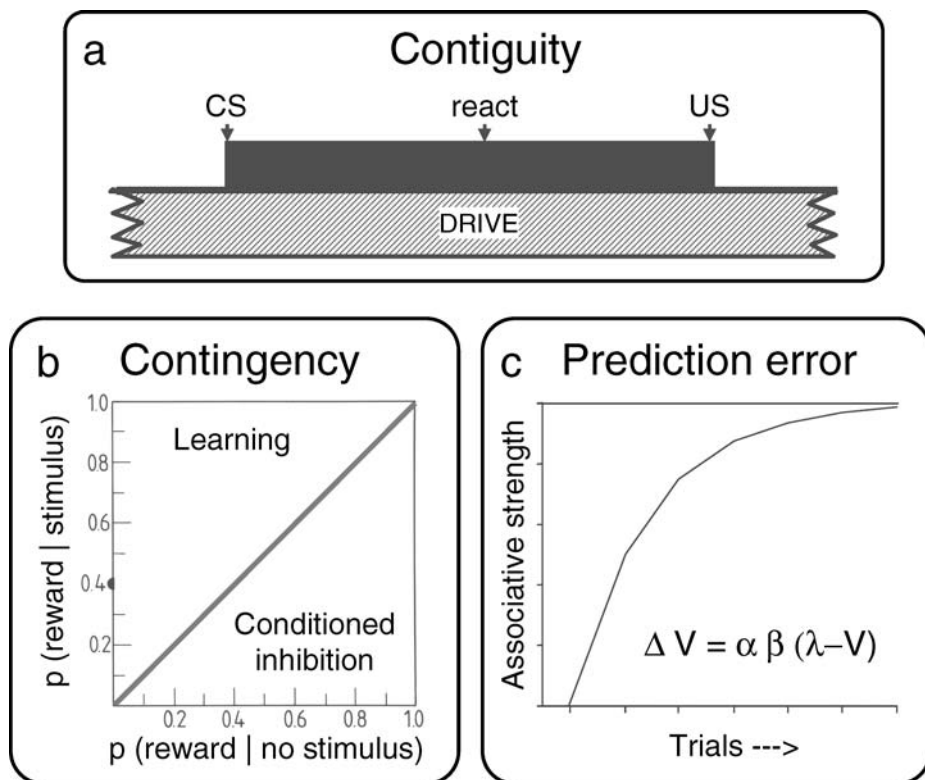


Figure 1 Basic assumptions of animal learning theory defining the behavioral functions of rewards. (a) Contiguity refers to the temporal proximity of a conditioned stimulus (CS), or action, and the reward. (b) Contingency refers to the conditional probability of reward occurring in the presence of a conditioned stimulus as opposed to its absence (modified from Dickinson 1980). (c) Prediction error denotes the discrepancy between an actually received reward and its prediction. Learning (ΔV , associative strength) is proportional to the prediction error ($\lambda - V$) and reaches its asymptote when the prediction error approaches zero after several learning trials. All three requirements need to be fulfilled for learning to occur. US, unconditioned stimulus.

requirement is fulfilled. The crucial role of prediction error is derived from Kamin's (1969) blocking effect, which postulates that a reward that is fully predicted does not contribute to learning, even when it occurs in a contiguous and contingent manner. This is conceptualized in the associative learning rules (Rescorla & Wagner 1972), according to which learning advances only to the extent to which a reinforcer is unpredicted and slows progressively as the reinforcer becomes more predicted (Figure 1c). The omission of a predicted reinforcer reduces the strength of the CS and produces extinction of behavior. So-called attentional learning rules in addition

relate the capacity to learn (associability) in certain situations to the degree of attention evoked by the CS or reward (Mackintosh 1975, Pearce & Hall 1980).

Approach Behavior

Rewards elicit two forms of behavioral reactions, approach and consumption. This is because the objects are labeled with appetitive value through innate mechanisms (primary rewards) or, in most cases, classical or instrumental conditioning, after which these objects constitute, strictly speaking, conditioned reinforcers (Wise 2002). Nutritional rewards can derive their value from hunger and thirst states, and satiation of the animal reduces the reward value and consequently the behavioral reactions.

Conditioned, reward-predicting stimuli also induce preparatory or approach behavior toward the reward. In Pavlovian conditioning, subjects automatically show nonconsummatory behavioral reactions that would otherwise occur after the primary reward and that increase the chance of consuming the reward, as if a part of the behavioral response has been transferred from the primary reward to the CS (Pavlovian response transfer).

In instrumental conditioning, a reward can become a goal for instrumental behavior if two conditions are met. The goal needs to be represented at the time the behavior is being prepared and executed. This representation should contain a prediction of the future reward together with the contingency that associates the behavioral action to the reward (Dickinson & Balleine 1994). Behavioral tests for the role of “incentive” reward-predicting mechanisms include assessing behavioral performance in extinction following devaluation of the reward by satiation or aversive conditioning in the absence of the opportunity to perform the instrumental action (Balleine & Dickinson 1998). A reduction of behavior in this situation indicates that subjects have established an internal representation of the reward that is updated when the reward changes its value. (Performing the action together with the devalued outcome would result in reduced behavior due to partial extinction, as the reduced reward value would diminish the strength of the association.) To test the role of action-reward contingencies, the frequency of “free” rewards in the absence of the action can be varied to change the strength of association between the action and the reward and thereby modulate instrumental behavior (Balleine & Dickinson 1998).

Motivational Valence

Punishers have opposite valence to rewards, induce withdrawal behavior, and act as negative reinforcers by increasing the behavior that results in decreasing the aversive outcome. Avoidance can be passive when subjects increasingly refrain from doing something that is associated with a punisher (don’t do it); active avoidance involves increasing an instrumental response that is likely to reduce the impact of a punisher (get away from it). Punishers induce negative emotional states of anger, fear, and panic.

NEUROPHYSIOLOGY OF REWARD BASED ON ANIMAL LEARNING THEORY

Primary Reward

Neurons responding to liquid or food rewards are found in a number of brain structures, such as orbitofrontal, premotor and prefrontal cortex, striatum, amygdala, and dopamine neurons (Amador et al. 2000, Apicella et al. 1991, Bowman et al. 1996, Hikosaka et al. 1989, Ljungberg et al. 1992, Markowitsch & Pritzel 1976, Nakamura et al. 1992, Nishijo et al. 1988, Pratt & Mizumori 1998, Ravel et al. 1999, Shidara et al. 1998, Thorpe et al. 1983, Tremblay & Schultz 1999). Satiation of the animal reduces the reward responses in orbitofrontal cortex (Critchley & Rolls 1996) and in the secondary gustatory area of caudal orbitofrontal cortex (Rolls et al. 1989), a finding that suggests that the responses reflect the rewarding functions of the objects and not their taste. Taste responses are found in the primary gustatory area of the insula and frontal operculum and are insensitive to satiation (Rolls et al. 1988).

Contiguity

Procedures involving Pavlovian conditioning provide simple paradigms for learning and allow the experimenter to test the basic requirements of contiguity, contingency, and prediction error. Contiguity can be tested by presenting a reward 1.5–2.0 seconds after an untrained, arbitrary visual or auditory stimulus for several trials. A dopamine neuron that responds initially to a liquid or food reward acquires a response to the CS after some tens of paired CS-reward trials (Figure 2) (Mireniewicz & Schultz 1994, Waelti 2000). Responses to conditioned, reward-predicting stimuli occur in all known reward structures of the brain, including the orbitofrontal cortex, striatum, and amygdala (e.g., Hassani et al. 2001, Liu & Richmond 2000, Nishijo et al. 1988, Rolls et al. 1996, Thorpe et al. 1983, Tremblay & Schultz 1999). (Figure 2 shows that the response to the reward itself disappears in dopamine neurons, but this is not a general phenomenon with other neurons.)

Contingency

The contingency requirement postulates that in order to be involved in reward prediction, neurons should discriminate between three kinds of stimuli, namely reward-predicting CSs (conditioned exciters), after which reward occurs more frequently compared with no CS (Figure 1*b*, top left); conditioned inhibitors, after which reward occurs less frequently compared with no CS (Figure 1*b*, bottom right); and neutral stimuli that are not associated with changes in reward frequency compared with no stimulus (diagonal line in Figure 1*b*). In agreement with these postulates, dopamine neurons are activated by reward-predicting CSs, show depressions of activity following conditioned inhibitors, which may be accompanied

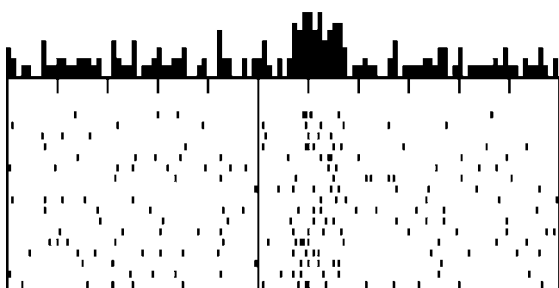
by small activations, and hardly respond to neutral stimuli when response generalization is excluded (Figure 3) (Tobler et al. 2003). The conditioned inhibitor in these experiments is set up by pairing the inhibitor with a reward-predicting CS while withholding the reward, which amounts to a lower probability of reward in the presence of the inhibitor compared with its absence (reward-predicting stimulus alone) and thus follows the scheme of Figure 1*b* (bottom right). Without conditioned inhibitors being tested, many studies find CS responses that distinguish between reward-predicting and neutral stimuli in all reward structures (e.g., Aosaki et al. 1994, Hollerman et al. 1998, Kawagoe et al. 1998, Kimura et al. 1984, Nishijo et al. 1988, Ravel et al. 1999, Shidara et al. 1998, Waelti et al. 2001).

