

Neural Coding of Distinct Statistical Properties of Reward Information in Humans

Jean-Claude Dreher^{1,2}, Philip Kohn¹ and Karen Faith Berman¹¹Unit on Integrative Neuroimaging, Clinical Brain Disorders Branch, NIMH, NIH, IRP, Bethesda, MD 892-365, USA and²Current address: Institute of Cognitive Science, CNRS-Université de Lyon 1, 67 Bd Pinel, 69675 Bron, France

Brain processing of reward information is essential for complex functions such as learning and motivation. Recent primate electrophysiological studies using concepts from information, economic and learning theories indicate that the midbrain may code two statistical parameters of reward information: a transient reward error prediction signal that varies linearly with reward probability and a sustained signal that varies highly non-linearly with reward probability and that is highest with maximal reward uncertainty (reward probability = 0.5). Here, using event-related functional magnetic resonance imaging, we disentangled these two signals in humans using a novel paradigm that systematically varied monetary reward probability, magnitude and expected reward value. The midbrain was activated both transiently with the error prediction signal and in a sustained fashion with reward uncertainty. Moreover, distinct activity dynamics were observed in post-synaptic midbrain projection sites: the prefrontal cortex responded to the transient error prediction signal while the ventral striatum covaried with the sustained reward uncertainty signal. These data suggest that the prefrontal cortex may generate the reward prediction while the ventral striatum may be involved in motivational processes that are useful when an organism needs to obtain more information about its environment. Our results indicate that distinct functional brain networks code different aspects of the statistical properties of reward information in humans.

Keywords: dopamine, error prediction, fMRI, gambling, information theory, uncertainty

Introduction

Detecting and predicting reward information are fundamental capabilities of simple life forms that have evolved in mammals into complex behavioral characteristics including learning and motivation. Reward information can be extracted from a large variety of stimuli and concerns the presence, qualities and magnitudes of rewards, their predictability and the motivational value associated with them. Of the many stimuli occurring in our natural environment during a certain time period, very few are actually paired with reward. It is thus essential that the brain represents the statistical properties of the stimuli leading to reward information, which provides a critical evolutionary advantage for survival in a changing environment.

Reward information processing depends upon a small number of brain regions, such as the prefrontal cortex (dorsolateral and orbitofrontal parts), the anterior cingulate cortex, the ventral striatum and the midbrain dopaminergic region (Schultz, 2000, 2002). Of particular importance for the normal functioning of this system are midbrain dopaminergic neurons that project in a divergent fashion to discrete targets, putting them in a unique position to broadcast their signals. Based on

series of elegant electrophysiological findings in monkeys, Schultz and colleagues have provided a theoretical framework for understanding the functions of these neurons. These studies suggest that the activity of dopaminergic neurons precisely codes two statistical parameters of reward information (Schultz *et al.*, 1997; Fiorillo *et al.*, 2003). The first parameter is coded by a phasic mode of dopamine neuronal activity that reflects a 'reward error prediction', which is the discrepancy between the probability with which reward is predicted and the actual outcome (Schultz *et al.*, 1997; Waelti *et al.*, 2001; Fiorillo *et al.*, 2003). This phasic dopamine signal may be used as a teaching signal by other brain structures for the learning of reward-directed behavior. After conditioning in a Pavlovian procedure in which distinct visual stimuli have specific reward probability, the phasic dopamine signal varies linearly with reward probability (Fiorillo *et al.*, 2003). That is, when reward probability increases, the phasic response of dopamine neurons increases at the time of the reward-predicting stimulus and decreases at the time of the reward (the reward magnitude being fixed).

