Dopamine responses comply with basic assumptions of formal learning theory

Pascale Waelti*, **Anthony Dickinson†** & **Wolfram Schultz***

* *Institute of Physiology and Programme in Neuroscience, University of Fribourg, CH-1700 Fribourg, Switzerland*
† *Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK*

According to contemporary learning theories, the discrepancy, or error, between the actual and predicted reward determines whether learning occurs when a stimulus is paired with a reward. The role of prediction errors is directly demonstrated by the observation that learning is blocked when the stimulus is paired with a fully predicted reward. By using this blocking procedure, we show that the responses of dopamine neurons to conditioned stimuli was governed differentially by the occurrence of reward prediction errors rather than stimulus–reward associations alone, as was the learning of behavioural reactions. Both behavioural and neuronal learning occurred predominantly when dopamine neurons registered a reward prediction error at the time of the reward. Our data indicate that the use of analytical tests derived from formal behavioural learning theory provides a powerful approach for studying the role of single neurons in learning.

Classic theories assume that predictive learning occurs whenever a stimulus is paired with a reward or punishment[1,2]. However, more recent analyses of associative learning argue that simple temporal contiguity between a stimulus and a reinforcer is not sufficient for learning and that a discrepancy between the reinforcer that is predicted by a stimulus and the actual reinforcer is also required[3–6]. This discrepancy can be characterized as a 'prediction error'[7]. Presentations of surprising or unpredicted reinforcers generate positive prediction errors, and thereby support learning, whereas omissions of predicted reinforcers generate negative prediction errors and lead to reduction or extinction of learned behaviour. Expected reinforcers do not generate prediction errors and therefore fail to support further learning even when the stimulus is consistently paired with the reinforcer. Modelling studies have shown that neuronal messages encoding prediction errors can act as explicit teaching signals for modifying the synaptic connections that underlie associative learning[8–17].

Current research suggests that one of the principal reward systems of the brain involves dopamine neurons[18–21]. Both psychopharmacological manipulations and lesions of the dopamine system impair reward-driven behaviour of animals[18,21], and drugs of abuse, which provide strong artificial rewards, act via dopamine neurons[19,20]. Neurobiological investigations of associative learning have shown that dopamine neurons respond phasically to rewards in a manner compatible with the coding of prediction errors[22–28], whereas slower dopamine changes are involved in a larger spectrum of motivating events[18,28,29]. Dopamine neurons show short-latency, phasic activations when rewards occur unpredictably, they are not modulated by fully predicted rewards and show phasically reduced activity when predicted rewards are omitted. During initial learning, when rewards occur unpredictably, dopamine neurons are activated by rewards. They gradually lose the response as the reward becomes increasingly predicted[24,25,27]. The conditioned, reward-predicting stimulus starts to drive the dopamine neurons as they lose their response to the reward itself[24,25]. Dopamine neurons also respond to novel, attention-generating and motivating stimuli[24], indicating that attentional mechanisms also contribute[28,30]. Results from neuronal modelling suggest that the responses of dopamine neurons to primary rewards and conditioned stimuli may constitute particularly effective teaching signals for associative learning[4,16,17,31,32] and embody the properties of the teaching signal of the temporal difference reinforcement model[33–36].

The present experiment explored how dopamine neurons of primates acquire responses to reward-predicting stimuli. In order to determine the concordance between dopamine responses and reward prediction errors, we employed the canonical paradigm for assessing the role of prediction errors in learning, the blocking test[37]. We demonstrated behavioural sensitivity to prediction errors, investigated the acquisition of neuronal responses to stimuli, and related behavioural and neuronal learning to the responses of dopamine neurons at the time of the reward.

Blocking and prediction errors

The blocking paradigm generated differential prediction errors by training a target stimulus in compound with a pretrained stimulus. As the pretrained stimulus has already been established as a predictor of the reinforcer, the reinforcer on compound trials is expected and therefore generates a minimal prediction error. Consequently, learning about the target stimulus is blocked. The blocking effect demonstrates that prediction errors, rather than simple stimulus–reinforcer pairings, play a crucial role in human conditioning[38], causal learning[39], animal conditioning[37] and artificial network learning[31].

We employed a visual stimulus A which predicted the delivery of juice to the monkey (A+ trials), whereas a control stimulus B was not followed by reward (B– trials) (Fig. 1a). Learning was assessed by licking at a spout in anticipation of the reward in A+ but not B– trials. Reward delivery was not dependent on the animal's anticipatory licks in this classical (Pavlovian) conditioning procedure. After training, the reward following stimulus A and the absence of reward following B were predicted and should not have generated prediction errors. During subsequent compound training, two stimuli (X and Y) were presented simultaneously with A and B, respectively, and both the AX and BY compounds were paired with reward in equivalent numbers of trials. In AX+ trials, the reward was already fully predicted by the pretrained stimulus A and, therefore, the presentation of the reward should not have generated a prediction error. By contrast, in the BY+ control trials, the reward was predicted by neither stimulus, and the occurrence of reward should have generated a prediction error. The animals continued to show anticipatory licking in AX+ trials, as in the prior A+ trials, and they learned to lick in anticipation of reward in BY+ trials (Fig. 1b).

The crucial test trials with the stimulus presented alone showed that learning about stimulus X was blocked relative to the control stimulus Y which had become an effective reward predictor (Fig. 1a bottom). Median durations of licking (50th percentile) during a 2.0-s period after stimulus onset were 0 ms after stimulus X but 323 ms after Y (400 trials; $P < 0.0001$; Wilcoxon test; X and Y tested without reward). The failure of the monkeys to respond to stimulus X in spite of its prior pairing with the reward suggests that learning depends upon the reward generating a prediction error.