Further tests assess the specificity of information contained in CS responses. In the typical behavioral tasks used in monkey experiments, the CS may contain several different stimulus components, namely spatial position; visual object features such as color, form, and spatial frequency; and motivational features such as reward prediction. It would be necessary to establish through behavioral testing which of these features is particularly effective in evoking a neural response. For example, neurons in the orbitofrontal cortex discriminate between different CSs on the basis of their prediction of different food and liquid rewards (Figure 4) (Critchley & Rolls 1996, Tremblay & Schultz 1999). By contrast, these neurons are less sensitive to the visual object features of the same CSs, and they rarely code their spatial position, although neurons in other parts of frontal cortex are particularly tuned to these nonreward parameters (Rao et al. 1997). CS responses that are primarily sensitive to the reward features are found also in the amygdala (Nishijo et al. 1988) and striatum (Hassani et al. 2001). These data suggest that individual neurons in these structures can extract the reward components from the multidimensional stimuli used in these experiments as well as in everyday life.

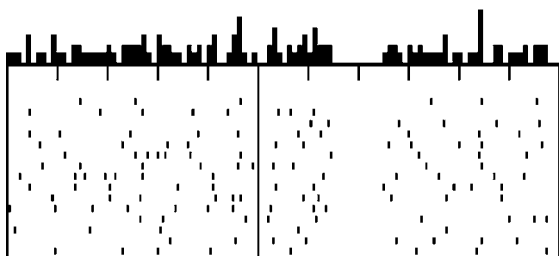
Reward neurons should distinguish rewards from punishers. Different neurons in orbitofrontal cortex respond to rewarding and aversive liquids (Thorpe et al. 1983). Dopamine neurons are activated preferentially by rewards and reward-predicting stimuli but are only rarely activated by aversive air puffs and saline (Mirenowicz & Schultz 1996). In anesthetized animals, dopamine neurons show depressions following painful stimuli (Schultz & Romo 1987, Ungless et al. 2004). Nucleus accumbens neurons in rats show differential activating or depressing responses to CSs predicting rewarding sucrose versus aversive quinine solutions in a Pavlovian task (Roitman et al. 2005). By contrast, the group of tonically active neurons of the striatum responds to both rewards and aversive air puffs, but not to neutral stimuli (Ravel et al. 1999). They seem to be sensitive to reinforcers in general, without specifying their valence. Alternatively, their responses might reflect the higher attention-inducing effects of reinforcers compared with neutral stimuli.

The omission of reward following a CS moves the contingency toward the diagonal line in Figure 1*b* and leads to extinction of learned behavior. By analogy, the withholding of reward reduces the activation of dopamine neurons by CSs within several tens of trials (Figure 5) (Tobler et al. 2003).

Conditioned stimulus
predicting reward



Conditioned stimulus
predicting absence of reward



Known neutral stimulus

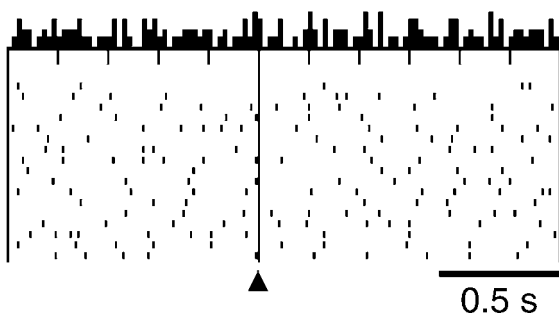
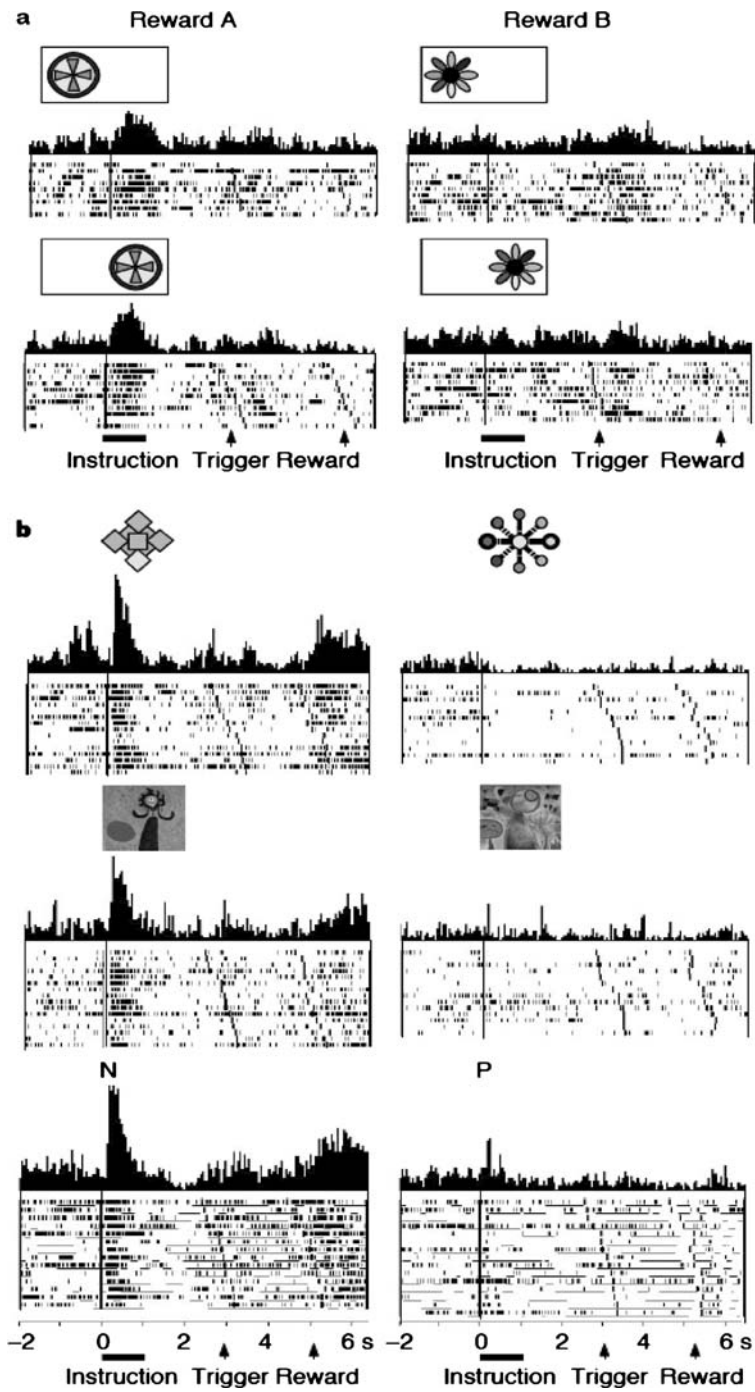


Figure 3 Testing the contingency requirement for associative learning: responses of a single dopamine neuron to three types of stimuli. (*Top*) Activating response to a reward-predicting stimulus (higher occurrence of reward in the presence as opposed to absence of stimulus). (*Middle*) Depressant response to a different stimulus predicting the absence of reward (lower occurrence of reward in the presence as opposed to absence of stimulus). (*Bottom*) Neutral stimulus (no change in reward occurrence after stimulus). Vertical line and arrow indicate time of stimulus.



Prediction Error

Just as with behavioral learning, the acquisition of neuronal responses to reward-predicting CSs should depend on prediction errors. In the prediction error–defining blocking paradigm, dopamine neurons acquire a response to a CS only when the CS is associated with an unpredicted reward, but not when the CS is paired with a reward that is already predicted by another CS and the occurrence of the reward does not generate a prediction error (Figure 6) (Waelti et al. 2001). The neurons fail to learn to respond to reward predictors despite the fact that contiguity and contingency requirements for excitatory learning are fulfilled. These data demonstrate the crucial importance of prediction errors for associative neural learning and suggest that learning at the single-neuron level may follow similar rules as those for behavioral learning. This suggests that some behavioral learning functions may be carried by populations of single neurons.

Neurons may not only be sensitive to prediction errors during learning, but they may also emit a prediction error signal. Dopamine neurons, and some neurons in orbitofrontal cortex, show reward activations only when the reward occurs unpredictably and fail to respond to well-predicted rewards, and their activity is depressed when the predicted reward fails to occur (Figure 7) (Mirenowicz & Schultz 1994, Tremblay & Schultz 2000a). This result has prompted the notion that dopamine neurons emit a positive signal (activation) when an appetitive event is better than predicted, no signal (no change in activity) when an appetitive event occurs as predicted, and a negative signal (decreased activity) when an appetitive event is worse than predicted (Schultz et al. 1997). In contrast to this bidirectional error signal, some neurons in the prefrontal, anterior, and posterior cingulate cortex show a unidirectional error signal upon activation when a reward fails to occur because of a behavioral error of the animal (Ito et al. 2003, McCoy et al. 2003, Watanabe 1989; for review of neural prediction errors, see Schultz & Dickinson 2000).

More stringent tests for the neural coding of prediction errors include formal paradigms of animal learning theory in which prediction errors occur in specific situations. In the blocking paradigm, the blocked CS does not predict a reward. Accordingly, the absence of a reward following that stimulus does not produce a prediction error nor a response in dopamine neurons, and the delivery of a reward does produce a positive prediction error and a dopamine response (Figure 8a; left) (Waelti et al. 2001). By contrast, after a well-trained,

←
Figure 4 Reward discrimination in orbitofrontal cortex. (a) A neuron responding to the instruction cue predicting grenadine juice (*left*) but not apple juice (*right*), irrespective of the left or right position of the cue in front of the animal. (b) A different neuron responding to the cue predicting grape juice (*left*) but not orange juice (*right*), irrespective of the picture object predicting the juice. From Tremblay & Schultz 1999, © Nature MacMillan Publishers.

