



HAL
open science

Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context

Paul Apicella, Marc Deffains, Sabrina Ravel, Eric Legallet

► To cite this version:

Paul Apicella, Marc Deffains, Sabrina Ravel, Eric Legallet. Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context. *European Journal of Neuroscience*, 2009, 30 (3), pp.515-526. 10.1111/j.1460-9568.2009.06872.x . hal-03838714

HAL Id: hal-03838714

<https://hal.science/hal-03838714v1>

Submitted on 9 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

COGNITIVE NEUROSCIENCE

Tonically active neurons in the striatum differentiate between delivery and omission of expected reward in a probabilistic task context



Paul Apicella, Marc Deffains, Sabrina Ravel and Eric Legallet

Laboratoire de Neurobiologie de la Cognition, Université de Provence – CNRS, 3 place Victor Hugo, 13331 Marseille Cedex 3, France

Keywords: error detection, learning, prediction, primate, putamen

Abstract

Tonically active neurons (TANs) in the primate striatum are responsive to rewarding stimuli and they are thought to be involved in the storage of stimulus–reward associations or habits. However, it is unclear whether these neurons may signal the difference between the prediction of reward and its actual outcome as a possible neuronal correlate of reward prediction errors at the striatal level. To address this question, we studied the activity of TANs from three monkeys trained in a classical conditioning task in which a liquid reward was preceded by a visual stimulus and reward probability was systematically varied between blocks of trials. The monkeys' ability to discriminate the conditions according to probability was assessed by monitoring their mouth movements during the stimulus–reward interval. We found that the typical TAN pause responses to the delivery of reward were markedly enhanced as the probability of reward decreased, whereas responses to the predictive stimulus were somewhat stronger for high reward probability. In addition, TAN responses to the omission of reward consisted of either decreases or increases in activity that became stronger with increasing reward probability. It therefore appears that one group of neurons differentially responded to reward delivery and reward omission with changes in activity into opposite directions, while another group responded in the same direction. These data indicate that only a subset of TANs could detect the extent to which reward occurs differently than predicted, thus contributing to the encoding of positive and negative reward prediction errors that is relevant to reinforcement learning.

Introduction

Several lines of evidence have implicated the striatum, especially its dorsal part, in procedural learning processes, such as habit formation and acquisition of motor skills (Graybiel, 1998; Salmon & Butters, 1995; Packard & Knowlton, 2002). Single-neuron recordings in the striatum of awake animals have shown that phasically active neurons, which correspond to medium spiny projection neurons, display changes in firing during the acquisition of action–outcome relations and their retention after extensive training (Jog *et al.*, 1999; Barnes *et al.*, 2005; Pasupathy & Miller, 2005; Brasted & Wise, 2004; Tang *et al.*, 2007; Tremblay *et al.*, 1998), whereas tonically active neurons (TANs), presumed to be cholinergic interneurons, may contribute to the storage of stimulus–reward associations or habits (Aosaki *et al.*, 1994; Apicella, 2002). The response properties of midbrain dopamine (DA) neurons that innervate the striatum are thought to reflect their capacity to carry information about differences between predictions and outcomes, which are called prediction errors and are crucial for reinforcement learning (Waelti *et al.*, 2001; Morris *et al.*, 2004; Bayer & Glimcher, 2005; Roesch *et al.*, 2007). The detection of the extent to

which a rewarding outcome is better or worse than expected corresponds to, respectively, a positive or negative reward prediction error, which have correlates in phasic changes in activity of DA neurons. Because the sensitivity of TANs to rewarding events shares some features with that described for DA neurons (Apicella, 2007), it is conceivable that TANs contribute to the processing of prediction error signals. Indeed, we previously found that TANs respond more frequently to unpredicted rewards than to predicted ones (Apicella *et al.*, 1997; Ravel *et al.*, 2001), indicating that they may encode at least positive reward prediction errors. Recently, Joshua *et al.* (2008) found evidence, in a probabilistic conditioning task, that TANs respond to reward delivery and reward omission, and both responses can be modulated by reward probability. However, the reported changes in TAN activity essentially consisted of decreases in firing after both reward and no reward, suggesting that they do not meet criteria for prediction error signaling similar to those described for DA neurons (Niv & Schoenbaum, 2008). It is possible that the coding of a full reward prediction error by these neurons depends on the specific characteristics of the learning situation in which animals experienced the stimulus–outcome associations, as it is known that TAN responses are most prominent under situations in which predictions are mediated by extensive practice with the same context (Sardo *et al.*, 2000; Ravel *et al.*, 2001). In the present study, the capacity of TANs to carry

Correspondence: Dr P. Apicella, as above.
E-mail: paul.apicella@univ-provence.fr

Received 30 March 2009, revised 15 June 2009, accepted 29 June 2009

prediction error signals was further examined in a classical conditioning task in which the probability of reward (P_r) was changed between blocks of trials. Under this condition, reward predictions are not explicitly driven by external stimuli, but rather are internally generated through repeated experience of stimulus–reward pairings, and we found that a subset of TANs appear to encode positive and negative errors in prediction of reward whereas other TANs respond to infrequent events irrespective of their valence.

Materials and methods

Animals

Three adult male macaque monkeys (*Macaca fascicularis*), monkeys 1, 2, and 3, weighing between 6 and 10 kg, were used in the present experiments. The monkeys were extensively trained in a classical conditioning task with a visual stimulus indicating the delivery of a liquid reward at the end of a constant delay. The monkeys had several months of experience with the same visual stimulus always preceding the delivery of reward by an interval of 1 s. All experiments were in compliance with the National Institutes of Health's *Guide for the Care and Use of Laboratory Animals* and the French laws on animal experimentation, under the supervision of a permanent veterinary (Dr. I. Balansard) in order to check that they were in agreement with the guidelines for ethical treatment of animals.

Behavioral task

The behavioral apparatus was similar to that described in Apicella *et al.* (1997). Briefly, monkeys were seated in a restraining box and faced a panel 30 cm in front of them. A red light-emitting diode (LED) was located at the center of the panel, at the eye level of the animal. A tube positioned in front of the monkey's mouth dispensed small amounts of fruit juice (0.3 mL) as a reward. The tube was equipped with force transducers with which the contact between the lips and tongue and the spout was monitored. The task involved pairing a visual stimulus with reward. We used a trace conditioning procedure in which the central LED was illuminated for 0.3 s (monkey 1) or 0.5 s (monkeys 2 and 3) and was followed 1 s after onset by the delivery of reward (Fig. 1). The interval between the delivery of reward on one trial and the presentation of the visual stimulus on the next trial varied between 5

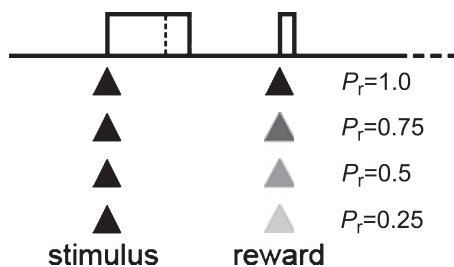


FIG. 1. Probabilistic reward structure of the classical conditioning procedure. We used a classical conditioning task in which a visual stimulus of 0.3–0.5 s duration was followed by a liquid reward. The interval between the onset of the visual stimulus and the delivery of reward was kept constant at 1 s and intertrial intervals varied between 5 and 8 s. A block presentation design was used, with probabilities of reward varying between blocks (1.0, 0.75, 0.5 or 0.25) and each block consisting of 40–70 trials. The same visual stimulus preceded the delivery of reward in the different probability conditions and no external cues were given to indicate transition from one block to the other. This testing procedure was based on the assumption that monkeys could estimate the degree of reward probability available within each block through experience, thus corresponding to internally attributed types of prediction of outcome.

and 8 s. The probability of reward was manipulated under this condition with probabilities varying from one block of trials to the other ($P_r = 1.0, 0.75, 0.5$ and 0.25) and each block consisting of 40–70 trials, so that the monkeys could estimate the degree of probability available in each block across trials.

Surgery

After several weeks of experience in the classical conditioning task when reward was delivered with a probability of $P_r = 1.0$, animals were implanted, under pentobarbital sodium anaesthesia (Sanofi, Libourne, France, 35 mg/kg i.v.) and sterile surgical conditions, with a recording chamber stereotaxically positioned above a craniotomy centered on the anterior commissure. This location was chosen to permit vertical access with microelectrodes to the putamen, mostly its posterior part. Two stainless steel cylinders were also fixed to the skull with surgical screws and dental acrylic for subsequent head fixation during recording sessions. Antibiotics (Ampicillin; Bristol-Myer Squibb, Paris, France, 17 mg/kg every 12 h) and analgesics (Tolfe-dine®; Vetoquinol, Lure, France, 2 mg/kg) were administered on the day of surgery and for the following 5 days. In the first two weeks after surgery, and before neuronal recording started, the monkeys were gradually habituated to accept head restraint.