An assessment of eye movements revealed that both rewarded and unrewarded stimuli were detected with comparable saccadic latencies (means in A+ and B– trials were 191 and 192 ms, respectively; $n = 133/123$; $P > 0.2$, t-test). Comparisons between stimulus and eye positions (Fig. 1c and d) demonstrate that the monkeys fixated all four stimuli despite their differential associations with reward prediction errors, including the blocked stimulus X. These data suggest that blocking was not due to a failure of the monkeys to detect stimulus X.

Conditions for neuronal learning

Based on the results from behavioural learning, we investigated how dopamine neurons acquired responses to conditioned stimuli in relation to prediction errors. We first assessed how 286 dopamine neurons discriminated between rewarded and unrewarded stimuli (8, 201 and 77 neurons in substantia nigra and ventral tegmental area groups A8, A9 and A10, respectively). We found that 200 dopamine neurons were activated by stimulus A, and that 150 of them discriminated between the reward-predicted stimulus A and nonpredictive stimulus B, either in an all-or-none fashion (75 neurons; Fig. 2a) or with weaker activations to stimulus B than A (75 neurons). Fifty of the 200 neurons showed comparable responses to the two stimuli, none of them being stronger to stimulus B than A. The incidence of nondiscriminating B responses increased slightly from 22% to 29% over several consecutively employed picture sets. Most activations to stimulus B were followed by depressions (68 of 75 neurons), thus producing characteristic biphasic responses (see below). Responding neurons were found insignificantly more frequently in the medial of three equidistant mediolateral regions of dopamine neurons ($P > 0.1$; chi-squared test). The rewarded compounds AX and BY activated 94 of 137 dopamine neurons (5, 96 and 36 neurons in groups A8, A9 and A10, respectively), none of them being activated in only one trial type (Fig. 2b). The latencies of neuronal responses were less than 100 ms and thus were shorter than the onset of saccadic eye movements towards the stimuli, indicating that the discriminations may have been based more on stimulus positions than on visual attributes requiring fixation.

Blocking was tested on responses to conditioned stimuli in 85 dopamine neurons (2, 73 and 10 neurons in groups A8, A9 and A10, respectively) (Fig. 1e). None of them showed exclusive responses to stimulus X but, more importantly, 39 were not activated by stimulus X while remaining responsive to stimulus Y (Fig. 2c). A further 16