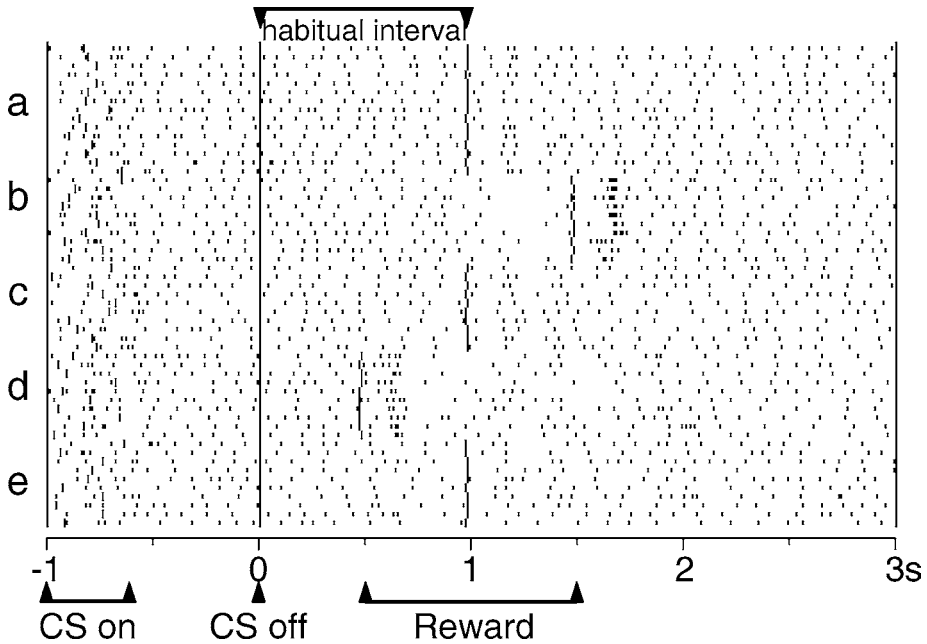


Figure 7 Dopamine response codes temporal reward prediction error. (a, c, e) No response to reward delivered at habitual time. (b) Delay in reward induces depression at previous time of reward, and activation at new reward time. (d) Precocious reward delivery induces activation at new reward time, but no depression at previous reward time. Trial sequence is from top to bottom. Data from Hollerman & Schultz (1998). CS, conditioned stimulus.

reward-predicting CS, reward omission produces a negative prediction error and a depressant neural response, and reward delivery does not lead to a prediction error or a response in the same dopamine neuron (Figure 8a; right). In a conditioned inhibition paradigm, the conditioned inhibitor predicts the absence of reward, and the absence of reward after this stimulus does not produce a prediction error or a response in dopamine neurons, even when another, otherwise reward-predicting stimulus is added (Figure 8b) (Tobler et al. 2003). By contrast, the occurrence of reward after an inhibitor produces an enhanced prediction error, as the prediction error represents the difference between the actual reward and the negative prediction from the inhibitor, and the dopamine neuron shows a strong response (Figure 8b; bottom). Taken together, these data suggest that dopamine neurons show bidirectional coding of reward prediction errors, following the equation

$$\text{Dopamine response} = \text{Reward occurred} - \text{Reward predicted.}$$

This equation may constitute a neural equivalent for the prediction error term of ($\lambda - V$) of the Rescorla-Wagner learning rule. With these characteristics, the

bidirectional dopamine error response would constitute an ideal teaching signal for neural plasticity.

The neural prediction error signal provides an additional means to investigate the kinds of information contained in the representations evoked by CSs. Time apparently plays a major role in behavioral learning, as demonstrated by the unblocking effects of temporal variations of reinforcement (Dickinson et al. 1976). Figure 7 shows that the prediction acting on dopamine neurons concerns the exact time of reward occurrence. Temporal deviations induce a depression when the reward fails to occur at the predicted time (time-sensitive reward omission response), and an activation when the reward occurs at a moment other than predicted (Hollerman & Schultz 1998). This time sensitivity also explains why neural prediction errors occur at all in the laboratory in which animals know that they will receive ample quantities of reward but without knowing when exactly the reward will occur. Another form of time representation is revealed by tests in which the probability of receiving a reward after the last reward increases over consecutive trials. Thus, the animal's reward prediction should increase after each unrewarded trial, the positive prediction error with reward should decrease, and the negative prediction error with reward omission should increase. In line with this reasoning, dopamine neurons show progressively decreasing activations to reward delivery as the number of trials since the last reward increases, and increasing depressions in unrewarded trials (Figure 9) (Nakahara et al. 2004). The result suggests that, for the neurons, the reward prediction in the CS increases after every unrewarded trial, due to the temporal profile of the task evoked by the CS, and contradicts an assumption from temporal difference reinforcement modeling that the prediction error of the preceding unrewarded trial would reduce the current reward prediction in the CS, in which case the neural prediction error responses should increase, which is the opposite to what is actually observed (although the authors attribute the temporal conditioning to the context and have the CS conform to the temporal difference model). The results from the two experiments demonstrate that dopamine neurons are sensitive to different aspects of temporal information evoked by reward-predicting CSs and demonstrate how experiments based on specific behavioral concepts, namely prediction error, reveal important characteristics of neural coding.

The uncertainty of reward is a major factor for generating the attention that determines learning according to the associability learning rules (Mackintosh 1975, Pearce & Hall 1980). When varying the probability of reward in individual trials from 0 to 1, reward becomes most uncertain at $p = 0.5$, as it is most unclear whether or not a reward will occur. (Common perception might say that reward is even more uncertain at $p = 0.25$; however, at this low probability, it is nearly certain that reward will not occur.) Dopamine neurons show a slowly increasing activation between the CS and reward that is maximal at $p = 0.5$ (Fiorillo et al. 2003). This response may constitute an explicit uncertainty signal and is different in time and occurrence from the prediction error response. The response might contribute to a teaching signal in situations defined by the associability learning rules.

Approach Behavior and Goal Directedness

Many behavioral tasks in the laboratory involve more than a CS and a reward and comprise instrumental ocular or skeletal reactions, mnemonic delays between instruction cues and behavioral reactions, and delays between behavioral reactions and rewards during which animals can expect the reward.

Appropriately conditioned stimuli can evoke specific expectations of reward, and phasic neural responses to these CSs may reflect the process of evocation (see above). Once the representations have been evoked, their content can influence the behavior during some time. Neurons in a number of brain structures show sustained activations after an initial CS has occurred. The activations arise usually during specific epochs of well-differentiated instrumental tasks, such as during movement preparation (Figure 10*a*) and immediately preceding the reward (Figure 10*b*), whereas few activations last during the entire period between CS and reward. The activations differentiate between reward and no reward, between different kinds of liquid and food reward, and between different magnitudes of reward. They occur in all trial types in which reward is expected, irrespective of the type of behavioral action (Figure 10). Thus, the activations appear to represent reward expectations. They are found in the striatum (caudate, putamen, ventral

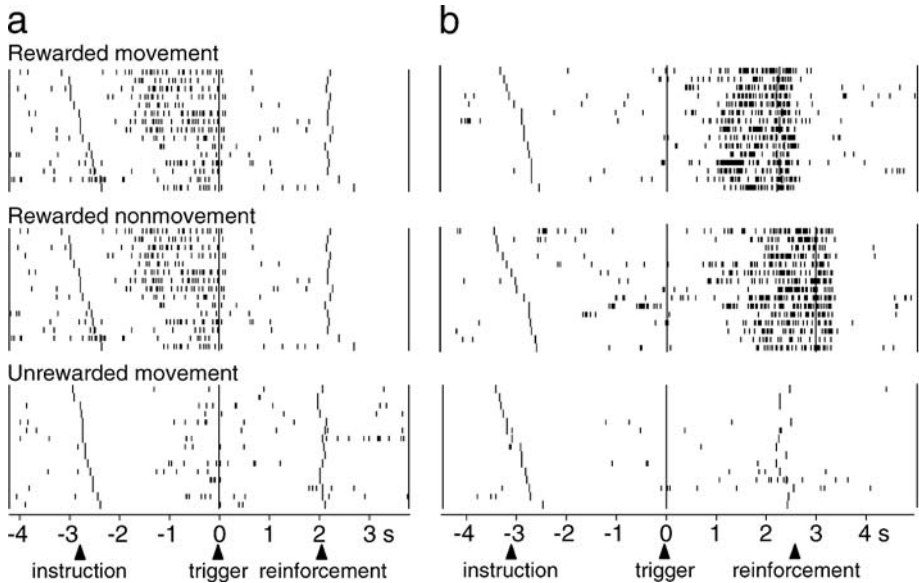


Figure 10 Reward expectation in the striatum. (*a*) Activation in a caudate neuron preceding the stimulus that triggers the movement or nonmovement reaction in both rewarded trial types irrespective of movement, but not in unrewarded movement trials. (*b*) Activation in a putamen neuron preceding the delivery of liquid reward in both rewarded trial types, but not before the reinforcing sound in unrewarded movement trials. Data from Hollerman et al. (1998).

striatum), amygdala, orbitofrontal cortex, dorsolateral prefrontal cortex, anterior cingulate, and supplementary eye field (Amador et al. 2000, Apicella et al. 1992, Cromwell & Schultz 2003, Hikosaka et al. 1989, Hollerman et al. 1998, Pratt & Mizumori 2001, Schoenbaum et al. 1998, Schultz et al. 1992, Shidara & Richmond 2002, Tremblay & Schultz 1999, 2000a, Watanabe 1996, Watanabe et al. 2002). Reward expectation-related activity in orbitofrontal cortex and amygdala develops as the reward becomes predictable during learning (Schoenbaum et al. 1999). In learning episodes with pre-existing reward expectations, orbitofrontal and striatal activations occur initially in all situations but adapt to the currently valid expectations, for example when novel stimuli come to indicate rewarded versus unrewarded trials. The neural changes occur in parallel with the animal's behavioral differentiation (Tremblay et al. 1998, Tremblay & Schultz 2000b).