Neuronal recordings

We used custom-made glass-coated tungsten electrodes for recording extracellular activity of single neurons. They were passed inside a guide canula (0.6 mm outer diameter) at the beginning of each recording session. After penetration of the dura, the electrode was advanced toward the striatum with a manual hydraulic microdrive (MO95; Narishige, Tokyo, Japan) until the activity of a neuron was isolated. The signal from neuronal activity was amplified 5000 times, filtered at 0.3–1.5 kHz and converted to digital pulses through a window discriminator (NeuroLog; Digitimer, Hertfordshire, UK). Continuous monitoring of the spike waveform on a digital oscilloscope during recording allowed us to check spikes of isolated neurons. Presentation of visual stimuli, delivery of reward, and digital pulses from neuronal activity were controlled by a computer using a custom-made software (E. Legallet). The task relationships of neuronal discharges were assessed on-line in the forms of rasters aligned on each task event. Single neurons were generally isolated while the monkey performed the task with reward at $P_r = 1.0$. We then tested them in separate blocks of trials using lower probability levels, the change in condition not being indicated by any explicit cues. The order of the $P_r < 1.0$ conditions was counterbalanced across sessions to avoid order effects. Mouth contacts with the spout were digitized at 100 Hz and stored during each block of trials, concomitantly with neuronal activity, for off-line quantitative analysis of the oral behavior.

Data analysis

We evaluated the effects of changes in reward probability on behavior by assessing the timing features of the mouth movements that monkeys performed in the different conditions. We measured the frequency of licking movements occurring during the 1-s period between the onset of the visual stimulus and the receipt of liquid and the latency of these movements from onset of the visual stimulus to onset of licking. The Mann–Whitney *U*-test served to compare lick latencies between conditions. Correlations between the latency of

licking movements and the probability of reward were determined with Spearman's rank correlation coefficient.

We analyzed neuronal activity by detecting changes in TAN firing on the basis of a Wilcoxon signed-rank test (Apicella et al., 1997). Only neurons with statistically significant changes against control activity were considered responsive. The baseline activity was calculated during the 0.5 s before the presentation of the visual stimulus. A test window of 100 ms duration was moved in steps of 10 ms, starting at the onset of the visual stimulus or the delivery of reward. The onset of a response was taken to be the beginning of the first of five consecutive steps showing a significant difference ($P < 0.05$) from the baseline activity. The end of a response was defined as the first of five consecutive steps in which activity had returned to control levels. The magnitude of the response was determined individually for each responding neuron and is expressed as a percentage below or above baseline activity. Differences in fractions of responding neurons among the conditions were tested with the chi-square test. The latency, duration and magnitude of TAN responses were compared among the four probability conditions using one-way ANOVA, with probability level as factor. Response parameters were also compared between rewarded and unrewarded trials with the Mann–Whitney U -test. A linear regression analysis was used to analyze the relationship between magnitudes of changes in TAN activity and the different reward probabilities.

To give a description of the properties of the population of TANs sampled, we calculated the ensemble average activity of all neurons recorded in each probability condition. For each neuron, a normalized perievent time histogram was obtained by dividing the content of each bin by the number of trials, and the population histogram was constructed by averaging all normalized histograms. We also used a time window analysis to statistically assess and compare changes in the average population response between conditions. First, latency and duration of these changes were determined for each population histogram. This analysis was performed in 10-ms bins to identify whether and when the population significantly changed its activity. The onset time of a change was defined as the first bin from which a significant difference (paired t -test, $P < 0.05$) continued consecutively for at least three bins (i.e., 30 ms). End time was defined in a similar fashion for the return to control. Then standard time windows were defined on the basis of onset and end times for distinct components of the population response. The magnitude of activity change was determined in every time window, by comparing the number of spikes between the time window (normalized for durations of time windows) and a control period of 100 ms preceding the visual stimulus. The magnitudes of activity changes obtained from this standard time window method were compared with one-way ANOVA using probability level as a factor.

Histology

After completion of the experiments, over a period of 8–12 months, small electrolytic lesions were made in the striatum of monkeys 1 and 2 by passing negative currents through the microelectrode (20 μ A for 20–30 s). These marking lesions were used as landmarks for the reconstruction of recording sites. Animals were given a lethal dose of sodium pentobarbital and perfused transcardially with isotonic saline followed by a fixative (4% paraformaldehyde in pH 7.4 phosphate buffer). The brains were removed, and frozen coronal sections were cut at a thickness of 50 μ m and stained with Cresyl violet. The recording sites were not localized in monkey 3, which is still used in recording experiments, but we identified the recording region as being mainly located in the posterior putamen on the basis of systematic

mapping of the striatum and adjacent structures (globus pallidus) with 1-mm-spaced electrode penetrations.

Results

Behavioral data

Examples of mouth movement patterns recorded in the different probability conditions are shown in Fig. 2A. Mouth movements starting prior to the delivery of reward remained relatively stable in $P_r = 1.0$ and 0.75 conditions, whereas their timing became more variable in the $P_r = 0.5$ and 0.25 conditions, indicating that the visual stimulus was a less reliable predictor of reward delivery. All three monkeys produced anticipatory licking movements in $> 95\%$ of the trials, except when the reward was delivered at the lowest probability (Fig. 2B). There were also differences in mean lick latencies between the various reward probability conditions. As seen in Fig. 2B, animals showed significantly shorter latencies in the $P_r = 1.0$ condition than in the other three conditions (Mann–Whitney U -test, $P < 0.01$), with the exception of monkey 3 in which the difference between the latencies at $P_r = 1.0$ and 0.75 was not significant ($P > 0.05$). We analyzed, separately for the three monkeys, the correlation between the latency of licking movements and the probability of reward and found that this latency decreased linearly as probability decreased ($r = 0.288, 0.652$ and 0.390 in monkeys 1, 2, and 3, respectively; Spearman rank-correlation test; $P < 0.01$). It therefore appears that the monkeys reacted differently to the onset of the visual stimulus, indicating that they were able to distinguish between the probabilities of reward delivery available in the different conditions.

Neuronal data

A total of 56 electrode penetrations (22, 21 and 13, in monkeys 1, 2 and 3, respectively) were made in the striatum. We tested a total of 85 TANs (monkey 1, 26; monkey 2, 31; monkey 3, 28) identified on the basis of well-established electrophysiological characteristics (Kimura et al., 1984; Aosaki et al., 1994; Apicella et al., 1997). Their mean \pm SD firing rate was 6.3 ± 1.6 spikes/s, $n = 85$. In agreement with previous studies, most neurons showed a brief decrease in activity after the presentation of the visual stimulus and/or the delivery of reward. This depression of the tonic firing is usually referred to as the TAN pause response. A rebound activation often occurred immediately after the pause, this biphasic response pattern being similar to that previously reported (Kimura et al., 1984; Aosaki et al., 1994; Apicella et al., 1997; Morris et al., 2004; Yamada et al., 2004). We also found a few neurons that displayed a brief activation before the pause in firing. In addition, we observed TANs which only displayed an increase in firing occurring at a relatively long latency after the time of expected reward delivery. These late activations appear similar to those occasionally observed following task events (Yamada et al., 2004).

Responses to the predictive signal and to reward

The responsiveness of TANs was tested in the classical conditioning task when the probability of reward was changed between blocks. As shown in Fig. 3A, the proportion of TANs displaying a pause response to the visual stimulus presented in the different probability conditions did not vary significantly (χ^2 with d.f. = 3, i.e. $\chi^2_3 = 1.33$, $P > 0.05$), whereas the proportion of neurons showing a pause response to reward was significantly lower at $P_r = 1.0$, as compared

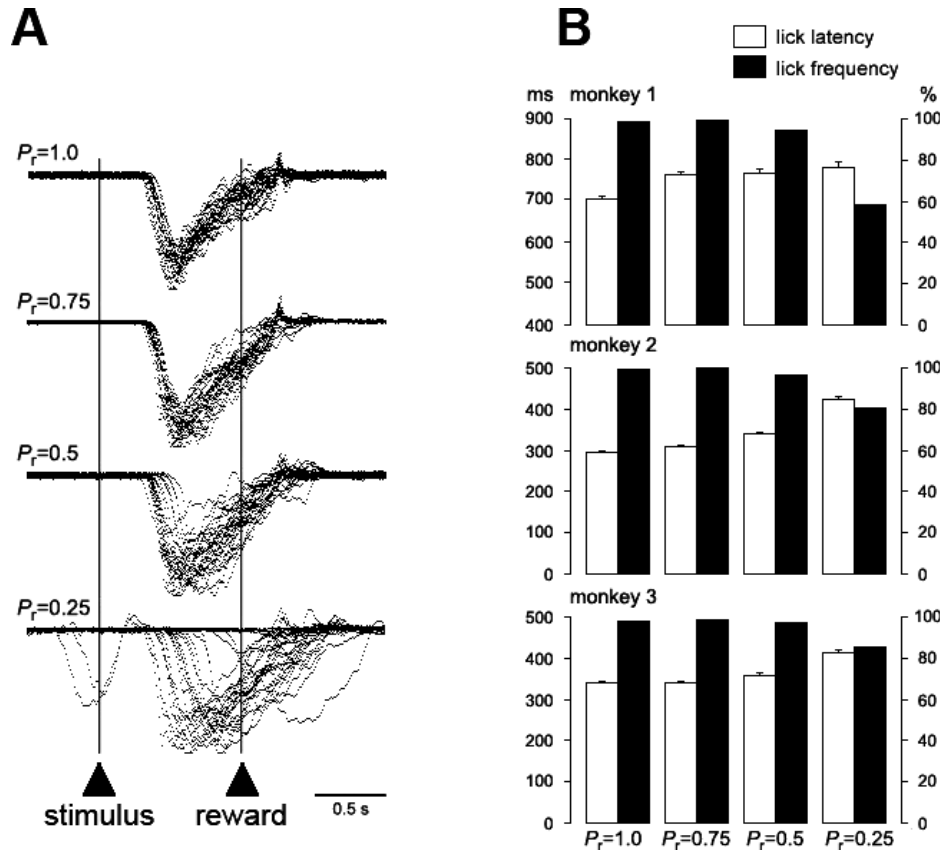


FIG. 2. Licking behavior at different reward probabilities. (A) Timing characteristics of the mouth movements. At each probability level, superimposed traces of mouth movement records from monkey 3 are shown from 30 to 40 consecutive trials, aligned on the onset of the visual stimulus. The change of probability condition occurred over four successive blocks of trials. (B) Frequencies and latencies of anticipatory licking movements for each probability condition. Frequencies denote the occurrence of licks in percentage of trials and values of latencies are means (\pm SEM). Data were obtained from 200 to 250 trials at each probability level in each monkey.