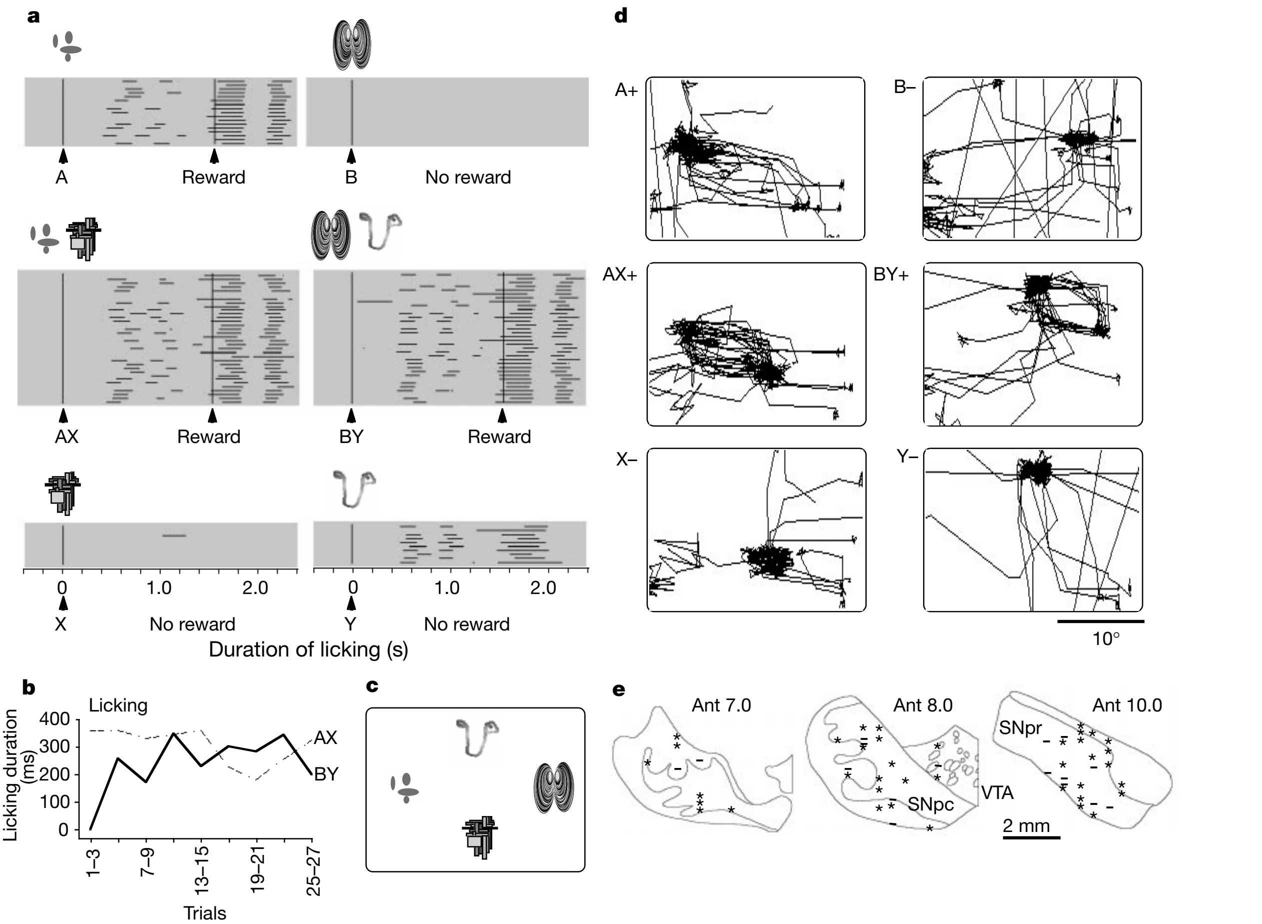

Figure 1 Behavioural performance in the blocking paradigm and neuronal localizations. **a**, Licking behaviour in the six trial types: A, B, pretraining; AX, BY, compound learning; X, Y, learning test. Horizontal lines indicate periods of licking. Note the licking in late Y-test trials in the absence of reward. **b**, Learning curves of durations of reward-anticipatory licking to compounds BY and AX during the 1.5-s stimulus–reward interval. **c**, Positions of the four stimuli on the computer screen in front of the animal. **d**, Eye positions during the 1.5-s stimulus–reward interval in the six trial types after learning (up is upward, right is rightward; eight trials are superimposed in each type). Darker areas indicate increased eye fixations on parts of the stimuli. The eyes fixated within a 2° diameter in A+, X– and Y– trials for mean durations of 444, 367, 551 and 698 ms per trial of 1,500 ms, respectively ($F = 1.630$, degrees of freedom, d.f. $= 3$, $P > 0.2$, one-way ANOVA). **e**, Histologically reconstructed positions of dopamine neurons responding to stimuli X or Y (asterisks indicate exclusive activations to Y but not to X, $n = 39$; dashes indicate activations to both X and Y, $n = 16$). SNpc, substantia nigra pars compacta, SNpr, substantia nigra pars reticulata, VTA, ventral tegmental area. Ant 7.0–10.0: levels anterior to the interaural stereotaxic line.

neurons were driven by both X and Y. Of these, four showed significantly weaker activations to X than Y, 10 showed comparable activations, and two responded more strongly to X than to Y. Eight of the 16 neurons had biphasic activation–depression responses (Fig. 2d). A further 30 neurons responded to neither X nor Y. The distribution of responding neurons varied insignificantly among groups A8, A9 and A10 or among three equidistant mediolateral regions of dopamine neurons ($P > 0.3$). Thus the dopamine neurons acquired stronger responses to the reward-predicting stimulus Y than to the redundant stimulus X (Fig. 2e), although both stimuli had received the same numbers of pairings with reward during compound training. Both behavioural and neuronal learning about stimulus X was blocked in the absence of a reward prediction error.

Learning and the dopamine reward signal

Subsequently we investigated whether the acquisition of behavioural and neuronal responses to the conditioned stimuli was related to the dopamine response at the time of the reward. After pretraining, the behavioural measures indicated that the reward following stimulus A and the absence of reward following stimulus B were predicted and should not have generated prediction errors. Correspondingly, dopamine activity remained at the baseline after the reward in A trials and at the time of no reward in B trials (Fig. 3a). By contrast, a surprising reward following the usually unrewarded stimulus B should have generated a positive prediction error and did, in fact, activate 11 of 13 dopamine neurons in occasional tests (Fig. 3a bottom). Similarly, the unexpected omission of reward after stimulus A, which should generate a negative prediction error, depressed 8 of 10 dopamine neurons at the predicted time of reward. These data suggest that the previously demonstrated dopamine reward prediction error signal[26–28] was operational in the current paradigm.

We then assessed whether dopamine neurons differentially reported reward prediction errors during the learning of behavioural and neuronal responses. The prediction error was low during initial AX+ trials with the reward already predicted by stimulus A. Accordingly, statistically significant activations following the reward were seen in none of six dopamine neurons tested in initial blocks of AX trials alone and in only 9 of 38 dopamine neurons tested in initial AX–BY trials (Fig. 3b). By contrast, 19 of the 38 neurons showed significant reward activations in initial BY+ trials when, on the basis of pretraining with unrewarded stimulus B, the reward was unpredicted and therefore should have generated a strong prediction error. The reward activations disappeared gradually during learning and were observed as late as 234 and 610 trials after onset of compound learning with two picture sets, respectively. All six AX-tested neurons were in group A9; of the 38 neurons, 29 were in group A9 and 9 in A10. Thus the dopamine neurons showed stronger reward activations with larger positive prediction errors and weaker activations with smaller prediction errors.

The responses at the time of the reward should also reflect the differential predictive status of stimuli X and Y. As predictive learning to stimulus X was blocked, the omission of reward following this stimulus should not have generated a negative prediction error. Accordingly, only one of 85 neurons was depressed by reward omission following stimulus X, whereas 30 neurons showed a depression after reward omission following stimulus Y (Fig. 3c). The occasional test presentation of a reward after stimulus X, but