In addition to this phasic mode of activity, important new results indicate that dopamine neurons exhibit a sustained mode of activity after learning, that covaries with a second statistical parameter of reward information, reward uncertainty (maximal when reward probability = 0.5), and grows from the onset of the conditioned stimulus to the expected time of reward delivery (Fiorillo *et al.*, 2003). Moreover, the growth of this sustained activity is steeper when the discrepancy between potential reward magnitudes is higher (e.g. a stimulus predicting either a small or large reward with maximal uncertainty shows higher sustained activity than a stimulus predicting a small or medium reward magnitude). The sustained mode of activity occurring with maximal reward uncertainty may be related to a specific form of attention (Pearce and Hall, 1980), to motivational processes in the context of reward uncertainty, or to the expectation of reward information following rules from information theory (Shannon, 1948). A basic concept of information theory is Shannon's entropy, which measures an ensemble's average information content or its uncertainty, and which is maximal for outcomes having a 50% chance of occurrence (the more uncertain the outcome, the more information conveyed by it). Similarly, the Pearce-Hall psychological model of attention proposed that attention to a stimulus and associability are enhanced if there is uncertainty about the predictions associated with this stimulus (while a conditioned stimulus loses its association with a reinforcer when its consequences are accurately predicted) (Pearce and Hall, 1980).

Regardless of the exact function of the sustained mode of activity, a major question that remains is whether similar phasic

and sustained activity dynamics can be observed in the human reward system. If both the phasic and sustained signals reported in single-cell monkey recordings can be observed in humans, this would support a unified cross-species view in which midbrain neurons obey common basic principles of neural computation and provide important new insights into human reward information processing. Another critical question is whether post-synaptic targets of midbrain neurons respond differentially to the phasic error prediction signal and the sustained reward uncertainty signal. If activity patterns in these projection sites were found to covary differently with each of the two types of signals, it would implicate differential neural mechanisms involved in higher-order processing of the error prediction and reward uncertainty signals. We hypothesized that the transient reward error prediction signal should be processed in brain regions that make or maintain the prediction, such as the lateral prefrontal cortex (Schultz *et al.*, 1997; Kobayashi *et al.*, 2002; Watanabe *et al.*, 2002). Conversely, we hypothesized that the sustained reward uncertainty signal should covary with brain regions involved in the expectation of reward information in terms of information theory.

To test predictions about reward-related neural activity in humans that emerge from monkey electrophysiological data and to investigate whether post-synaptic targets of midbrain dopaminergic neurons respond differentially to the phasic error

prediction signal and the sustained reward uncertainty signal, we designed a new paradigm for event-related functional magnetic resonance imaging (fMRI) (see Materials and Methods and Fig. 1, Table 1). Thirty-one subjects were paid for responding to different 'slot machines' that systematically varied monetary reward probability, magnitude and expected reward value (i.e. product probability \times magnitude, which is a crucial parameter of reward information) (Glimcher, 2003). This allowed us to distinguish the influence of these parameters of reward information on brain regions responding *transiently* with the error prediction signal at the time of the reward-predicting stimulus and at the time of the rewarded outcome versus in a *sustained* fashion during anticipation of rewards of maximal uncertainty.

Materials and Methods

Subjects

Thirty-one young right-handed subjects (mean age = 27.6; SD 5.7, 16 males) were recruited following procedures approved by the National Institute of Mental Health Institutional Review Board and provided written informed consent. All subjects were free of neurological, psychiatric and substance abuse problems. They had no history of gambling and no medical problems or medical treatment that could affect cerebral metabolism and blood flow. Subjects were paid for participating and earned extra money for performing the task described