to the other three probability levels ($\chi^2_3 = 30.78$, $P < 0.01$). The fractions of reward responses were not significantly different ($P > 0.05$) when comparing $P_r = 0.75$, 0.5 and 0.25 conditions. Most of the TAN pause responses to either the visual stimulus or reward were followed by a rebound activation, whereas a brief initial activation only rarely occurred before the pause. No differences were observed in the fraction of rebound activations following the pause response to the visual stimulus across the different conditions ($\chi^2_3 = 6.19$, $P > 0.05$), whereas the fraction of TANs showing a rebound activation after the pause response to reward was significantly increased as the probability of reward decreased ($\chi^2_3 = 22.01$, $P < 0.01$). Figure 3B shows a representative example of a TAN displaying a more pronounced response to reward as probability decreased, whereas the response to the preceding visual stimulus did not vary markedly. The effects of reward probability were further assessed by comparing the magnitudes of each component of the TAN response, i.e., pause and rebound activation, at the four probability levels (Fig. 4). The magnitudes of pauses ($F_{3,231} = 3.95$, $P < 0.01$, one-way ANOVA) and rebound activations ($F_{3,185} = 5.23$, $P < 0.01$) after the visual stimulus were significantly higher at $P_r = 0.75$ than in the other conditions. On the other hand, magnitudes of pauses ($F_{3,176} = 15.57$, $P < 0.01$) and rebound activations ($F_{3,145} = 4.77$, $P < 0.01$) following reward delivery were significantly higher at $P_r = 0.25$ than in the other conditions. There were no other significant differences in the magnitude of reward responses among conditions. We then conducted a linear regression analysis of the magnitudes of neuronal responses to reward in

relation to the four probability levels. This was done for each component of the TAN response. The correlation was negative and significant for the pause ($r = -0.433$, $n = 180$, $P < 0.01$) and the rebound activation ($r = -0.285$, $n = 149$, $P < 0.01$), indicating that response magnitude increased linearly with decreasing reward probability. In contrast, the magnitude of the pause response to the visual stimulus showed no consistent relation to the probability of reward ($r = 0.116$, $n = 235$, $P > 0.05$) and this was also the case for the rebound activation ($r = 0.070$, $n = 189$, $P > 0.05$). Finally, durations of responses to reward delivery were not significantly different when comparing conditions of different reward probabilities, their common means being 124 ± 42 ms (range 50–280 ms) and 185 ± 83 ms (range 60–500 ms) for the pause and rebound activation, respectively. Also, no significant differences were observed in the duration of TAN responses to the visual stimulus among reward probability conditions.

In summary, both components of TAN responses to reward, namely pause and following rebound activation, were more numerous and pronounced with decreasing reward probability, whereas the responsiveness of TANs to the stimulus predictive of reward was enhanced in terms of magnitude of responses with increasing reward probability.

Responses at the time of reward omission

To determine whether the absence of expected reward modulates the firing of TANs, we examined neuronal activity for unrewarded trials in

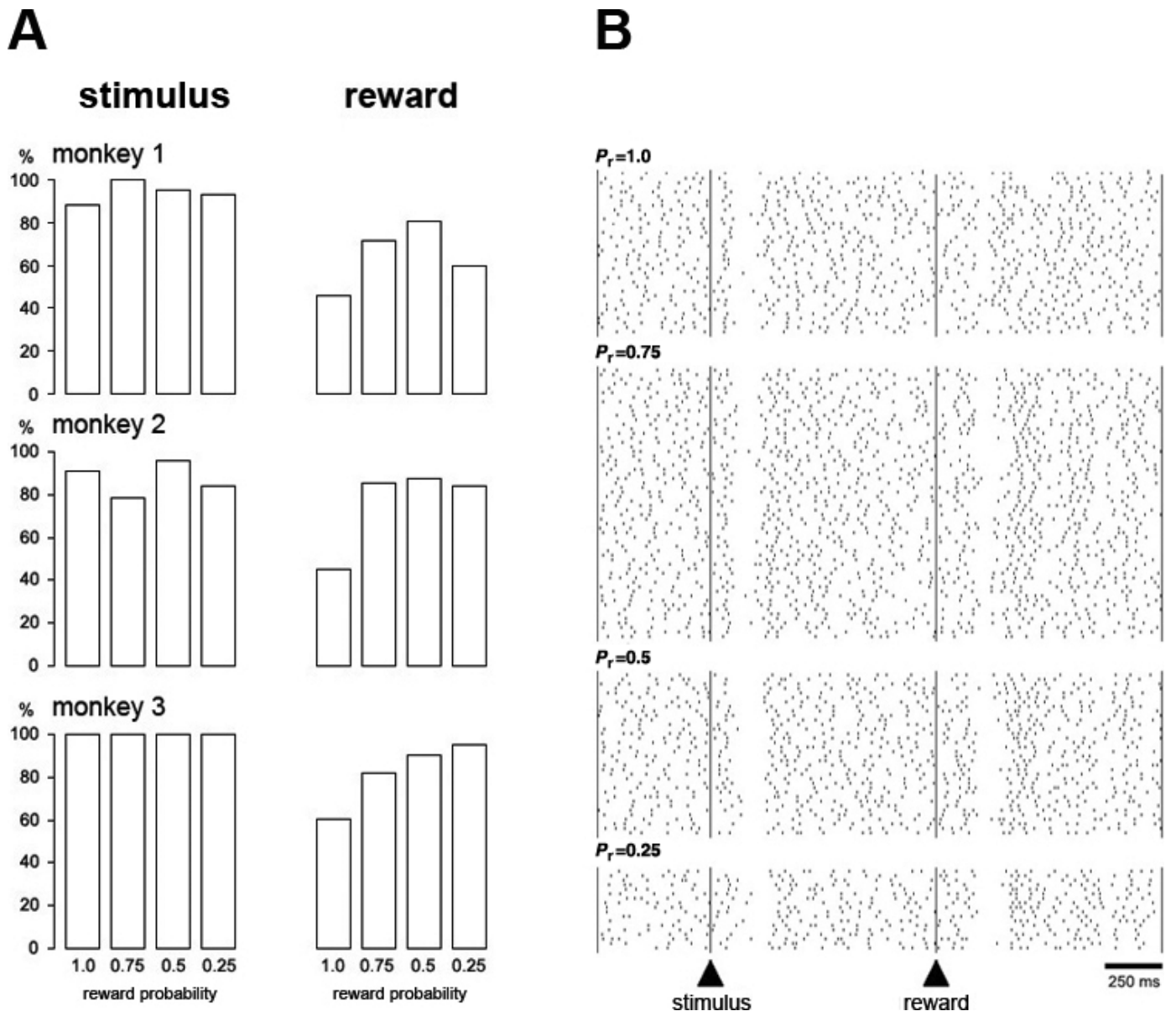


FIG. 3. Influence of reward probability on TAN responses. (A) Changes in the relative proportions of TANs responding to the visual stimulus and reward with changing reward probability. Bar plots for each monkey show proportions of responses across the four probability conditions. Left, percentage of neurons responding to the visual stimulus; right, percentage of neurons responding to reward. In all three monkeys, differences in the fraction of TANs responding to the visual stimulus were not significantly different between conditions, whereas the fraction of TANs responding to reward was lowest at $P_r = 1.0$. Numbers of neurons tested at $P_r = 1.0, 0.75, 0.5$ and 0.25 are as follows: monkey 1, $n = 26, 7, 21$ and 15 , respectively; monkey 2, $n = 33, 14, 24$ and 19 ; monkey 3, $n = 28, 22, 20$ and 21 . (B) Responses to reward influenced by probability in one TAN. The change in reward probability occurred over four successive blocks of trials and only rewarded trials are shown. Each dot indicates a neuronal impulse and each line of dots gives the neuronal activity recorded during a single trial. Dot displays are aligned on the onset of the visual stimulus and reward in each raster display. The raster in each block is shown in chronological order with the first trial at the top.

the $P_r < 1.0$ conditions. A number of TANs had detectable changes in their activity if the reward did not occur at the expected time. These changes consisted of two types of modulation: decreases (34.8, 26.1 and 16.3% at $P_r = 0.75, 0.5$ and 0.25 , respectively) or increases (30.2, 32.3 and 41.8% at $P_r = 0.75, 0.5$ and 0.25 , respectively) in activity. It is noteworthy that increases in TAN firing produced by reward omission occurred at a relatively long latency after the usual time of reward. The two neurons whose activity is shown in Fig. 5 are representative examples of changes in TAN firing after reward omission, while they exhibited the typical pause-rebound response to reward delivery. In the first example (Fig. 5A), firing rate was weakly

but significantly depressed following the omission of reward while, in the second example (Fig. 5B), firing rate began to be significantly elevated at 200 ms after the omitted reward.