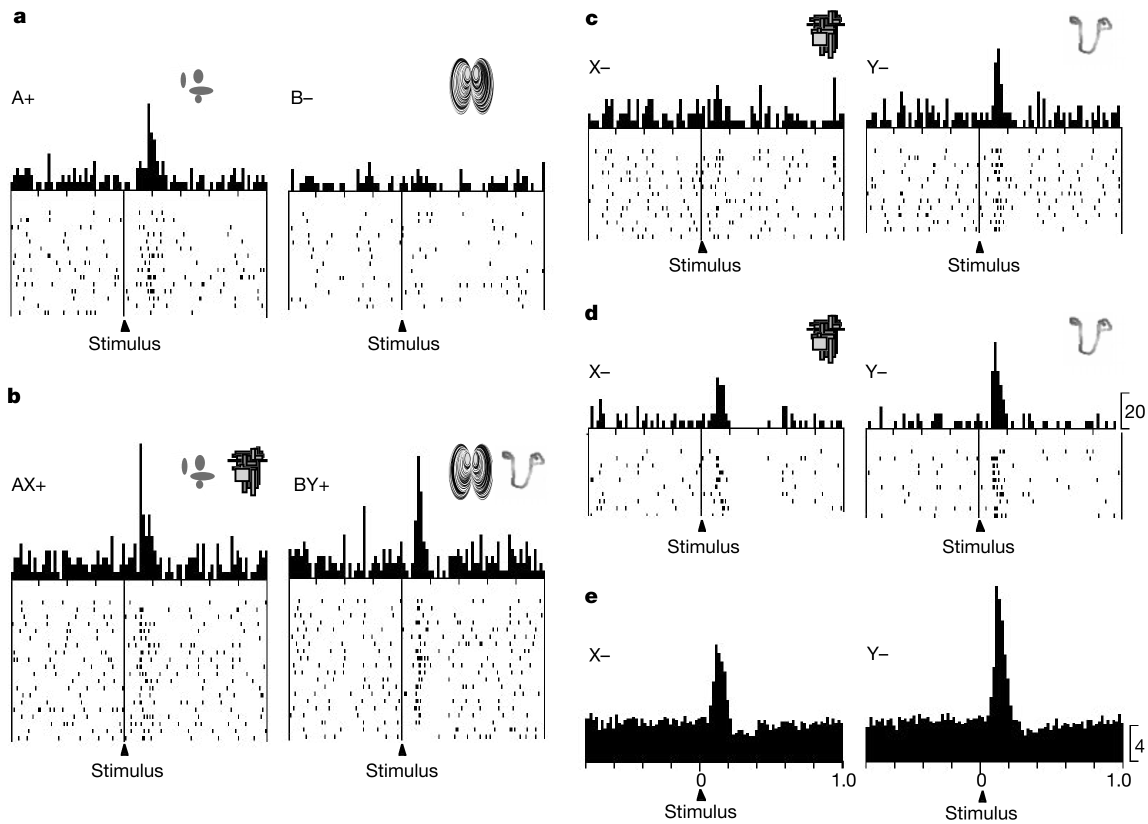

Figure 2 Acquisition of dopamine responses to conditioned stimuli depends on prediction errors in the blocking paradigm. **a**, Pretraining; **b**, after compound learning; **c**–**e**, learning tests. **a**, Differential activations following reward-predicting stimulus A but not unrewarded stimulus B. **b**, Maintained activation to reward-predicting compounds AX and acquired activation to BY. **c**, Absent (blocked) neuronal response to stimulus X but acquired response to stimulus Y (same neuron as in **b**). **d**, Occasional small activation followed by depression to blocked stimulus X, probably due to stimulus generalization. **e**, Averaged population histograms of the 85 dopamine neurons tested with stimuli X and Y after compound learning (normalized for trial numbers). In **a**–**d**, dots denote neuronal impulses, referenced in time to the stimuli. Each line of dots shows one trial, the original sequence being from top to bottom in each panel. Histograms contain the sums of raster dots. Vertical calibrations indicate 20 impluses s^{-1} (in **a**–**d**) and 4 impulses s^{-1} (in **e**).

not stimulus Y, should have generated a positive prediction error and, indeed, induced an activation in all three dopamine neurons tested (Fig. 3d). Thus the contrasting activations and depressions at the time of the reward in X and Y trials corresponded to the different prediction errors generated by the presentations and omissions of the reward in these trials.

Response generalization

As noted above, 16 neurons showed activations to the blocked stimulus X, which in eight were followed by depressions (Fig. 2d). We investigated the nature of these responses by comparing them with responses to stimuli A and B, which had well defined differential reward predictions. Thirteen of the 16 X-responding neurons also showed responses to stimulus B before this stimulus had ever been paired with reward (Fig. 4a). Moreover, the strong activation by a surprising reward on occasional B test trials (Fig. 4b) confirmed that B did not predict the reward in spite of its activation of the dopamine neurons. The reward activation contrasted with the absence of any reward response following the consistently reward-predicting stimulus A (Fig. 4a). The neuronal responses to reward omission revealed similar contrasting predictive functions. Although neuronal activity was depressed by reward omission on test trials with the usually rewarded stimulus A (Fig. 4b), the activity remained at the baseline level at the time of reward on unrewarded B trials (Fig. 4a). Thus neuronal prediction errors occurred differentially at the time of the reward on A versus B trials, although both stimuli triggered neuronal activations (Fig. 4c). Therefore, the responses to stimulus B were probably not related to reward prediction but may have reflected response generalization from the reward-predicting stimulus A (ref. 28). The notable response similarities between stimuli X and B (Figs 2d versus 4a, b) may suggest that the X responses were also unrelated to explicit reward prediction but possibly reflected generalization from the reward-predicting stimulus Y to stimulus X.