In some neurons, the differential reward expectation-related activity discriminates in addition between different behavioral responses, such as eye and limb movements toward different spatial targets and movement versus nonmovement reactions (Figure 11). Such neurons are found in the dorsolateral prefrontal cortex (Kobayashi et al. 2002, Matsumoto et al. 2003, Watanabe 1996) and striatum

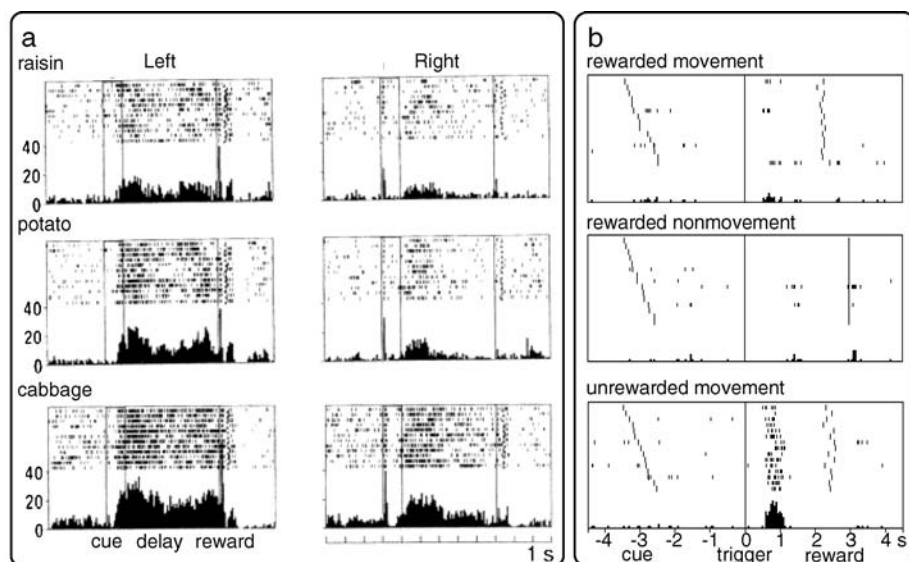


Figure 11 Potential neural mechanisms underlying goal-directed behavior. (a) Delay activity of a neuron in primate prefrontal cortex that encodes, while the movement is being prepared, both the behavioral reaction (left versus right targets) and the kind of outcome obtained for performing the action. From Watanabe (1996), © Nature MacMillan Publishers. (b) Response of a caudate neuron to the movement-triggering stimulus exclusively in unrewarded trials, thus coding both the behavioral reaction being executed and the anticipated outcome of the reaction. Data from Hollerman et al. (1998).

(Cromwell & Schultz 2003, Hassani et al. 2001, Hollerman et al. 1998, Kawagoe et al. 1998). The activations occur during task epochs related to the preparation and execution of the movement that is performed in order to obtain the reward. They do not simply represent outcome expectation, as they differentiate between different behavioral reactions despite the same outcome (Figure 11*a*, left versus right; Figure 11*b*, movement versus nonmovement), and they do not simply reflect different behavioral reactions, as they differentiate between the expected outcomes (Figure 11*a,b*, top versus bottom). Or, expressed in another way, the neurons show differential, behavior-related activations that depend on the outcome of the trial, namely reward or no reward and different kinds and magnitudes of reward. The differential nature of the activations develops during learning while the different reward expectations are being acquired, similar to simple reward expectation-related activity (Tremblay et al. 1998).

It is known that rewards have strong attention-inducing functions, and reward-related activity in parietal association cortex might simply reflect the known involvement of these areas in attention (Maunsell 2004). It is often tedious to disentangle attention from reward, but one viable solution would be to test neurons for specificity for reinforcers with opposing valence while keeping the levels of reinforcement strength similar for rewards and punishers. The results of such tests suggest that dopamine neurons and some neurons in orbitofrontal cortex discriminate between rewards and aversive events and thus report reward-related but not attention-related stimulus components (Mireniewicz & Schultz 1996, Thorpe et al. 1983). Also, neurons showing increasing activations with decreasing reward value or magnitude are unlikely to reflect the attention associated with stronger rewards. Such inversely related neurons exist in the striatum and orbitofrontal cortex (Hassani et al. 2001, Hollerman et al. 1998, Kawagoe et al. 1998, Watanabe 1996).

General learning theory suggests that Pavlovian associations of reward-predicting stimuli in instrumental tasks relate either to explicit CSs or to contexts. The neural correlates of behavioral associations with explicit stimuli may not only involve the phasic responses to CSs described above but also activations at other task epochs. Further neural correlates of Pavlovian conditioning may consist of the sustained activations that occur during the different task periods preceding movements or rewards (Figure 10), which are only sensitive to reward parameters and not to the types of behavioral reactions necessary to obtain the rewards.

Theories of goal-directed instrumental behavior postulate that in order to consider rewards as goals of behavior, there should be (*a*) an expectation of the outcome at the time of the behavior that leads to the reward, and (*b*) a representation of the contingency between the instrumental action and the outcome (Dickinson & Balleine 1994). The sustained, reward-discriminating activations may constitute a neural mechanism for simple reward expectation, as they reflect the expected reward without differentiating between behavioral reactions (Figure 10). However, these activations are not fully sufficient correlates for goal-directed behavior, as the reward expectation is not necessarily related to the specific action that results in the goal being attained; rather, it might refer to an unrelated reward

that occurs in parallel and irrespective of the action. Such a reward would not constitute a goal of the action, and the reward-expecting activation might simply reflect the upcoming reward without being involved in any goal mechanism. By contrast, reward-expecting activations might fulfill the second, more stringent criterion if they are also specific for the action necessary to obtain the reward. These reward-expecting activations differentiate between different behavioral acts and arise only under the condition that the behavior leading to the reward is being prepared or executed (Figure 11). Mechanistically speaking, the observed neural activations may be the result of convergent neural coding of reward and behavior, but from a theoretical point, the activations could represent evidence for neural correlates of goal-directed mechanisms. To distinguish between the two possibilities, it would be helpful to test explicitly the contingency requirement by varying the probabilities of reward in the presence versus absence of behavioral reactions. Further tests could employ reward devaluations to distinguish between goal-directed and habit mechanisms, as the relatively more simple habits might also rely on combined neural mechanisms of expected reward and behavioral action but lack the more flexible representations of reward that are the hallmark of goal mechanisms.

REWARD FUNCTIONS DEFINED BY MICROECONOMIC UTILITY THEORY

How can we compare apples and pears? We need a numerical scale in order to assess the influence of different rewards on behavior. A good way to quantify the value of individual rewards is to compare them in choice behavior. Given two options, I would choose the one that at this moment has the higher value for me. Give me the choice between a one-dollar bill and an apple, and you will see which one I prefer and thus my action will tell you whether the value of the apple for me is higher or lower or similar compared with one dollar. To be able to put a quantifiable, numerical value onto every reward, even when the value is short-lived, has enormous advantages for getting reward-related behavior under experimental control.

To obtain a more complete picture, we need to take into account the uncertainty with which rewards frequently occur. One possibility would be to weigh the value of individual rewards with the probability with which they occur, an approach taken by Pascal ca. 1650. The sum of the products of each potential reward and its probability defines the expected value (EV) of the probability distribution and thus the theoretically expected payoff of an option, according to

$$EV = \sum_i (p_i \cdot x_i); \quad i = 1, n; \quad n = \text{number of rewards.}$$

With increasing numbers of trials, the measured mean of the actually occurring distribution will approach the expected value. Pascal conjectured that human choice behavior could be approximated by this procedure.

Despite its advantages, expected value theory has limits when comparing very small with very large rewards or when comparing values at different start positions. Rather than following physical sizes of reward value in a linear fashion, human choice behavior in many instances increases more slowly as the values get higher, and the term of utility, or in some cases prospect, replaces the term of value when the impact of rewards on choices is assessed (Bernoulli 1738, Kahneman & Tversky 1984, Savage 1954, von Neumann & Morgenstern 1944). The utility function can be modeled by various equations (for detailed descriptions, see Gintis 2000, Huang & Litzenberger 1988), such as

1. The logarithmic utility function, $u(x) = \ln(x)$, yields a concave curve similar to the Weber (1850) function of psychophysics.
2. The power utility function, $u(x) = x^a$. With $a \in (0,1)$, and often $a \in [0.66, 0.75]$, the function is concave and resembles the power law of psychophysics (Stevens 1957). By contrast, $a = 1.0$ produces a linear function in which utility (value) = value. With $a > 1$, the curve becomes convex and increases faster toward higher values.
3. The exponential utility function, $u(x) = 1 - e^{-bx}$, produces a concave function for $b \in (0,1)$.
4. With the weighted reward value being expressed as utility, the expected value of a gamble becomes the expected utility (EU) according to

$$EU = \sum_i (p_i \cdot u(x_i)); i = 1, n; n = \text{number of rewards}.$$

Assessing the expected utility allows comparisons between gambles that have several outcomes with different values occurring at different probabilities. Note that a gamble with a single reward occurring at a $p < 1$ actually has two outcomes, the reward occurring with p and the nonreward with $(1 - p)$. A gamble with only one reward at $p = 1.0$ is called a safe option. Risk refers simply to known probabilities of < 1.0 and does not necessarily involve loss. Risky gambles have known probabilities; ambiguous gambles have probabilities unknown to the agent.