Although the number of decreases in activity after reward omission tended to increase with reward probability, chi-square analyses revealed that there were no significant differences among the three probability conditions ($\chi^2_2 = 4.45, P > 0.05$). Also, the fractions of neurons showing a late activation after reward omission were not significantly different when comparing conditions of different probabilities ($\chi^2_2 = 1.76, P > 0.05$). On the other hand, the magnitude of both decreases ($F_{2,43} = 15.08, P < 0.01$) and increases ($F_{2,58} = 5.58,$

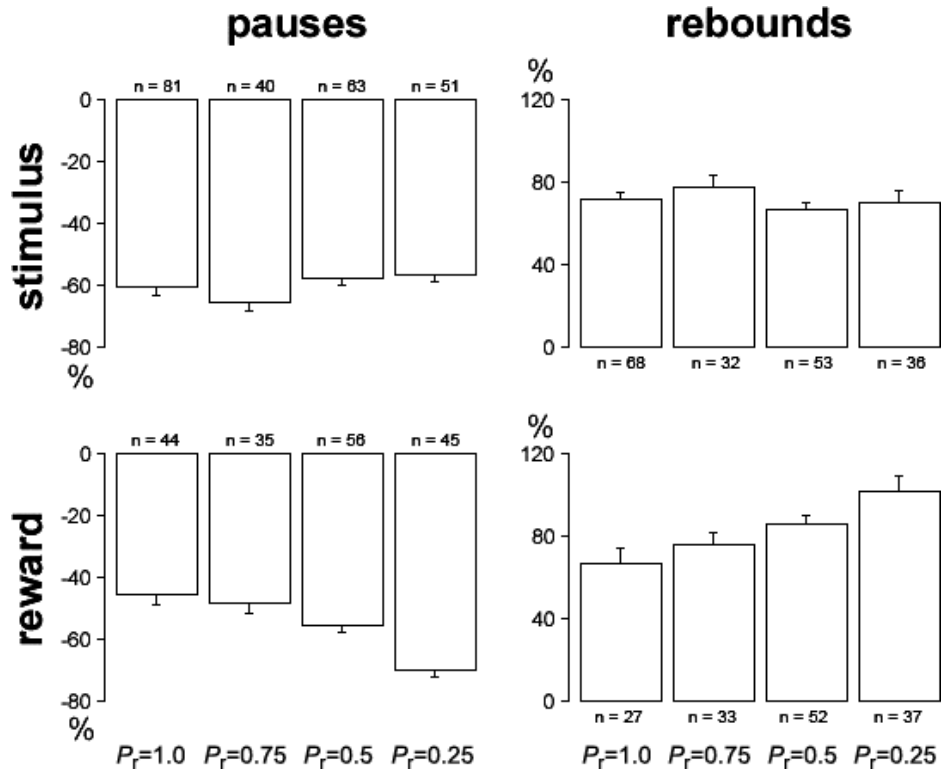


FIG. 4. Comparison of magnitudes of each component of TAN responses to the visual stimulus and reward between the four probability conditions. Magnitudes of changes are indicated as decreases (pauses) or increases (rebounds) in percentage below baseline activity. Each value contributing to n is the magnitude of response for a particular neuron at a probability level. Results are pooled for the three monkeys. Values are given as means \pm SEM.

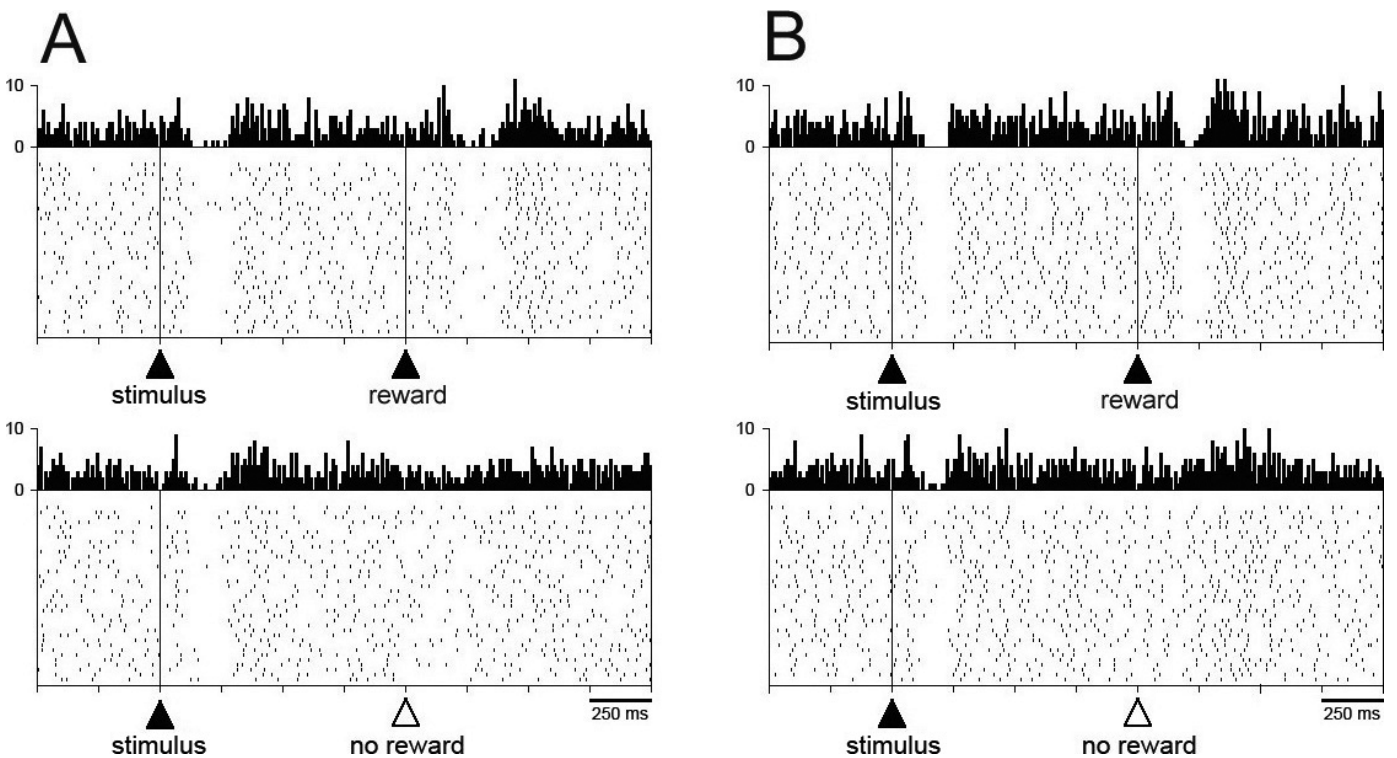


FIG. 5. Changes in activity of two TANs when predicted rewards were omitted. (A) In the first example, firing rate was decreased following the omission of reward, while (B) in the second example, firing rate was increased in the same time period (lower panels). Both neurons responded with a pause to reward delivery (upper panels). Data were collected at $P_r = 0.5$ and were separated offline according to the presence or absence of reward. Same conventions as in Fig. 2B, except that perievent time histograms are added on each raster. Histogram scale is in impulses/bin. Bin width for histograms, 10 ms.

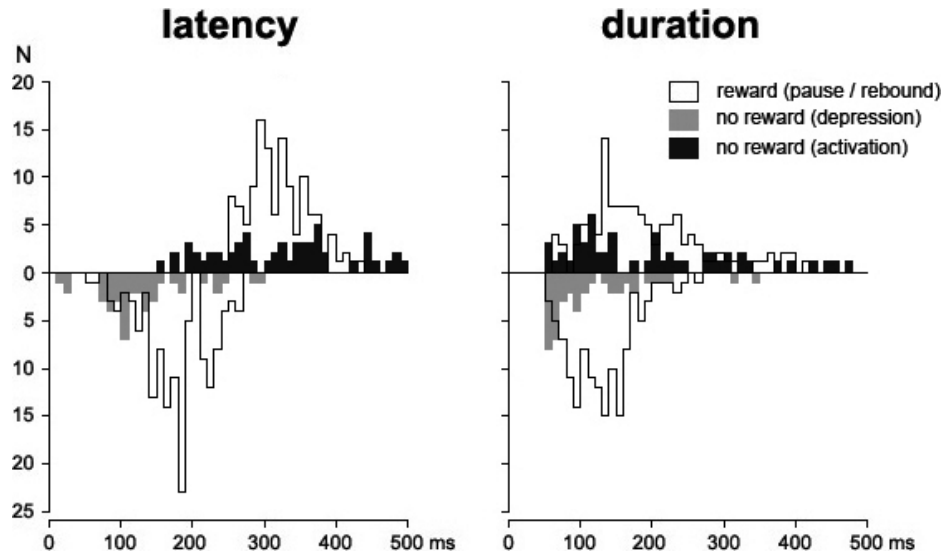


FIG. 6. Distribution of times of change in TAN activity relative to the delivery and omission of reward. Latency and duration histograms of different components of TAN modulations are presented, namely pause ($n = 136$) and rebound ($n = 122$) after reward delivery, and depression ($n = 34$) and activation ($n = 57$) after reward omission. The data of the three $P_r < 1.0$ conditions were pooled.