Discussion

Our data demonstrate that the activity of single dopamine neurons in the blocking paradigm conformed to basic assumptions of formal learning rules. The acquisition of neuronal responses to conditioned, reward-predicting stimuli was governed by the stringent criteria for learning developed in behavioural theories. Neuronal learning, like behavioural learning, depended crucially on the reward prediction error and occurred less readily with stimulus–reward associations alone. Furthermore, the dopamine neurons themselves coded the reward prediction error during the differential progress of behavioural and neuronal learning. Thus learning depended crucially on the presence of a reward prediction error which was coded by the dopamine neurons, and rewards that produced reduced dopamine signals failed to support both behavioural and neuronal learning. The observed concordance between cellular and behavioural processes in dopamine neurons contributes to establishing a neurobiological basis for error-driven learning rules that are derived from behavioural analyses[2–6]. The application of formal behavioural learning rules to the study of single neurons in the mammalian brain may provide a powerful approach for studying the cellular foundations of reward-directed learning.

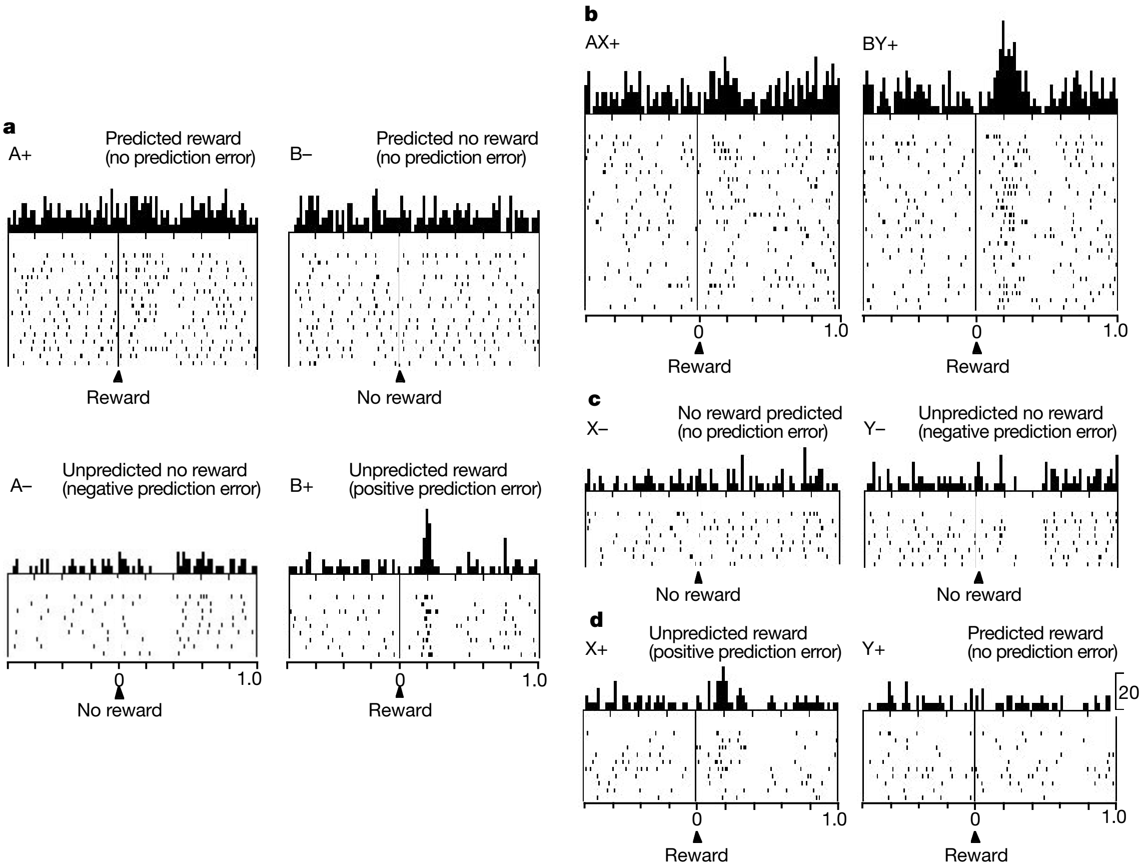

Figure 3 Dopamine prediction error response at the time of the reward in the blocking paradigm. **a**, Pretraining; **b**, during compound learning; **c**, **d**, learning tests. **a**, Lack of responses in regular trials (top); occasional test trials show neuronal depression with omitted reward in A trials and activation with surprising reward in B trials (bottom). **b**, Lack of substantial response of dopamine neuron to predicted reward in AX+ trials, but activation to surprising reward in BY+ trials. **c**, Depressant dopamine response with omitted reward in Y but not X trials. **d**, Dopamine activation to reward in X but not Y trials. Four different neurons are shown in **a**–**d**, respectively. Vertical calibration indicates 20 impulses s^{-1} (in **a**–**d**).

The design of the blocking paradigm allowed us to differentiate the influence of reward prediction errors from that of simple stimulus–reward pairings. Although the stimuli X and Y received comparable numbers of pairings with reward during compound training, only stimulus Y was associated with a reward prediction error. The behavioural reactions demonstrated that only stimulus Y was learned as a predictor of reward, whereas X was blocked from learning. These data accord with numerous behavioural demonstrations of blocking as well as contemporary learning theory[2–6,37,38] and suggest that behavioural learning in the present primate paradigm depends on reward prediction errors rather than on simply stimulus–reward pairings.

Dopamine neurons acquired stronger activations to stimulus Y than X, similar to the acquisition of behavioural reactions. Neuronal learning to stimulus X, like the behavioural reaction, was subject to blocking. The fewer, comparatively minor and biphasic responses to stimulus X were possibly due to response generalization, rather than neuronal reward prediction. All four stimuli used in our experiment induced ocular reactions. Thus the differential acquisition of dopamine responses to stimuli X and Y reflected the degree to which the stimuli became valid reward predictors, rather than differences in stimulus detection.