The shape of the utility function allows us to deal with the influence of uncertainty on decision-making. Let us assume an agent whose decision making is characterized by a concave utility function, as shown in Figure 12, who performs in a gamble with two outcomes of values 1 and 9 at $p = 0.5$ each (either the lower or the higher outcome will occur, with equal probability). The EV of the gamble is 5 (vertical dotted line), and the utility $u(\text{EV})$ (horizontal dotted line) lies between $u(1)$ and $u(9)$ (horizontal lines). Interestingly, $u(\text{EV})$ lies closer to $u(9)$ than to $u(1)$, suggesting that the agent foregoes more utility when the gamble produces $u(1)$ than she wins with $u(9)$ over $u(\text{EV})$. Given that outcomes 1 and 9 occur with the same frequency, this agent would profit more from a safe reward at EV, with $u(\text{EV})$, over the gamble. She should be risk averse. Thus, a concave utility function suggests risk aversion, whereas a convex function, in which an

agent foregoes less reward than she wins, suggests risk seeking. Different agents with different attitudes toward risk have differently shaped utility functions.

A direct measure of the influence of uncertainty is obtained by considering the difference between $u(EV)$ and the EU of the gamble. The EU in the case of equal probabilities is the mean of $u(1)$ and $u(9)$, as marked by $EU(1-9)$, which is considerably lower than $u(EV)$ and thus indicates the loss in utility due to risk. By comparison, the gamble of 4 and 6 involves a smaller range of reward magnitudes and thus less risk and less loss due to uncertainty, as seen by comparing the vertical bars associated with $EU(4-6)$ and $EU(1-9)$. This graphical analysis suggests that value and uncertainty of outcome can be considered as separable measures.

A separation of value and uncertainty as components of utility can be achieved mathematically by using, for example, the negative exponential utility function often employed in financial mathematics. Using the exponential utility function for EU results in

$$EU = \sum_i (p_i \cdot -e^{-b x_i}),$$

which can be developed by the Laplace transform into

$$EU = -e^{-b(EV - b/2 \cdot \text{var})},$$

where EV is expected value, var is variance, and the probability distribution p_i is Gaussian. Thus, EU is expressed as $f(EV, \text{variance})$. This procedure uses variance as a measure of uncertainty. Another measure of uncertainty is the entropy of information theory, which might be appropriate to use when dealing with information processing in neural systems, but entropy is not commonly employed for describing decision making in microeconomics.

Taken together, microeconomic utility theory has defined basic reward parameters, such as magnitude, probability, expected value, expected utility, and variance, that can be used for neurobiological experiments searching for neural correlates of decision making under uncertainty.

NEUROPHYSIOLOGY OF REWARD BASED ON ECONOMIC THEORY

Magnitude

The easiest quantifiable measure of reward for animals is the volume of juice, which animals can discriminate in submilliliter quantities (Tobler et al. 2005). Neurons show increasing responses to reward-predicting CSs with higher volumes of reward in a number of reward structures, such as the striatum (Cromwell & Schultz 2003) (Figure 13a), dorsolateral and orbital prefrontal cortex (Leon & Shadlen 1999, Roesch & Olson 2004, Wallis & Miller 2003), parietal and posterior cingulate cortex (McCoy et al. 2003, Musallam et al. 2004, Platt & Glimcher 1999),

and dopamine neurons (Sato et al. 2003, Tobler et al. 2005). Similar reward magnitude–discriminating activations are found in these structures in relation to other task events, before and after reward delivery. Many of these studies also report decreasing responses with increasing reward magnitude (Figure 13*b*), although not with dopamine neurons. The decreasing responses are likely to reflect true magnitude discrimination rather than simply the attention induced by rewards, which should increase with increasing magnitude.

Recent considerations cast doubt on the nature of some of the reward magnitude–discriminating, behavior-related activations, in particular in structures involved in motor and attentional processes, such as the premotor cortex, frontal eye fields, supplementary eye fields, parietal association cortex, and striatum. Some reward-related differences in movement-related activations might reflect the differences in movements elicited by different reward magnitudes (Lauwereyns et al. 2002, Roesch & Olson 2004). A larger reward might make the animal move faster, and increased neural activity in premotor cortex with larger reward might reflect the higher movement speed. Although a useful explanation for motor structures, the issue might be more difficult to resolve for areas more remote from motor output, such as prefrontal cortex, parietal cortex, and caudate nucleus. It would be helpful to correlate reward magnitude–discriminating activity in single neurons with movement parameters, such as reaction time and movement speed, and, separately, with reward parameters, and see where higher correlations are obtained. However, the usually measured movement parameters may not be sensitive enough to make these distinctions when neural activity varies relatively little with reward magnitude. On the other hand, inverse relationships, such as higher neural activity for slower movements associated with smaller rewards, would argue against a primarily motor origin of reward-related differences, as relatively few neurons show higher activity with slower movements.

Probability

Simple tests for reward probability involve CSs that differentially predict the probability with which a reward, as opposed to no reward, will be delivered for trial completion in Pavlovian or instrumental tasks. Dopamine neurons show increasing phasic responses to CSs that predict reward with increasing probability (Fiorillo et al. 2003, Morris et al. 2004). Similar increases in task-related activity occur in parietal cortex and globus pallidus during memory and movement-related task periods (Arkadir et al. 2004, Musallam et al. 2004, Platt & Glimcher 1999). However, reward-responsive tonically active neurons in the striatum do not appear to be sensitive to reward probability (Morris et al. 2004), indicating that not all neurons sensitive to reward may code its value in terms of probability. In a decision-making situation with varying reward probabilities, parietal neurons track the recently experienced reward value, indicating a memory process that would provide important input information for decision making (Sugrue et al. 2004).

Expected Value

Parietal neurons show increasing task-related activations with both the magnitude and probability of reward that do not seem to distinguish between the two components of expected value (Musallam et al. 2004). When the two value parameters are tested separately and in combination, dopamine neurons show monotonically increasing responses to CSs that predict increasing value (Tobler et al. 2005). The neurons fail to distinguish between magnitude and probability and seem to code their product (Figure 14*a*). However, the neural noise inherent in the stimulus-response relationships makes it difficult to determine exactly whether dopamine neurons encode expected value or expected utility. In either case, it appears as if neural responses show a good relationship to theoretical notions of outcome value that form a basis for decision making.

Uncertainty

Graphical analysis and application of the Laplace transform on the exponential utility function would permit experimenters to separate the components of expected value and utility from the uncertainty inherent in probabilistic gambles. Would the

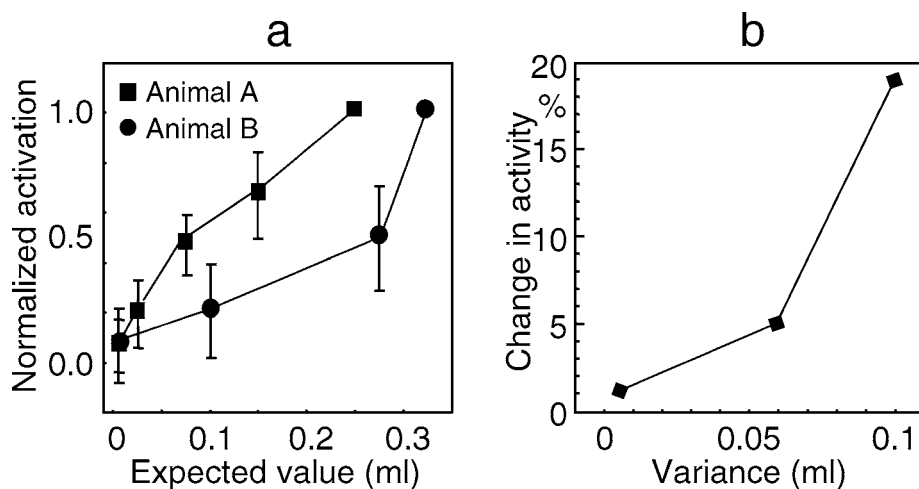


Figure 14 Separate coding of reward value and uncertainty in dopamine neurons. (*a*) Phasic response to conditioned, reward-predicting stimuli scales with increasing expected value (EV, summed magnitude \times probability). Data points represent median responses normalized to response to highest EV (animal A, 57 neurons; animal B, 53 neurons). Data from Tobler et al. (2005). (*b*) Sustained activation during conditioned stimulus–reward interval scales with increasing uncertainty, as measured by variance. Two reward magnitudes are delivered at $p = 0.5$ each (0.05–0.15, 0.15–0.5 ml, 0.05–0.5 ml). Ordinate shows medians of changes above background activity from 53 neurons. Note that the entropy stays 1 bit for all three probability distributions. Data from Fiorillo et al. (2003).

brain be able to produce an explicit signal that reflects the level of uncertainty, similar to producing a reward signal? For both reward and uncertainty, there are no specialized sensory receptors. A proportion of dopamine neurons show a sustained activation during the CS-reward interval when tested with CSs that predict reward at increasing probabilities, as opposed to no reward. The activation is highest for reward at $p = 0.5$ and progressively lower for probabilities further away from $p = 0.5$ in either direction (Fiorillo et al. 2003). The activation does not occur when reward is substituted by a visual stimulus. The activations appear to follow common measures of uncertainty, such as statistical variance and entropy, both of which are maximal at $p = 0.5$. Most of the dopamine neurons signaling reward uncertainty also show phasic responses to reward-predicting CSs that encode expected value, and the two responses coding different reward terms are not correlated with each other. When in a refined experiment two different reward magnitudes alternate randomly (each at $p = 0.5$), dopamine neurons show the highest sustained activation when the reward range is largest, indicating a relationship to the statistical variance and thus to the uncertainty of the reward (Figure 14*b*). In a somewhat comparable experiment, neurons in posterior cingulate cortex show increased task-related activations as animals choose among rewards with larger variance compared with safe options (McCoy & Platt 2003). Although only a beginning, these data suggest that indeed the brain may produce an uncertainty signal about rewards that could provide essential information when making decisions under uncertainty. The data on dopamine neurons suggest that the brain may code the expected value separately from the uncertainty, just as the two terms constitute separable components of expected utility when applying the Laplace transform on the exponential utility function.