$P < 0.01$) in TAN firing was increased with reward probability. Magnitudes of decreases at $P_r = 0.75$ were significantly higher than those at $P_r = 0.5$ and 0.25 . A difference in the same direction was also evident between the $P_r = 0.5$ and 0.25 conditions. There were significantly higher magnitudes of increases at $P_r = 0.75$ than at $P_r = 0.25$ ($P < 0.01$) and $P_r = 0.5$ ($P < 0.05$). Linear regression analysis of the magnitudes of TAN responses to reward omission in relation to the probability of reward showed a significant positive correlation for depressions ($r = 0.633$, $n = 46$, $P < 0.01$) and a significant negative correlation for activations ($r = -0.406$, $n = 61$, $P < 0.01$). Durations of the depressions and activations after reward omission were not significantly different when comparing conditions of different reward probabilities, their common means being 111 ± 71 ms (range 50–340 ms) and 191 ± 114 ms (range 50–470 ms), respectively.

To further characterize the distinct profiles of TAN response to reward and no reward, we compared the latency and duration distributions of the activity changes for rewarded and unrewarded trials (Fig. 6). The mean latency and duration of the depression after reward delivery (173 ± 45 and 124 ± 42 ms, respectively) were significantly longer ($P < 0.01$, Mann–Whitney U) than those of the depression after reward omission (125 ± 64 and 111 ± 71 ms, respectively). On the other hand, the mean latency and duration of activation after pause response to reward delivery (314 ± 42 and 185 ± 83 ms, respectively) differed nonsignificantly ($P > 0.05$) from those of activation following reward omission (315 ± 90 and 191 ± 114 ms, respectively). It therefore appears that the pause response to reward delivery and the depression in activity after reward omission did not occur with a similar time course, whereas the increases in firing rate occurring after reward delivery and reward omission largely overlapped. It is noteworthy that the latency and duration of the activity changes following reward omission were more variable than those following reward delivery, suggesting that the TAN response to reward omission was less well temporally coordinated than the response to reward delivery.

To summarize, TANs showed decreases or increases in firing rate following the omission of reward with the magnitude of these activity changes increasing with reward probability. This suggests

that TANs can be separated into at least two distinct groups according to the characteristics of their sensitivity to the presence and absence of reward: a subset of TANs responded differently to reward and no reward, with decreases and increases in activity, respectively, while the other subset of TANs responded to both in a similar manner, namely with decreases in activity after both reward and no reward.

Comparisons between responses to reward and no reward

Among 136 TAN responses consisting of a pause in firing rate to reward delivery in the $P_r < 1.0$ conditions, 34 (25%) were associated with a depression in firing for reward omission and 57 (42%) were associated with late activations after reward omission. The remaining 45 pause responses to reward delivery (33%) occurred without any significant change after reward omission. To investigate the relationship between TAN responses and the presence or absence of reward, we examined the effect of probability level on the magnitude of the mean changes in activity during rewarded and unrewarded trials, separately for the two types of TANs responding to reward omission, i.e., with depression or activation. As can be seen in Fig. 7, the magnitude of pause responses to reward for both groups of neurons was significantly correlated with the probability of reward ($r = -0.440$ and -0.463 , $P < 0.01$). Following reward omission, the correlation between the magnitude of activity changes and reward probability was also significant for both groups ($r = 0.668$ and 0.428 , $P < 0.01$). In contrast, the magnitude of pause responses to the visual stimulus was not correlated with the probability of reward ($r = 0.177$ and 0.024 , $P > 0.05$).

We looked at the activity of the entire population of TANs recorded in the different probability conditions, regardless of the responsiveness of individual neurons. As shown in Fig. 8, left, the magnitude of the population response to reward increased with decreasing reward probability. Although an excitation preceding the TAN pause response to the visual stimulus was rarely observed in the data from single neurons, such an early response component was visible on population histograms, except at the lowest reward probability. To demonstrate

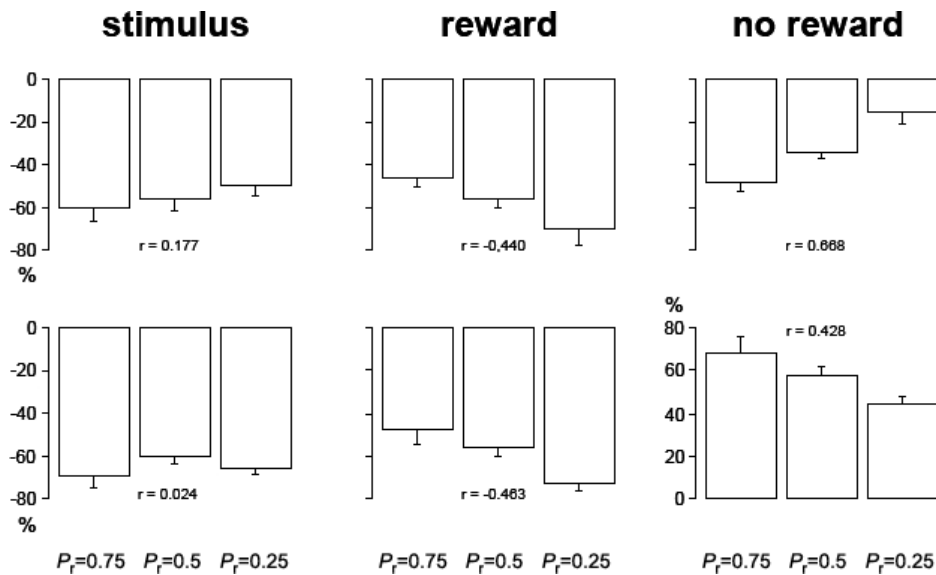


FIG. 7. Comparison of magnitudes of TAN responses to the visual stimulus, reward delivery and reward omission in relation to three different reward probability levels. Neurons were separated into two groups according to the direction of their response to reward omission, namely increase or decrease in activity. Each bar represents the average response magnitude as a function of reward probability. The data are taken from the $P_r < 1.0$ conditions. Numbers of values at $P_r = 0.75, 0.5$ and 0.25 for neurons showing a decrease in activity after reward omission are as follows: $n = 12, 16$ and 5 , respectively (top); for neurons showing an increase in activity after reward omission, $n = 11, 22$ and 22 (bottom). Results are pooled for the three monkeys. Values are given as means \pm SEM.