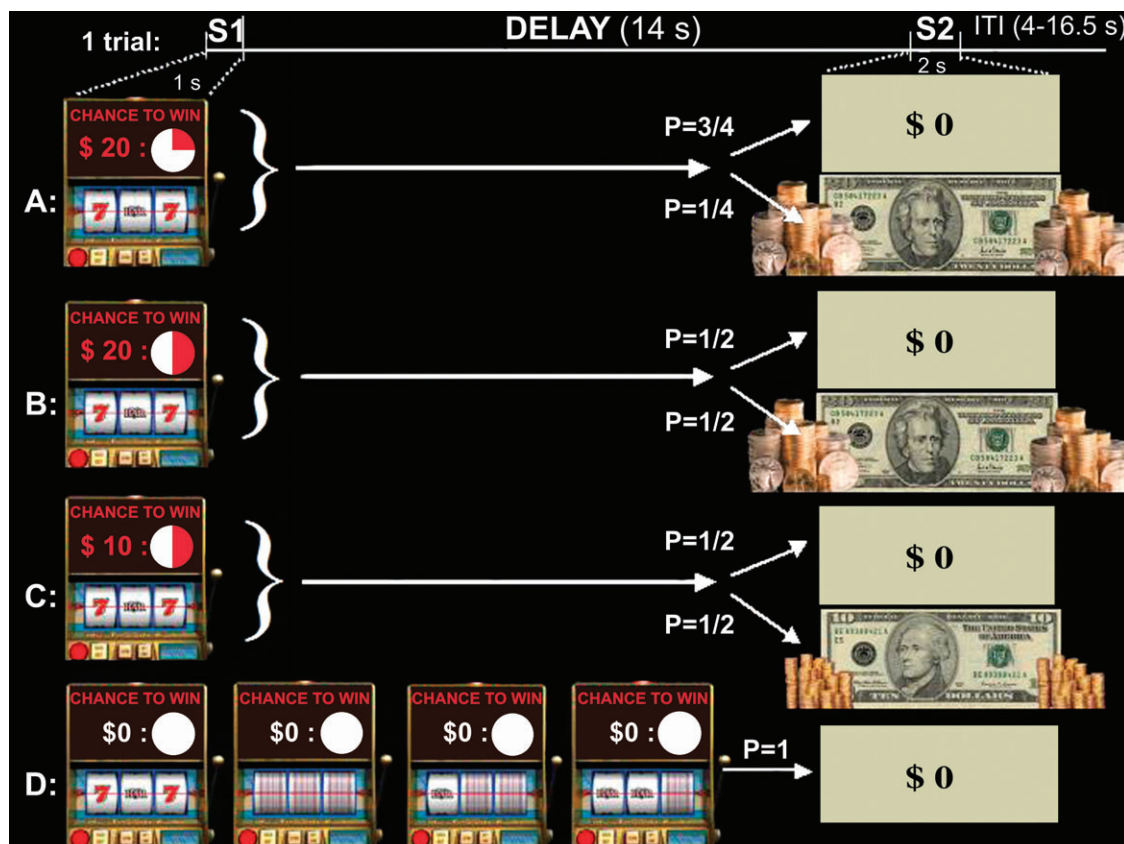


Figure 1. Task design. Four types of 'slot machines' (types A–D) were presented pseudo-randomly. The probabilities of winning different amounts of money or nothing were indicated, respectively, by the red and white portions of a pie chart above the slot machines. The slot machine and pie chart remained on the screen throughout the delay duration (as shown for slot D). Each trial consisted of a brief (1 s) presentation of the cue (stimulus S1, one of the four slot machines), followed after a fixed delay (14 s) by the outcome S2 (either \$0 or a picture of a \$10 or \$20 bill, lasting 2 s). This long fixed delay allowed us to distinguish transient hemodynamic signals associated with the error prediction signal at S1 and S2 from the sustained signal associated with reward uncertainty during the delay. During each trial, subjects indicated which 'slot machine' was presented by pressing a response button both at the cue S1 and the outcome S2 (regardless of winning or not). Reward delivery was not contingent upon subject response.

Table 1

Task design. In bold are outlined the contrasts revealing transient (S1: Prediction and S2: error prediction) and sustained activation (Delay: reward uncertainty). ↑ higher; ↓ lower; ↔ equal.

Phase of the trial	Reward probability: P	Potential rewards magnitudes: M	Expected reward value: E
S1: Prediction (S1: $P = 0.5$ \$20 > $P = 0.25$ \$20)	$P \uparrow$	$M \leftrightarrow$	$E \uparrow$
Delay: Reward uncertainty (Delay: $P = 0.5$ \$20 > $P = 0.25$ \$20) Delay: Maximal discrepancy between potential reward magnitudes (Delay: $P = 0.5$ \$20 > $P = 0.5$ \$10)	$P \uparrow$	$M \leftrightarrow$	$E \uparrow$
	$P \leftrightarrow$	$M \uparrow$	$E \uparrow$
Delay: equal expected reward value but variable probability and magnitude (Delay: $P = 0.5$ \$10 > $P = 0.25$ \$20)	$P \uparrow$	$M \downarrow$	$E \leftrightarrow$
S2: Positive error prediction (S2: receive \$20 in the context of $P = 0.25$ > receive \$20 in the context of $P = 0.5$)	$P \downarrow$	$M \leftrightarrow$	$E \downarrow$
S2: Negative error prediction (S2: receive \$0 in the context of reward probability $P = 0.5$ > receive \$0 in the context of reward probability $P = 0.25$)	$P \uparrow$	$M \leftrightarrow$	$E \uparrow$

below. Subjects were told that they would earn only a percentage of each of the \$10/\$20 bills presented on the screen.