CONCLUSIONS

It is intuitively simple to understand that the use of well-established behavioral theories can only be beneficial when working with mechanisms underlying behavioral reactions. Indeed, these theories can very well define the different functions of rewards on behavior. It is then a small step on firm ground to base the investigation of neural mechanisms underlying the different reward functions onto the phenomena characterized by these theories. Although each theory has its own particular emphasis, they deal with the same kinds of outcome events of behavior, and it is more confirmation than surprise to see that many neural reward mechanisms can be commonly based on, and understood with, several theories. For the experimenter, the use of different theories provides good explanations for an interesting spectrum of reward functions that may not be so easily accessible by using only a single theory. For example, it seems that uncertainty plays a larger role in parts of microeconomic theory than in learning theory, and the investigation of neural mechanisms of uncertainty in outcomes of behavior can rely on several hundred years of thoughts about decision making (Pascal 1650 in Glimcher 2003, Bernoulli 1738).

ACKNOWLEDGMENTS

This article is based on lectures delivered at a Society for Neuroscience Meeting in October 2004 in San Diego and on a Max-Planck Symposium in January 2005 in Frankfurt, Germany. The author wishes to thank Drs. Anthony Dickinson and Peter Bossaerts for illuminating discussions on behavioral theories. Our work was supported by grants from the Wellcome Trust, Swiss National Science Foundation, Human Frontiers, and European Community.

The *Annual Review of Psychology* is online at <http://psych.annualreviews.org>

LITERATURE CITED

- Amador N, Schlag-Rey M, Schlag J. 2000. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J. Neurophysiol.* 84:2166–70
- Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. 1994. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci.* 14:3969–84
- Apicella P, Ljungberg T, Scarnati E, Schultz W. 1991. Responses to reward in monkey dorsal and ventral striatum. *Exp. Brain Res.* 85:491–500
- Apicella P, Scarnati E, Ljungberg T, Schultz W. 1992. Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J. Neurophysiol.* 68:945–60
- Arkadir D, Morris G, Vaadia E, Bergman H. 2004. Independent coding of movement direction and reward prediction by single pallidal neurons. *J. Neurosci.* 24:10047–56
- Balleine B, Dickinson A. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–19
- Barraclough D, Conroy ML, Lee DJ. 2004. Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* 7:405–10
- Bernoulli J. (1738) 1954. Exposition of a new theory on the measurement of risk. *Econometrica* 22:23–36
- Bowman EM, Aigner TG, Richmond BJ. 1996. Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J. Neurophysiol.* 75:1061–73
- Critchley HG, Rolls ET. 1996. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.* 75:1673–86
- Cromwell HC, Schultz W. 2003. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.* 89:2823–38
- Delaney K, Gelperin A. 1986. Post-ingestive food-aversion learning to amino acid deficient diets by the terrestrial slug *Limax maximus*. *J. Comp. Physiol. A* 159:281–95
- Dickinson A. 1980. *Contemporary Animal Learning Theory*. Cambridge, UK: Cambridge Univ. Press
- Dickinson A, Balleine B. 1994. Motivational control of goal-directed action. *Anim. Learn. Behav.* 22:1–18
- Dickinson A, Hall G, Mackintosh NJ. 1976. Surprise and the attenuation of blocking. *J. Exp. Psychol. Anim. Behav. Process.* 2:313–22
- Dorris MC, Glimcher PW. 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365–78
- Fiorillo CD, Tobler PN, Schultz W. 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–902

- Gintis H. 2000. *Game Theory Evolving*. Princeton, NJ: Princeton Univ. Press
- Glimcher PW. 2003. *Decisions, Uncertainty and the Brain*. Cambridge, MA: MIT Press
- Hassani OK, Cromwell HC, Schultz W. 2001. Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J. Neurophysiol.* 85:2477–89
- Hikosaka K, Watanabe M. 2000. Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb. Cortex* 10:263–71
- Hikosaka O, Sakamoto M, Usui S. 1989. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.* 61:814–32
- Hollerman JR, Schultz W. 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* 1:304–9
- Hollerman JR, Tremblay L, Schultz W. 1998. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* 80:947–63
- Hrupka BJ, Lin YM, Gietzen DW, Rogers QR. 1997. Small changes in essential amino acid concentrations alter diet selection in amino acid-deficient rats. *J. Nutr.* 127:777–84
- Huang C-F, Litzenberger RH. 1988. *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice Hall
- Ito S, Stuphorn V, Brown JW, Schall JD. 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302:120–22
- Kahneman D, Tversky A. 1984. Choices, values, and frames. *Am. Psychol.* 4:341–50
- Kamin LJ. 1969. Selective association and conditioning. In *Fundamental Issues in Instrumental Learning*, ed. NJ Mackintosh, WK Honig, pp. 42–64. Halifax, NS: Dalhousie Univ. Press
- Kawagoe R, Takikawa Y, Hikosaka O. 1998. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1:411–16
- Kimura M, Rajkowski J, Evarts E. 1984. Tonically discharging putamen neurons exhibit set-dependent responses. *Proc. Natl. Acad. Sci. USA* 81:4998–5001
- Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O. 2002. Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J. Neurophysiol.* 87:1488–98
- Lauwereyns J, Watanabe K, Coe B, Hikosaka O. 2002. A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–17
- Leon MI, Shadlen MN. 1999. Effect of expected reward magnitude on the responses of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–25
- Liu Z, Richmond BJ. 2000. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J. Neurophysiol.* 83:1677–92
- Ljungberg T, Apicella P, Schultz W. 1992. Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 67:145–63
- Mackintosh NJ. 1975. A theory of attention: variations in the associability of stimulus with reinforcement. *Psychol. Rev.* 82:276–98
- Markowitsch HJ, Pritzel M. 1976. Reward-related neurons in cat association cortex. *Brain Res.* 111:185–88
- Matsumoto K, Suzuki W, Tanaka K. 2003. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301:229–32
- Maunsell JHR. 2004. Neuronal representations of cognitive state: reward or attention? *Trends Cogn. Sci.* 8:261–65
- McCoy AN, Crowley JC, Haghghian G, Dean HL, Platt ML. 2003. Saccade reward signals in posterior cingulate cortex. *Neuron* 40:1031–40
- Mirenowicz J, Schultz W. 1994. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* 72:1024–27
- Mirenowicz J, Schultz W. 1996. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379:449–51

- Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. 2004. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43:133–43
- Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. 2004. Cognitive control signals for neural prosthetics. *Science* 305:258–62
- Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. 2004. Dopamine neurons can represent context-dependent prediction error. *Neuron* 41:269–80
- Nakamura K, Mikami A, Kubota K. 1992. Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J. Neurophysiol.* 67:1447–63
- Nishijo H, Ono T, Nishino H. 1988. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J. Neurosci.* 8:3570–83
- O'Doherty JP. 2004. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14:769–76
- Pavlov PI. 1927. *Conditioned Reflexes*. London: Oxford Univ. Press
- Pearce JM, Hall G. 1980. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87:532–52
- Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400:233–38
- Pratt WE, Mizumori SJY. 1998. Characteristics of basolateral amygdala neuronal firing on a spatial memory task involving differential reward. *Behav. Neurosci.* 112:554–70
- Pratt WE, Mizumori SJY. 2001. Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task. *Behav. Brain Res.* 123:165–83
- Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. *Science* 276:821–24
- Ravel S, Legallet E, Apicella P. 1999. Tonicly active neurons in the monkey striatum do not preferentially respond to appetitive stimuli. *Exp. Brain Res.* 128:531–34
- Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, ed. AH Black, WF Prokasy, pp. 64–99. New York: Appleton-Century-Crofts
- Roesch MR, Olson CR. 2004. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 304:307–10
- Rogers QR, Harper AE. 1970. Selection of a solution containing histidine by rats fed a histidine-imbalanced diet. *J. Comp. Physiol. Psychol.* 72:66–71
- Roitman MF, Wheeler RA, Carelli RM. 2005. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45:587–97
- Rolls ET, Critchley HD, Mason R, Wakeman EA. 1996. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J. Neurophysiol.* 75:1970–81
- Rolls ET, Scott TR, Sienkiewicz ZJ, Yaxley S. 1988. The responsiveness of neurones in the frontal opercular gustatory cortex of the macaque monkey is independent of hunger. *J. Physiol.* 397:1–12
- Rolls ET, Sienkiewicz ZJ, Yaxley S. 1989. Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *Eur. J. Neurosci.* 1:53–60
- Satoh T, Nakai S, Sato T, Kimura M. 2003. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* 23:9913–23
- Savage LJ. 1954. *The Foundations of Statistics*. New York: Wiley
- Schoenbaum G, Chiba AA, Gallagher M. 1998. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* 1:155–59
- Schoenbaum G, Chiba AA, Gallagher M. 1999.

- Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* 19:1876–84
- Schultz W, Apicella P, Scarnati E, Ljungberg T. 1992. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* 12:4595–10
- Schultz W, Dayan P, Montague RR. 1997. A neural substrate of prediction and reward. *Science* 275:1593–99
- Schultz W, Dickinson A. 2000. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23:473–500
- Schultz W, Romo R. 1987. Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* 57:201–17
- Shidara M, Aigner TG, Richmond BJ. 1998. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* 18:2613–25
- Shidara M, Richmond BJ. 2002. Anterior cingulate: single neuron signals related to degree of reward expectancy. *Science* 296:1709–11
- Stevens SS. 1957. On the psychophysical law. *Psychol. Rev.* 64:153–81
- Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–87
- Thorndike EL. 1911. *Animal Intelligence: Experimental Studies*. New York: MacMillan
- Thorpe SJ, Rolls ET, Maddison S. 1983. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp. Brain Res.* 49:93–115
- Tobler PN, Dickinson A, Schultz W. 2003. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J. Neurosci.* 23:10402–10
- Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–45
- Tremblay L, Hollerman JR, Schultz W. 1998. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* 80:964–77
- Tremblay L, Schultz W. 1999. Relative reward preference in primate orbitofrontal cortex. *Nature* 398:704–8
- Tremblay L, Schultz W. 2000a. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *J. Neurophysiol.* 83:1864–76
- Tremblay L, Schultz W. 2000b. Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *J. Neurophysiol.* 83:1877–85
- Ungless MA, Magill PJ, Bolam JP. 2004. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* 303:2040–42
- von Neumann J, Morgenstern O. 1944. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press
- Waelti P. 2000. *Activité phasique des neurones dopaminergiques durant une tâche de discrimination et une tâche de blocage chez le primate vigile*. PhD thesis. Univ. de Fribourg, Switzerland
- Waelti P, Dickinson A, Schultz W. 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48
- Wallis JD, Miller EK. 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* 18:2069–81
- Wang Y, Cummings SL, Gietzen DW. 1996. Temporal-spatial pattern of c-fos expression in the rat brain in response to indispensable amino acid deficiency. I. The initial recognition phase. *Mol. Brain Res.* 40:27–34
- Watanabe M. 1989. The appropriateness of behavioral responses coded in post-trial activity of primate prefrontal units. *Neurosci. Lett.* 101:113–17
- Watanabe M. 1996. Reward expectancy in primate prefrontal neurons. *Nature* 382:629–32
- Watanabe M, Hikosaka K, Sakagami M, Shirakawa SI. 2002. Coding and monitoring

- of behavioral context in the primate prefrontal cortex. *J. Neurosci.* 22:2391–400
- Weber EH. 1850. Der Tastsinn und das Gemeingefuehl. In *Handwoerterbuch der Physiologie*, Vol. 3, Part 2, ed. R Wagner, pp. 481–588. Braunschweig, Germany: Vieweg
- Wise RA. 2002. Brain reward circuitry: insights from unsensed incentives. *Neuron* 36:229–40

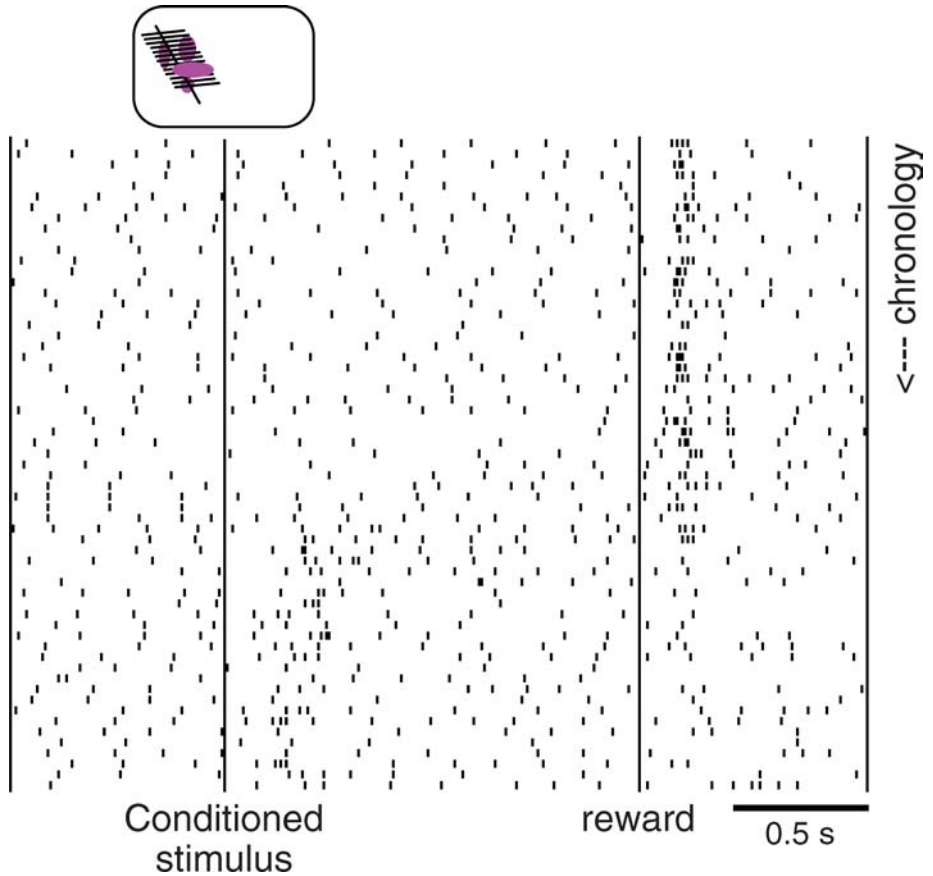


Figure 2 Testing the contiguity requirement for associative learning: acquisition of neural response in a single dopamine neuron during a full learning episode. Each line of dots represents a trial, each dot represents the time of the discharge of the dopamine neuron, the vertical lines indicate the time of the stimulus and juice reward, and the picture above the raster shows the visual conditioned stimulus presented to the monkey on a computer screen. Chronology of trials is from top to bottom. The top trial shows the activity of the neuron while the animal saw the stimulus for the first time in its life, whereas it had previous experience with the liquid reward. Data from Waelti (2000).

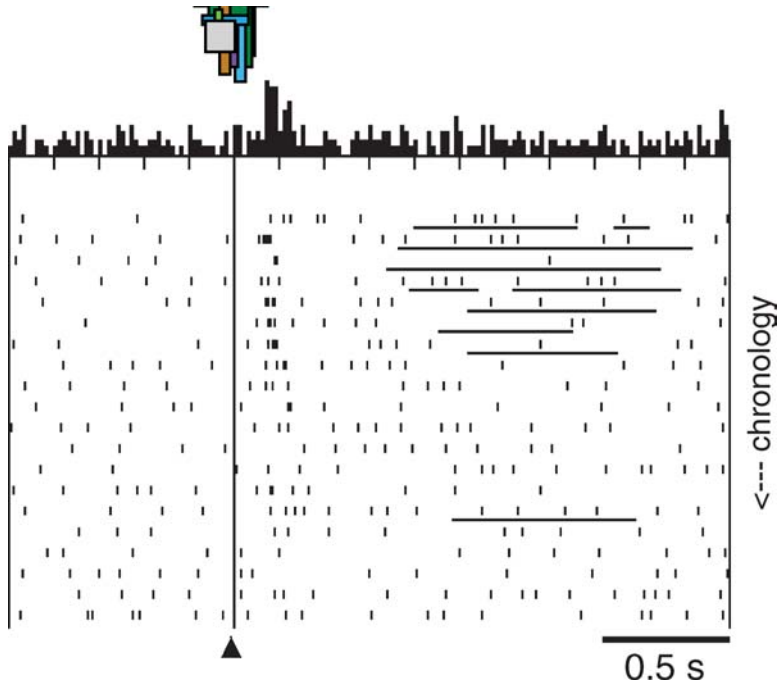
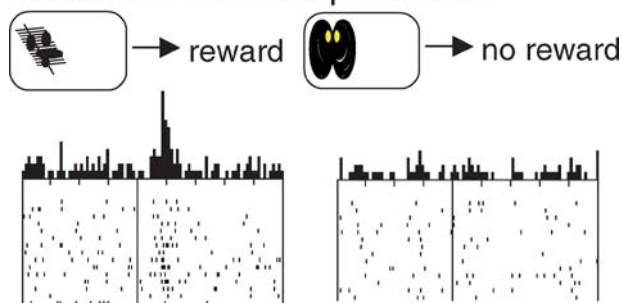
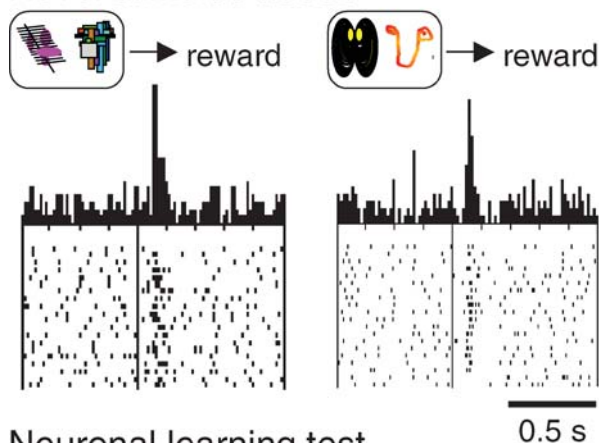


Figure 5 Loss of response in dopamine neuron to the conditioned stimulus following withholding of reward. This manipulation violates the contiguity requirement (co-occurrence of reward and stimulus) and produces a negative prediction error that brings down the associative strength of the stimulus. The contingency moves toward the neutral situation. Data from Tobler et al. (2003).