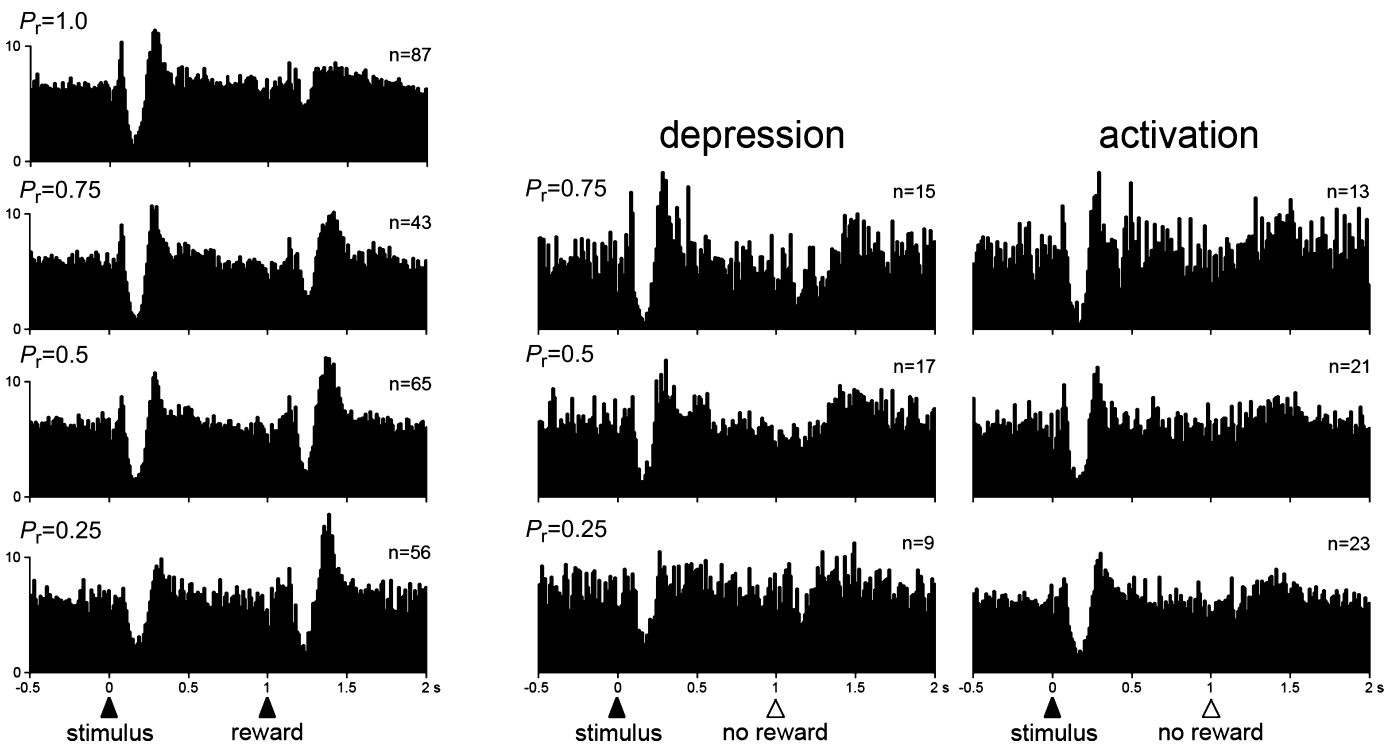


FIG. 8. Population responses of TANs to task events occurring in rewarded and unrewarded trials at different levels of probability. Population histograms with reward probabilities ranging from 1.0 (top) to 0.25 (bottom) included neurons of all three monkeys. Left, histograms obtained from all recorded TANs in rewarded trials. Middle and right, histograms shown separately for TANs showing decreases (depression) or increases (activation) in firing rate after the omission of reward in unrewarded trials. Vertical scale denotes impulses.

this quantitatively, we rated the magnitude of changes during a period spanning 50–80 ms after visual stimulus onset, which we selected on the basis of an analysis of population response latency and duration (see Materials and methods). We found that response magnitude in this

time window was significantly higher at $P_r = 1.0$ ($P < 0.05$, one-way ANOVA followed by Fisher's test) than at $P_r = 0.25$. We performed the same analysis for subsequent components of the response to the visual stimulus, using time windows specifically determined for each

response component (pause, 100–210 ms; rebound, 230–360 ms), and found that the magnitudes of pause responses at $P_r = 0.75$ and 0.5 were significantly higher than magnitudes at $P_r = 0.25$ ($P < 0.01$). A difference in the same direction was found between the $P_r = 1.0$ and 0.25 conditions ($P < 0.05$). Magnitudes of rebound activations following the pause were also significantly higher at $P_r = 1.0$ than at $P_r = 0.25$ ($P < 0.01$) and $P_r = 0.5$ ($P < 0.05$). It therefore appeared that the sensitivity of TANs to the conditioned stimulus, with respect to reward probability, was more evident at the level of the population average than at the level of individual neurons, with increased neuronal responsiveness for high reward probability. The same analysis for the population response to reward (time windows for pause and rebound following reward delivery were 100–210 and 230–380 ms, respectively) revealed that the magnitudes of both pause and rebound were significantly stronger at $P_r = 0.25$, as compared to the other probability levels ($P < 0.01$), just as described in the single-neuron data, without other significant differences in the magnitudes of responses among probability conditions. This indicates that TANs were markedly responsive to reward at low probabilities, this effect being demonstrated at the level of both the population and the individual neuron responses.

We also examined the average activity of all TANs recorded in the $P_r < 1.0$ conditions and found that the modulation of activity following the omission of reward was less apparent than that following reward delivery. This may result at least in part from the combination of decreases and increases in firing demonstrated at the single neuron level that would attenuate the response of the whole population. The population histograms were then constructed separately for TANs showing decreases or increases in their firing rate (Fig. 8, middle and right) and we assessed changes in the average activity of these two samples of TANs within time windows determined for each population histogram (time windows for depression and activation following reward omission were 100–190 and 460–530 ms, respectively). We found that the magnitude of depressions after reward omission was significantly higher in the $P_r = 0.75$ condition than the $P_r = 0.5$ and 0.25 conditions ($P < 0.05$). Visual inspection of the population histograms also indicates that an activation appeared to follow depressions to reward omission (Fig. 8, middle), but the magnitude of this late increase in activity was not significantly influenced by probability ($P > 0.05$). Also, the magnitude of late activations after reward omission in the $P_r = 0.75$ condition was significantly higher than in the $P_r = 0.5$ and 0.25 conditions ($P < 0.01$). Thus, the sensitivity of TANs to reward omission, with respect to reward probability, was evident both at the level of the population average and at the level of individual neurons, with increased neuronal responsiveness for high reward probability.

We finally examined whether our findings depended on the fact that trials were presented in blocks. Notably, the activity of TANs might adapt to different probabilities of reward in a gradual manner as the animal experienced the situation over successive trials in order to gain insight into probabilistic reward structure of any given block. To assess this possibility, we analyzed the time course of changes in the latency of licking movements on a trial-by-trial basis, following transitions to different probabilities of reward. Within 15–20 trials of a change in reward probabilities, monkeys reached a stationary phase in which lick latencies remained constant in the following trials, suggesting that animals could estimate the probability of reward at that point during testing. For example, when monkeys received reward in the $P_r = 0.25$ condition, we found a significant difference between the first 20 trials of each block and the subsequent 20 trials for lick latencies in monkeys 2 and 3 (Mann–Whitney U -test, $P < 0.01$) and, to a lesser extent, in monkey 1 ($P < 0.05$). We then rated the

magnitude of activity changes at the level of population average, after the first 20 trials of each block were excluded from analysis, using time windows previously determined for each component of TAN responses. Statistically, the significant effect of reward probability on the magnitudes of the different response components was left unchanged (one-way ANOVA followed by Fisher's test), with increased responsiveness to the conditioned stimulus for high reward probability and increased responsiveness to reward delivery as the reward probability decreased, just as described in data from all trials. This suggests that the early behavioral adaptation that followed the switch in reward probabilities did not influence the probability-dependent modulation of TAN responses at the population level.

Location of the recording sites

The recording sites were confirmed histologically in monkeys 1 and 2. As seen in Fig. 9, most neurons were distributed over the dorsal and middle parts of the postcommissural putamen. In both monkeys, the implanted chamber did not allow the most anterior and medial portions of the striatum to be targeted by our electrode tracks, thus explaining that the recording sites did not entail the caudate nucleus and ventral striatum, i.e., the nucleus accumbens and adjacent

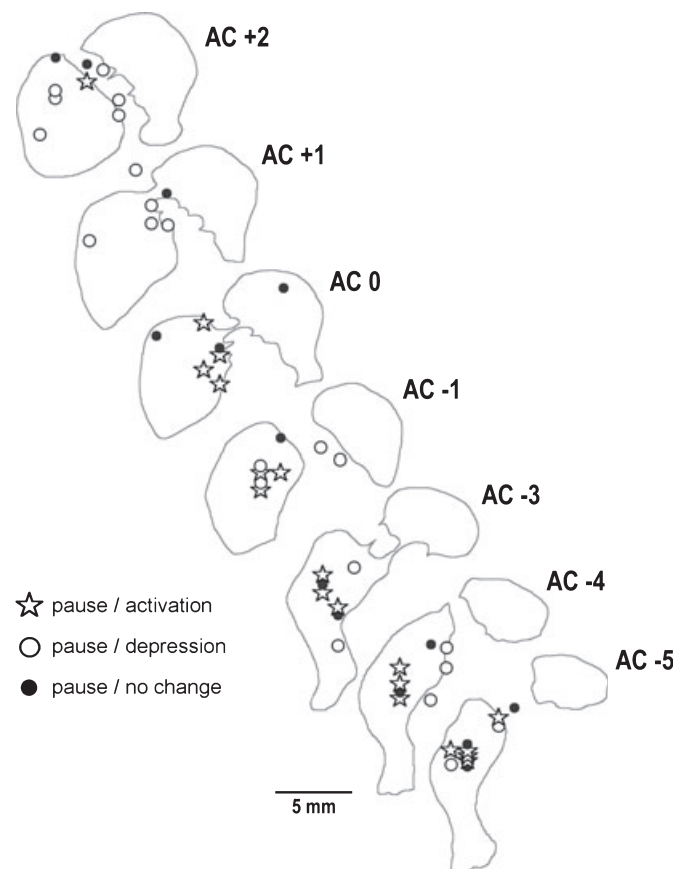


FIG. 9. Histological reconstruction of TANs tested with the different reward probabilities. Neurons from monkeys 1 and 2 are superimposed on coronal sections of the striatum. AC -5 to +2, levels posterior and anterior to the anterior commissure. Symbols indicate properties of recorded TANs: stars, neurons that responded with a decrease and an increase in firing to reward delivery and reward omission, respectively; circles, neurons that responded with a decrease in firing to both delivery and omission of reward; dots, neurons that responded with a decrease in firing to reward and no change to reward omission.

putamen and caudate nucleus. The rostrocaudal extent of the regions from which we recorded was subdivided into three levels. TANs that showed pauses to reward delivery and late activations to reward omission were found significantly more frequently in the posterior parts of the sampled regions, as compared with anterior parts ($\chi^2 = 12.0, P < 0.01$), whereas the fraction of neurons that decreased their firing rate following both reward delivery and reward omission ($\chi^2 = 5.04, P > 0.05$) and those that showed a pause to reward without change to no reward ($\chi^2 = 2.78, P > 0.05$) varied nonsignificantly along the rostrocaudal extent.