The dopamine activation at the time of the reward was stronger in BY than in AX trials. This response coincided with substantial behavioural and neuronal learning to stimulus Y. By contrast, pairings between stimulus X and the predicted reward in AX trials produced only a small dopamine response to the reward, no behavioural learning, and little acquisition of neuronal response to stimulus X. Whereas earlier work showed that dopamine reward responses diminished with increasing reward prediction during learning[27], the present experiments showed a differential relationship to learning. It thus appears that the differential capacity of the reward to support behavioural and neuronal learning depends on prediction errors which are signalled by the dopamine response at the time of the reward. Rewards that produce larger reward prediction errors induce stronger dopamine activations and better behavioural and neuronal learning of a reward-predicting stimulus, thereby suggesting that dopamine reward responses and behavioural and neuronal learning are correlated.

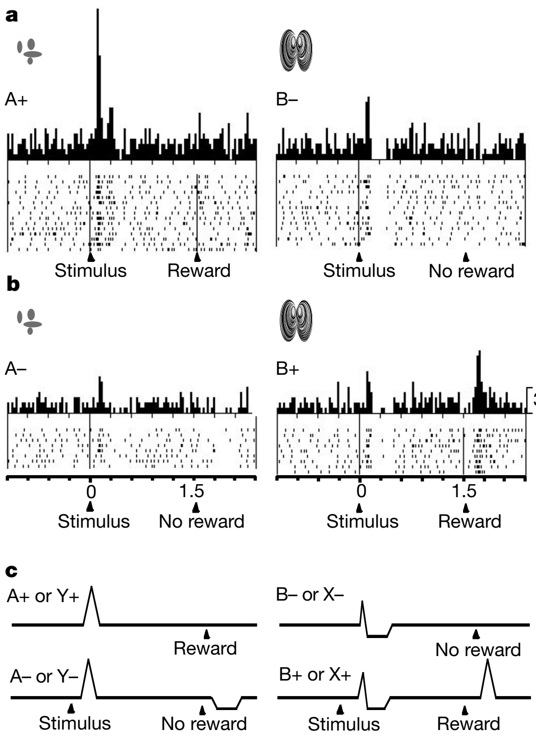

Figure 4 Dopamine responses to unrewarded stimuli may reflect stimulus generalization rather than reward prediction. **a**, Activation–depression response to unrewarded stimulus B. The absence of depression to predicted no reward in B trials indicates an absence of neuronal reward prediction. **b**, Occasional test trials in a different neuron. Activation to unpredicted reward indicates absence of reward prediction in B trials, whereas the depression at the habitual time of reward in unrewarded A trials indicates a reward prediction. Vertical calibration is 30 impulses s^{-1} for **a** and **b**. **c**, Schematic of generalizing dopamine responses to reward-predicting and non-predicting stimuli.

The present results suggest that the dopamine signal encodes the prediction error of formal theories of associative learning. These theories deploy prediction errors in two distinct ways[7]. The first assumes that the error generated on a trial has a direct impact on predictive learning in that trial[3,31]. By contrast, attentional theories argue that reward prediction errors modulate predictive learning indirectly by producing attentional learning to a stimulus that controls its subsequent associability with the reward[4,5]. Recent neurobiological research has begun to identify the brain structures involved in such attentional learning[40]. The bidirectional changes of dopamine neurons with prediction errors in A, B, X and Y trials might comply better with the $\lambda - V$ (reinforcer − predictor) error term of predictive learning than with the absolute value error term of attentional theories ($|\lambda - V|$) (ref. 7). Neurons carrying prediction errors are also found in brain structures other than the dopamine system[7], particularly in the climbing fibre projection to cerebellar Purkinje cells[11,12,14,15].

The generation and action of the hypothetical dopamine teaching signal may involve basal ganglia circuits linking the dopamine neurons with modifiable synapses in the striatum and cortex[28,35,36,41–47]. A dopamine reward-predicting teaching signal could sustain learning about extended chains of predictive stimuli leading to a primary reward by mediating the 'credit assignment' to earlier links of such chains through higher-order conditioning[16,30,48]. The interpretation of the dopamine response serving as an internal reinforcement signal accords with the finding that stimuli that are blocked from learning, like dopamine responses, are weakened in their function as conditioned reinforcers to support higher-order learning[49]. □

Methods

Two adult, male *Macaca fascicularis* monkeys were used in a classical conditioning procedure. Stimuli were presented for 1.5 s on a 13-inch computer monitor at 65 cm from the monkey's eyes (24 × 18°) (Fig. 1a, c). Each of them covered a rectangle on the screen, which had a surface area of 1,600 mm^2 with side lengths of 9.7 and 61.7 mm. We employed consecutively five sets of four structured, coloured visual stimuli (A, B, X, Y) and recorded from similar numbers of neurons in each set. Each neuron was tested with one stimulus set. Each stimulus set was used for several months and discarded when the next set started. Animals showed faster learning with increasing numbers of experienced stimulus sets ('learning set' behaviour). Stimuli were presented on a structured, coloured monitor background which was identical for each stimulus set but different between sets. In order to make the stimuli as dissimilar as possible and thus reduce potential neuronal response generalization[28], each stimulus had a distinctively different and constant form, mix of colours and position on the screen, and was not overlapped by other stimuli. Owing to their constant positions, animals were able to identify each stimulus most easily by its position when it came up, probably without even fixating it. In one set, stimuli X and Y were sounds delivered from a small loudspeaker below the computer monitor (2 kHz and white noise, respectively). The data obtained with the sounds were comparable to those obtained with the respective visual stimuli, including the differential responses to conditioned stimuli and rewards; they were therefore treated together. A constant 0.1–0.2-ml volume of fruit juice was delivered at the end of the 1.5-s stimulus duration through a spout at the animal's mouth. Interstimulus intervals varied randomly between 12 and 20 s. There was no specific action required by the animal for having reward delivered following a stimulus. In free liquid trial blocks, animals received 0.1–0.2 ml of fruit juice at irregular intervals of 12–20 s outside of any specific task. Lick responses of the animal were monitored by interruptions by the animal's tongue of an infrared light beam 4 mm below the spout. Eye positions were monitored through an infrared oculometer (Iscan).