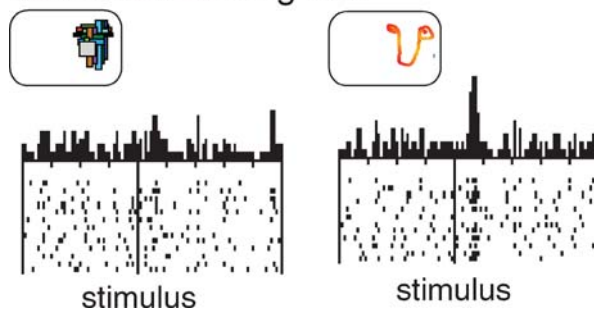
Established reward prediction



Novel stimulus added



Neuronal learning test



See legend on next page

Figure 6 Acquisition of dopamine response to reward-predicting stimulus is governed by prediction error. Neural learning is blocked when the reward is predicted by another stimulus (*left*) but is intact in the same neuron when reward is unpredicted in control trials with different stimuli (*right*). The neuron has the capacity to respond to reward-predicting stimuli (*top left*) and discriminates against unrewarded stimuli (*top right*). The addition of a second stimulus results in maintenance and acquisition of response, respectively (*middle*). Testing the added stimulus reveals absence of learning when the reward is already predicted by a previously conditioned stimulus (*bottom left*). Data from Waelti et al. (2001).

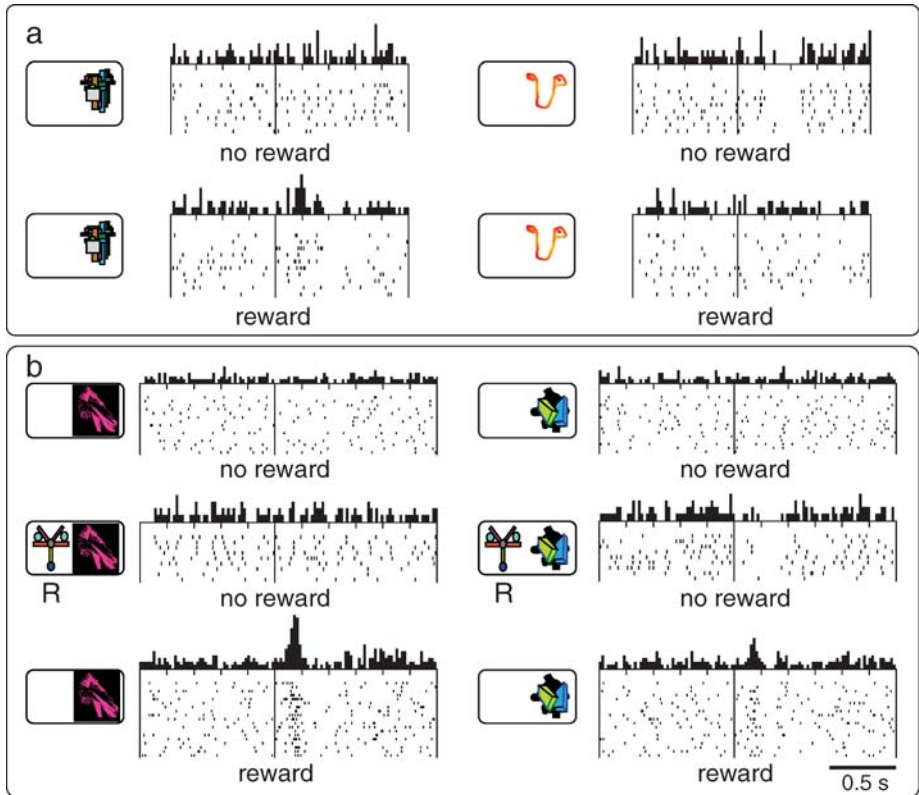


Figure 8 Coding of prediction errors by dopamine neurons in specific paradigms. (a) Blocking test. Lack of response to absence of reward following the blocked stimulus, but positive signal to delivery of reward (*left*), in contrast to control trials with a learned stimulus (*right*). Data from Waelti et al. 2001. (b) Conditioned inhibition task. Lack of response to absence of reward following the stimulus predicting no reward (*top*), even if the stimulus is paired with an otherwise reward-predicting stimulus (R, *middle*, summation test), but strong activation to reward following a stimulus predicting no reward (*bottom*). These responses contrast with those following the neutral control stimulus (*right*). Data from Tobler et al. (2003).

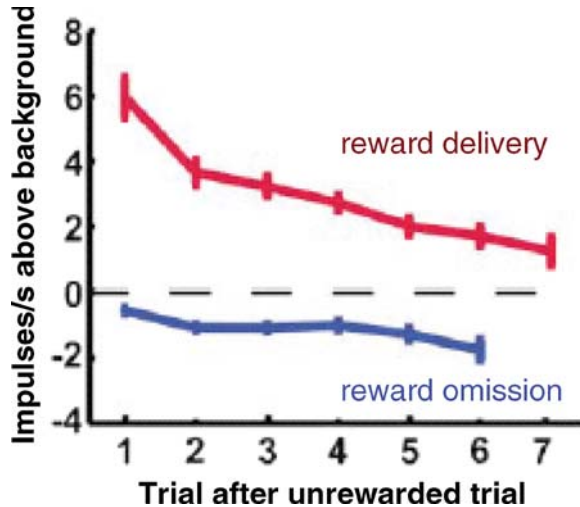


Figure 9 Time information contained in predictions acting on dopamine neurons. In the particular behavioral task, the probability of reward, and thus the reward prediction, increases with increasing numbers of trials after the last reward, reaching $p = 1.0$ after six unrewarded trials. Accordingly, the positive dopamine error response to a rewarding event decreases over consecutive trials (*upper curve*), and the negative response to a nonrewarding event becomes more prominent (*lower curve*). Data are averaged from 32 dopamine neurons studied by Nakahara et al. (2004), © Cell. Press.

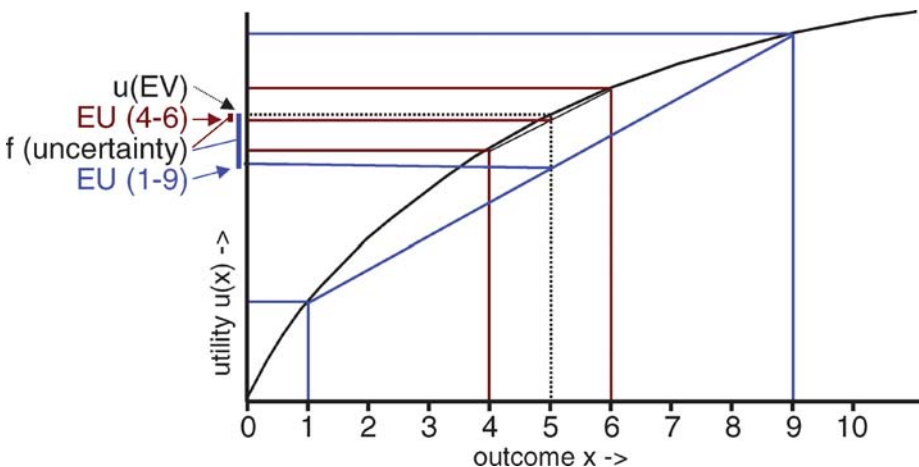


Figure 12 A hypothetical concave utility function. EV, expected value (5 in both gambles with outcomes of 1 and 9, and 4 and 6); EU, expected utility. See text for description.

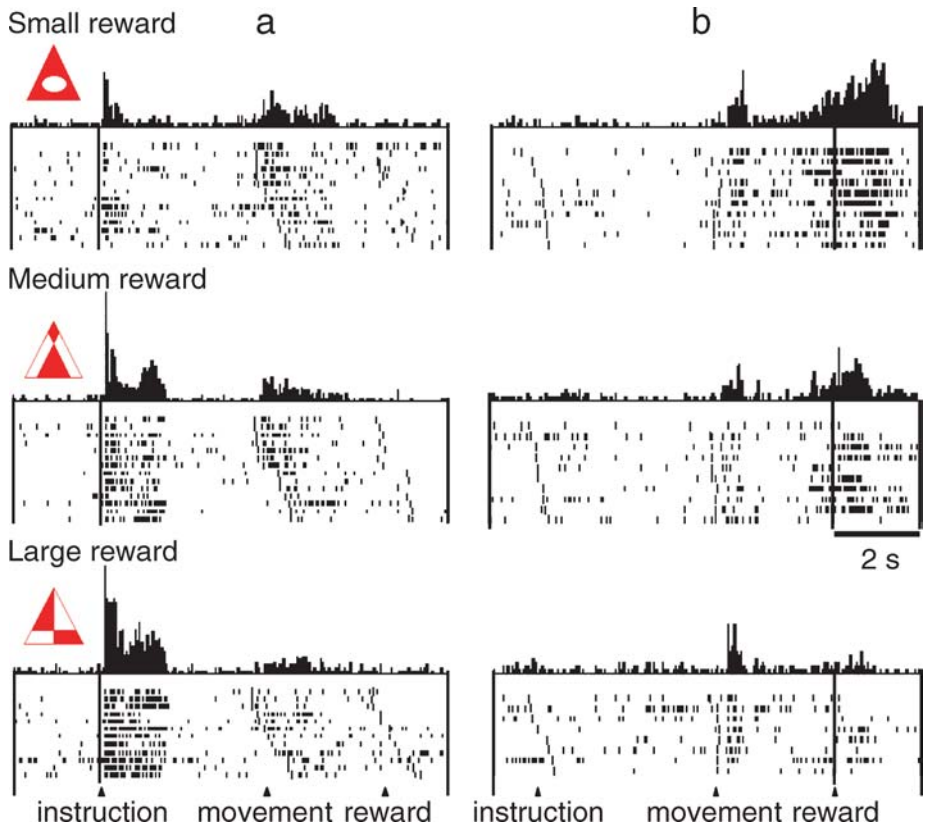


Figure 13 Discrimination of reward magnitude by striatal neurons. (a) Increasing response in a caudate neuron to instruction cues predicting increasing magnitudes of reward (0.12, 0.18, 0.24 ml). (b) Decreasing response in a ventral striatum neuron to rewards with increasing volumes. Data from Cromwell & Schultz 2003.