Discussion

This study was undertaken to assess the capacity of TANs to adapt their responsiveness to a change in the probability of reward. We found that the responses of TANs to reward were enhanced at lower probabilities, whereas responses to the stimulus that predicted reward appeared somewhat more pronounced for high reward probability. In addition, TANs were modulated following the omission of reward, these changes being expressed as either a decrease or increase in firing rate that became stronger with increasing reward probability. It therefore appears that only one group of TANs displayed opposite changes in activity that could reflect positive and negative errors in the prediction of reward, whereas another group may be involved in signaling unexpected outcomes, irrespective of their rewarding value, possibly reflecting an attentional feature of outcome.

Effects of probability on TAN responses to reward delivery

The modulation of TAN responses by reward probability described here, with the reward response becoming more prominent with lower probabilities, is in line with our previous study showing that TANs appear to signal the extent to which reward timing occurs differently than predicted (Ravel *et al.*, 2001). These results, taken together with recent findings from one other laboratory (Joshua *et al.*, 2008), suggest that TANs are capable of signaling differences between predictions and rewarding outcomes, which correspond to prediction errors thought to be crucial for reinforcement learning (Rescorla & Wagner, 1972; Sutton & Barto, 1981). In this regard, there are similarities in coding capabilities between TANs and midbrain DA neurons. It has been shown in both monkeys and rats that phasic activations in DA neurons are a possible neuronal substrate for the reward prediction error signal (Waelti *et al.*, 2001; Morris *et al.*, 2004; Bayer & Glimcher, 2005; Roesch *et al.*, 2007). Only a few neuronal recording experiments in monkeys have examined whether the coding of prediction errors could arise in the striatum and they have provided conflicting data. In a first study, Morris *et al.* (2004) did not find evidence for the influence of probability on the responses of TANs to reward in a task involving instrumental reactions to gain access to reward. In contrast, the findings of a subsequent study (Joshua *et al.*, 2008) showed that TAN responses to reward were stronger with decreasing reward probability when the monkey simply waited for reward to be delivered, with no requirement for action. The lack of sensitivity to reward probability may therefore reflect a decreased ability of TANs to process information about prediction errors if the reward is contingent on an instrumental response rather than passively received in a classical conditioning protocol (Joshua *et al.*, 2008). We have previously pointed out that the TAN response to reward is most prominent when the monkey is not attending to an instrumental task and instead receives the reward at unpredictable times outside a task (Apicella *et al.*, 1997; Ravel *et al.*, 2001). These results suggest that the capacity of TANs to generate a

prediction error signal depends on the particular situation in which animals experienced the stimulus–outcome associations. In particular, the sensitivity of TANs to basic reward parameters, such as probability and timing, may find expression under situations belonging to the category of procedural learning in which predictions are mediated by extensive training. This interpretation fits with the idea that TANs may be important for the performance of behaviors in specific learning modes, especially those subserving automatic responses or habits (Graybiel, 1998; Apicella, 2002).

Several neuroimaging studies in humans have reported activations related to errors in reward prediction in striatal and prefrontal cortical areas, which are the primary target structures of DA neurons (McClure *et al.*, 2003; Ullsperger & von Cramon, 2003; O'Doherty *et al.*, 2003; Seymour *et al.*, 2004; Abler *et al.*, 2006; Rodriguez *et al.*, 2006; Tobler *et al.*, 2006; Tanaka *et al.*, 2006). It is generally accepted that changes in DA transmission contribute to prediction error-related striatal activations observed with functional magnetic resonance imaging (fMRI). Because TANs most probably correspond to the cholinergic interneurons of the striatum, our findings suggest that the fMRI signal linked to errors in prediction of reward may also arise from local processing within the striatum. However, contrary to most neuroimaging results emphasizing activations localised to the ventral striatum, we found TAN responses modulated by reward probability in the posterior putamen that is associated with motor aspects of behavior (Parent & Hazrati, 1995). This indicates that the involvement of TANs in processing information about reward prediction errors is not restricted to ventral striatal regions.

Although the modulation of the response of TANs to the stimulus that predicted reward appeared less marked than that following reward delivery across the different probability conditions, at least at the level of single neurons, we found evidence that population TAN responses to the predictive stimulus were stronger for high reward probability. The weak modulation of individual TAN responses may have resulted from the use of a conditioned stimulus that did not contain specific information about probability, but merely served as a temporal cue for the upcoming reward. However, it has been reported that the response of TANs to a reward-predicting stimulus was not markedly influenced by the probability of reward, even when using distinct stimuli to indicate the probability explicitly (Morris *et al.*, 2004; Joshua *et al.*, 2008). This suggests that TAN responses to reward-predicting stimuli are less sensitive to changes in reward probability than those of DA neurons, thus emphasizing a difference in coding between the two populations.

TAN responses to the absence of expected reward

In the present study, changes in TAN activity were also observed at the time of reward omission in unrewarded trials and they became more prominent with increasing reward probability. These changes were reflected as either decreases or increases in firing, suggesting that TANs responding to the absence of reward were not equivalent but contain two subsets of neurons whose properties might be related to some functional distinction. These results contrast with the findings of Joshua *et al.* (2008) showing that the TAN response to reward omission was homogeneous, consisting of a phasic depression in firing. However, most previous investigations of the modulation of TAN activity have focused largely on the typical pause response as an index of TAN responsiveness, possibly leaving undetected more subtle increases in firing rate occurring in the absence of a depression in firing (Yamada *et al.*, 2004; Lee *et al.*, 2006). On the other hand, as we have pointed out, the particular learning situation in which our

monkeys experienced changes in reward probability may contribute to the expression of a distinct profile of TAN response following the omission of reward.

In humans, imaging studies have provided evidence of ventral striatal activity related to the processing of prediction error for the absence of expected reward or unexpected negative feedback. However, there is a substantial variability across different studies as to how the activation level of the ventral striatum changes with the detection of negative prediction errors. For example, decreasing activity has been found in some studies (Knutson *et al.*, 2001; O'Doherty *et al.*, 2003; Abler *et al.*, 2006) while other studies have reported increasing activity (Rodriguez *et al.*, 2006; Seymour *et al.*, 2007). It is conceivable that different directions of the striatal fMRI signal accompanying negative reward prediction errors are partly driven by TAN processing within the striatum.

In the present study, we found that the omission of reward delivery was less effective for modulating the activity of TANs than was reward delivery, suggesting that these neurons were more involved in positive prediction error processing than with negative prediction error processing. A possible reason for the weak response to reward omission may be due to a low degree of temporal coupling between changes in TAN firing and the absence of expected outcome in trace conditioning in which outcome timing was not indicated by an external cue. Accordingly, a lack of predictive accuracy on the time of potential reward could contribute to the variations in the latency of activity changes we observed after reward omission. Indeed, Lee *et al.* (2006) have emphasized that the responses of TANs are not well temporally linked to events whose occurrence must be internally timed. Also, Joshua *et al.* (2008) have indicated that the responsiveness of TANs to reward omission became more evident if the termination of the stimulus–reward interval was made more salient by adding a sound. It is therefore possible that the sensitivity of TANs to reward omission is more prominent when the presence of an external cue provides reliable information about the timing of outcome.

Are TANs capable of coding a reward prediction signal?

It is generally accepted that the encoding of a prediction error signal of the type required by reinforcement learning models takes the form of changes in neuronal activity into opposite directions (Schultz & Dickinson, 2000; Niv & Schoenbaum, 2008). This view has been exemplified by the observation of phasic changes in activity of DA neurons which are thought to encode a positive and negative reward prediction error by an increase and decrease in firing, respectively (Schultz, 2002). Recently, it has been shown that neurons in the lateral habenula have the capacity to differentially encode positive and negative reward prediction errors with activity changes elicited by reward omission in the opposite direction from those elicited by reward delivery (Matsumoto & Hikosaka, 2007). In the present study, we found that a group of TANs showed changes in firing that resemble the characteristics of a full prediction error signal, with decreased and increased firing in response to reward delivery and reward omission, respectively. These TANs showing directional changes in activity appeared to be located predominantly in the putamen caudal to the anterior commissure, an area known to be associated with the processing of motor information. Further studies are necessary to clarify whether different parts of the striatum make different contributions to the processing of information about reward prediction errors.