Behavioural task

Three consecutive phases were employed. During pretraining, stimulus A was followed by liquid reward, but B went unrewarded. Stimuli A and B alternated semirandomly. During compound conditioning, stimulus X was added to the established, reward-predicting stimulus A without changing reward delivery. Control stimulus B was paired with stimulus Y and followed by reward. AX+ and BY+ trials were semirandomly intermixed with A+ and B– trials to maintain the reward and nonreward associations of stimuli A and B, respectively. In the third phase, stimuli X and Y were tested in occasional unrewarded trials in semirandom alternation and intermixed with A+, B–, AX+ and BY+ trials (ratio about 1:5) to avoid conditioning.

Electrophysiological recording

Activity from single midbrain dopamine neurons was recorded extracellularly during 20–60 min in the two animals using standard electrophysiological techniques. As described previously[22–25,50] dopamine neurons discharged polyphasic, initially negative or positive impulses with relatively long durations (1.8–5.5 ms) and low frequencies (2.0–8.5 impulses s^{-1}). Impulses contrasted with those of pars reticulata neurons of substantia nigra (70–90 impulses s^{-1} and <1.1 ms duration), a few unknown neurons discharging impulses of <1.0 ms at low rates, and neighbouring fibres (<0.4 ms duration). Only dopamine neurons activated by reward delivery in free liquid trials were tested in the present experiments (about 75–80% of the dopamine neuronal population). Neuronal activations were compared against an 800-ms control period preceding the first task event by using the Wilcoxon test in at least seven trials, with a constant time window of 70–220 ms following the conditioned stimuli and 90–220 ms following reward. These time windows comprised 80% of onset and offset times of statistically significant increases ($P < 0.01$). Neuronal depressions were assessed with the Wilcoxon test during individual time windows. Responses were compared in individual neurons with the two-tailed Mann–Whitney U test ($P < 0.05$) between different stimuli, using impulse counts in single trials. Recording sites of dopamine neurons sampled from groups A8, A9 and A10 were marked with small electrolytic lesions and reconstructed from 40-μm-thick, tyrosine-hydroxylase-immunoreacted or cresyl-violet-stained, stereotaxically oriented, coronal brain sections (Fig. 1e). Experimental protocols conformed to the Swiss Animal Protection Law and were supervised by the Fribourg Cantonal Veterinary Office.

1. Thorndike, E. L. *Animal Intelligence: Experimental Studies* (MacMillan, New York, 1911).
2. Pavlov, I. P. *Conditional Reflexes* (Oxford Univ. Press, London, 1927).
3. Rescorla, R. A. & Wagner, A. R. in *Classical Conditioning II: Current Research and Theory* (eds Black, A. H. & Prokasy, W. F.) 64–99 (Appleton Century Crofts, New York, 1972).
4. Mackintosh, N. J. A theory of attention: Variations in the associability of stimulus with reinforcement. *Psychol. Rev.* **82,** 276–298 (1975).
5. Pearce, J. M. & Hall, G. A. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* **87,** 532–552 (1980).
6. Dickinson, A. *Contemporary Animal Learning Theory* (Cambridge Univ. Press, Cambridge, 1980).
7. Schultz, W. & Dickinson, A. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* **23,** 473–500 (2000).
8. Widrow, G. & Hoff, M. E. Adaptive switching circuits. *IRE Western Electron. Show Convention, Convention Record* Part 4, 96–104 (1960).
9. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng. Trans. ASME* **82,** 35–45 (1960).
10. Widrow, G. & Sterns, S. D. *Adaptive Signal Processing* (Prentice-Hall, Englewood Cliffs, 1985).
11. Marr, D. A theory of cerebellar cortex. *J. Physiol.* **202,** 437–470 (1969).
12. Ito, M. Long-term depression. *Ann. Rev. Neurosci.* **12,** 85–102 (1989).
13. Thompson, R. F. & Gluck, M. A. in *Perspectives on Cognitive Neuroscience* (eds Lister, R. G. & Weingartner, H.) 25–45 (Oxford Univ. Press, New York, 1991).
14. Kawato, M. & Gomi, H. The cerebellum and VOR/OKR learning models. *Trends Neurosci.* **15,** 445–453 (1992).
15. Kim, J. J., Krupa, D. J. & Thompson, R. F. Inhibitory cerebello-olivary projections and blocking effect in classical conditioning. *Science* **279,** 570–573 (1998).
16. Sutton, R. S. & Barto, A. G. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.* **88,** 135–170 (1981).
17. Sutton, R. S. & Barto, A. G. *Reinforcement Learning* (MIT Press, Cambridge, Massachusetts, 1998).
18. Fibiger, H. C. & Phillips, A. G. in *Handbook of Physiology—The Nervous System IV* (ed. Bloom, F. E.) 647–675 (Williams and Wilkins, Baltimore, 1986).
19. Wise, R. A. & Hoffman, C. D. Localization of drug reward mechanisms by intracranial injections. *Synapse* **10,** 247–263 (1992).
20. Robinson, T. E. & Berridge, K. C. The neural basis for drug craving: an incentive-sensitization theory of addiction. *Brain Res. Rev.* **18,** 247–291 (1993).
21. Robbins, T. W. & Everitt, B. J. Neurobehavioural mechanisms of reward and motivation. *Curr. Opin. Neurobiol.* **6,** 228–236 (1996).
22. Romo, R. & Schultz, W. Dopamine neurons of the monkey midbrain: Contingencies of responses to active touch during self-initiated arm movements. *J. Neurophysiol.* **63,** 592–606 (1990).
23. Schultz, W. & Romo, R. Dopamine neurons of the monkey midbrain: Contingencies of responses to stimuli eliciting immediate behavioural reactions. *J. Neurophysiol.* **63,** 607–624 (1990).
24. Ljungberg, T., Apicella, P. & Schultz, W. Responses of monkey dopamine neurons during learning of behavioural reactions. *J. Neurophysiol.* **67,** 145–163 (1992).
25. Schultz, W., Apicella, P. & Ljungberg, T. 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).
26. Schultz, W., Dayan, P. & Montague, R. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
27. Hollerman, J. R. & Schultz, W. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neurosci.* **1,** 304–309 (1998).
28. Schultz, W. Predictive reward signal of dopamine neurons. *J. Neurophysiol.* **80,** 1–27 (1998).
29. Salamone, J. D. The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav. Brain Res.* **61,** 117–133 (1994).
30. Horvitz, J. C. Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* **96,** 651–656 (2000).
31. Sutton, R. S. & Barto, A. G. in *Learning and Computational Neuroscience: Foundations of Adaptive Networks* (eds Gabriel, M. & Moore, J.) 497–537 (MIT Press, Cambridge, Massachusetts, 1990).
32. Mackintosh, N. J. *Conditioning and Associative Learning* (Oxford Univ. Press, New York, 1983).
33. Friston, K. J., Tononi, G., Reeke, G. N. Jr, Sporns, O. & Edelman, G. M. Value-dependent selection in the brain: simulation in a synthetic neural model. *Neuroscience* **59,** 229–243 (1994).
34. Montague, P. R., Dayan, P. & Sejnowski, T. J. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* **16,** 1936–1947 (1996).
35. Houk, J. C., Adams, J. L. & Barto, A. G. in *Models of Information Processing in the Basal Ganglia* (eds Houk, J. C., Davis, J. L. & Beiser, D. G.) 249–270 (MIT Press, Cambridge, Massachusetts, 1995).
36. Suri, R. & Schultz, W. A neural network with dopamine-like reinforcement signal that learns a spatial delayed response task. *Neuroscience* **91,** 871–890 (1999).
37. Kamin, L. J. in *Fundamental Issues in Instrumental Learning* (eds Mackintosh, N. J. & Honig, W. K.) 42–64 (Dalhousie Univ. Press, Dalhousie, 1969).
38. Martin, I. & Levey, A. B. Blocking observed in human eyelid conditioning. *Q. J. Exp. Psychol. B* **43,** 233–255 (1991).
39. Dickinson, A. Causal learning: An associative analysis. *Q. J. Exp. Psychol. B* **54,** 3–25 (2001).
40. Holland, P. C. Brain mechanisms for changes in processing of conditioned stimuli in Pavlovian conditioning: Implications for behavioural theory. *Anim. Learn. Behav.* **25,** 373–399 (1997).
41. Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B. & Bernardi, G. Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. *J. Neurosci.* **12,** 4224–4233 (1992).
42. Garcia-Munoz, M., Young, S. J. & Groves, P. Presynaptic long-term changes in excitability of the corticostriatal pathway. *NeuroReport* **3,** 357–360 (1992).
43. Wickens, J. R., Begg, A. J. & Arbuthnott, G. W. Dopamine reverses the depression of rat corticostriatal synapses which normally follows high-frequency stimulation of cortex in vitro. *Neuroscience* **70,** 1–5 (1996).
44. Calabresi, P. *et al*. Abnormal synaptic plasticity in the striatum of mice lacking dopamine D2 receptors. *J. Neurosci.* **17,** 4536–4544 (1997).
45. Otani, S., Blond, O., Desce, J. M. & Crepel, F. Dopamine facilitates long-term depression of glutamatergic transmission in rat prefrontal cortex. *Neuroscience* **85,** 669–676 (1998).
46. Otani, S., Auclair, N., Desce, J., Roisin, M. P. & Crepel, F. *J. Neurosci.* **19,** 9788–9802 (1999).
47. Centonze, D. *et al*. Unilateral dopamine denervation blocks corticostriatal LTP. *J. Neurophysiol.* **82,** 3575–3579 (1999).
48. Minsky, M. L. Steps toward artificial intelligence. *Proc. Inst. Radio Engineers* **49,** 8–30 (1961).
49. Rauhut, A. S., McPhee, J. E. & Ayres, J. J. B. Blocked and overshadowed stimuli are weakened in their ability to serve as blockers and second-order reinforcers in Pavlovian fear conditioning. *J. Exp. Psychol: Anim. Behav. Process* **25,** 45–67 (1999).
50. Schultz, W. & Romo, R. Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* **57,** 201–217 (1987).

Acknowledgements

We thank B. Aebischer, J. Corpataux, A. Gaillard, B. Morandi, A. Pisani and F. Tinguely for expert technical assistance. The study was supported by the Swiss NSF, the European Union (Human Capital and Mobility, and Biomed 2 programmes), the James S. McDonnell Foundation and the British Council.

Correspondence and requests for materials should be addressed to W.S. (e-mail: Wolfram.Schultz@unifr.ch).