In the other group of TANs we recorded, the detection of reward and no reward resulted in changes in activity in the same direction, i.e., a decrease in firing. These neurons appear similar to those

described by Joshua *et al.* (2008) in a probabilistic conditioning task. Although it is possible that positive and negative prediction errors can be signaled by specific patterns of decreasing activity (Bayer *et al.*, 2007), it must be emphasized that the experience of reward prediction error was not the only aspect to vary when we manipulated the probability of reward. In particular, changes in the attentional demands of the stimulus–reward pairing may have been an additional factor contributing to differences in TAN activity. In this regard, neurons responding in the same way to both reward and no reward may be related to some common process of event detection, such as enhanced arousal elicited by infrequent outcomes. Indeed, there is some evidence linking TAN responses to arousal (Aosaki *et al.*, 1994; Blazquez *et al.*, 2002) and an fMRI study has revealed increased striatal activity in response to the surprising presentation of salient stimuli, regardless of reward expectation (Zink *et al.*, 2003).

Conclusion

The present study is the first demonstration that some TANs are capable of encoding positively and negatively valued differences between the expected and obtained reward and thus may participate in reinforcement learning. In this regard, the response properties of at least a subset of TANs appear to correspond to the characteristics of the DA signal. As they are thought to be cholinergic interneurons, TANs might use prediction error signals to influence striatal output circuits. Little is known about the impact of changes in TAN activity on striatal projection neurons, but it seems likely that the TAN signal would exert a local effect on specific groups of projection neurons that are involved in the control of action. Although the role of striatal output circuits in processing error signals has not been investigated specifically, this interaction may be essential in the planning and execution of movements, particularly to determine the value of the stimulus that is selected to drive subsequent behavioral reactions. It remains to be clarified, however, to what extent the findings reported here are specific to procedural forms of learning that underlie the development of habits. Our assumption is that the TAN network participates in a type of error coding that is engaged when rewards are processed in an automatic manner, whereas tracking errors to optimize the acquisition of new action–outcome relations is processed by the DA system.

Acknowledgements

We thank Dr J. F. Espinosa-Parrilla for help with behavioral training and Dr I. Balansard for assistance with surgery. This research was supported by Centre National de la Recherche Scientifique.

Abbreviations

DA, dopamine; fMRI, functional magnetic resonance imaging; P_r , probability of reward; TAN, tonically active neuron.

References

- Abler, B., Walter, H., Erk, S., Kammerer, H. & Spitzer, M. (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *Neuroimage*, **31**, 790–795.
- Aosaki, T., Tsubokawa, H., Ishida, A., Watanabe, K., Graybiel, A.M. & Kimura, M. (1994) Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci.*, **14**, 3969–3984.

- Apicella, P. (2002) Tonicly active neurons in the primate striatum and their role in the processing of information about motivationally relevant events. *Eur. J. Neurosci.*, **16**, 2017–2026.
- Apicella, P. (2007) Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends Neurosci.*, **30**, 299–306.
- Apicella, P., Legallet, E. & Trouche, E. (1997) Responses of tonically discharging neurons in the monkey striatum to primary rewards delivered during different behavioral states. *Exp. Brain Res.*, **116**, 456–466.
- Barnes, T.D., Kubota, Y., Hu, D., Jin, D.Z. & Graybiel, A.M. (2005) Activity of striatal neurons reflects dynamic encoding and recoding of procedural memories. *Nature*, **437**, 1158–1161.
- Bayer, H.M. & Glimcher, P.W. (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, **47**, 129–141.
- Bayer, H.M., Lau, B. & Glimcher, P.W. (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. *J. Neurophysiol.*, **98**, 1428–1439.
- Blazquez, P.M., Fujii, N., Kojima, J. & Graybiel, A.M. (2002) A network representation of response probability in the striatum. *Neuron*, **33**, 973–982.
- Brasted, P.J. & Wise, S.P. (2004) Comparison of learning-related neuronal activity in the dorsal premotor cortex and striatum. *Eur. J. Neurosci.*, **19**, 721–740.
- Graybiel, A.M. (1998) Building action repertoires: memory and learning functions of the basal ganglia. *Curr. Opin. Neurobiol.*, **5**, 733–741.
- Jog, M.S., Kubota, Y., Connolly, C.I., Hillegaart, V. & Graybiel, A.M. (1999) Building neural representations of habits. *Science*, **286**, 1745–1749.
- Joshua, M., Adler, A., Mitelman, R., Vaadia, E. & Bergman, H. (2008) Midbrain dopaminergic neurons and striatal cholinergic interneurons encode the difference between reward and aversive events at different epochs of probabilistic classical conditioning trials. *J. Neurosci.*, **28**, 11673–11684.
- Kimura, M., Rajkowski, J. & Evarts, E.V. (1984) Tonicly discharging putamen neurons exhibit set dependent responses. *Proc. Natl Acad. Sci. USA*, **81**, 4998–5001.
- Knutson, B., Fong, G.W., Adams, C.M., Varner, J.L. & Hommer, D. (2001) Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport*, **12**, 3683–3687.
- Lee, I.H., Seitz, A.R. & Assad, J.A. (2006) Activity of tonically active neurons in the monkey putamen during initiation and withholding of movement. *J. Neurophysiol.*, **95**, 2391–2403.
- Matsumoto, M. & Hikosaka, O. (2007) Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature*, **447**, 1111–1115.
- McClure, S.M., Berns, G.S. & Montague, P.R. (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, **38**, 339–346.
- 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**, 133–143.
- Niv, Y. & Schoenbaum, G. (2008) Dialogues on prediction errors. *Trends Cogn. Sci.*, **12**, 265–272.
- O'Doherty, J.P., Dayan, P., Friston, K., Critchley, H. & Dolan, R.J. (2003) Temporal difference models and reward-related learning in the human brain. *Neuron*, **38**, 329–337.
- Packard, M.G. & Knowlton, B.J. (2002) Learning and memory functions of the basal ganglia. *Annu. Rev. Neurosci.*, **25**, 563–593.
- Parent, A. & Hazrati, L.N. (1995) Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Res. Rev.*, **20**, 91–127.
- Pasupathy, A. & Miller, E.K. (2005) Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*, **433**, 873–876.
- Ravel, S., Sardo, P., Legallet, E. & Apicella, P. (2001) Reward unpredictability inside and outside of a task context as a determinant of the responses of tonically active neurons in the monkey striatum. *J. Neurosci.*, **21**, 5730–5739.
- Rescorla, R. & Wagner, A. (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. & Prokasy, W. (Eds), *Classical Conditioning II: Current Research and Theory*. Appleton Century Crofts, New York, pp. 64–99.
- Rodriguez, P.F., Aron, A.R. & Poldrack, R.A. (2006) Ventral-striatal/nucleus accumbens sensitivity to prediction errors during classification learning. *Hum. Brain Mapp.*, **27**, 306–313.
- Roesch, M.R., Calu, D.J. & Schoenbaum, G. (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat. Neurosci.*, **10**, 1615–1624.
- Salmon, D.P. & Butters, N. (1995) Neurobiology of skill and habit learning. *Curr. Opin. Neurobiol.*, **5**, 184–190.
- Sardo, P., Ravel, S., Legallet, E. & Apicella, P. (2000) Influence of the predicted time of stimuli eliciting movements on responses of tonically active neurons in the monkey striatum. *Eur. J. Neurosci.*, **12**, 1801–1816.
- Schultz, W. (2002) Getting formal with dopamine and reward. *Neuron*, **36**, 241–263.
- Schultz, W. & Dickinson, A. (2000) Neuronal coding of prediction errors. *Annu. Rev. Neurosci.*, **23**, 473–500.
- Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J. & Frackowiak, R.S. (2004) Temporal difference models describe higher-order learning in humans. *Nature*, **429**, 664–667.
- Seymour, B., Daw, N., Dayan, P., Singer, T. & Dolan, R. (2007) Differential encoding of losses and gains in the human striatum. *J. Neurosci.*, **27**, 4826–4831.
- Sutton, R.S. & Barto, A.G. (1981) Toward a modern theory of adaptive networks: expectation and prediction. *Psychol. Rev.*, **88**, 135–170.
- Tanaka, S.C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S. & Doya, K. (2006) Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Netw.*, **19**, 1233–1241.
- Tang, C., Pawlak, A.P., Prokopenko, V. & West, M.O. (2007) Changes in activity of the striatum during formation of a motor habit. *Eur. J. Neurosci.*, **25**, 1212–1227.
- Tobler, P.N., O'Doherty, J.P., Dolan, R.J. & Schultz, W. (2006) Human neural learning depends on reward prediction errors in the blocking paradigm. *J. Neurophysiol.*, **95**, 301–310.
- Tremblay, L., Hollerman, J.R. & Schultz, W. (1998) Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.*, **80**, 964–977.
- Ullsperger, M. & von Cramon, D.Y. (2003) Error monitoring using external feedback: specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *J. Neurosci.*, **23**, 4308–4314.
- Waelti, P., Dickinson, A. & Schultz, W. (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, **412**, 43–48.
- Yamada, H., Matsumoto, N. & Kimura, M. (2004) Tonicly active neurons in the primate caudate nucleus and putamen differentially encode instructed motivational outcomes of action. *J. Neurosci.*, **24**, 3500–3510.
- Zink, C.F., Pagnoni, G., Martin, M.E., Dhamala, M. & Berns, G.S. (2003) Human striatal response to salient nonrewarding stimuli. *J. Neurosci.*, **23**, 8092–8097.