Experimental Paradigm

The stimuli representing the 'slot machines' were projected on a screen positioned at the foot of the subjects. The experiment was performed using Presentation software (version 6.0, <http://nbs.neuro-bs.com/presentation>) on a laptop PC running Windows 2000.

Experimental trials were divided into three phases (Fig. 1). During the first two phases (i.e. presence of the slot machine on the screen), the words: 'Chance to win \$ XX' (where XX stands for \$0, \$10 and \$20) remained visible on top of each slot machine with a pie chart indicating in red the probability of winning the indicated amount of money and in white the probability of receiving nothing.

The first phase of the trial consisted of the presentation of a cue, S1, lasting 1 s, representing one of the four slot machines (A, B, C or D):

Slot machine (A): $P = 1/4$; \$20, $P = 3/4$; \$0.

Slot machine (B): $P = 1/2$; \$20, $P = 1/2$; \$0.

Slot machine (C): $P = 1/2$; \$10, $P = 1/2$; \$0.

Slot machine (D): $P = 1$; \$0 (sure to get no reward).

These slot machines were designed to vary reward probability, magnitude and expected reward value (reward probability \times magnitude):

A and B: Reward values match, reward probabilities differ, expected reward values differ;

A and C: Reward values differ, reward probabilities differ, expected reward values equal;

B and C: Reward values differ, reward probabilities match, expected reward values differ.

During the second (delay) phase, the pie chart remained on the screen and each of the three spinners of the slot machine rotated successively before stopping on a fixed image ('bar' or '7') that lasted until the end of the trial. The overall delay duration was 14 s (starting after the cue S1), and was divided as indicated for the slot machine D in Figure 1 (the first 'bar' or '7' image appeared at 13.5/3 s, the second at $2 \times 13.5/3$ s and the third at 13.5 s). This fixed delay was necessary because dopamine neurons are known to be susceptible to the uncertainty of the timing delivery of rewards. Moreover, the long delay between S1 and S2 was required in order to distinguish phasic and sustained modes of activity. Therefore, the timing of our design made it possible to distinguish hemodynamic signals associated with the error

prediction signal at the time of S1 presentation from those associated with anticipation during the delay and those associated with the error prediction signal at the time of the outcome S2.

In the third 'outcome', or S2 phase, '\$0', '\$10' or '\$20' was projected for 2 s. Pictures of the \$10 and \$20 bills were surrounded, respectively, by a small and a large stack of gold pieces in order to produce direct visual experience of distinct reward magnitudes and reinforce the pleasantness of winning money (see Fig. 1). To equalize visual similarity between stimuli, the '\$0' outcome was presented in a grey rectangle that had the same dimensions as the bills. The inter-trial interval between slot machines varied between 4 and 16.5 s with a geometric distribution of mean = 6.8 s.

Subjects indicated which slot machine was presented by pressing a specific response button on a diamond-shaped four-response button device at the time of S1 and the time of S2 (the same response button press corresponding to the current slot machine at S1 was also required at S2 regardless of winning or not). The association between each slot machine and a specific response button was previously learned during a training session before scanning (see Training session). These motor responses ensured that subjects were attending to the specific types of slot machines as well as their outcomes and enabled us to keep the motor component equal between S1 and S2. No requirement was made to press the response button quickly. Importantly, the stimuli presentations were not contingent on the subject's response. The number of correctly responded trials was indicated at the end of each run to the subject during scanning.

We designed the 'slot machines' A, B, C by including the common null outcome '\$0' because the emotional response to the outcome of a 'gamble' depends on the perceived value and likelihood of both the obtained outcome and its alternatives (Mellers, 2000). Without this common null outcome, smaller wins could be experienced as more pleasurable than larger wins for the same reward probabilities between two slot machines (e.g. it would feel better to win \$10 if the alternative is \$1, than if the alternative is \$200, an effect called the 'counterfactual comparison'). For the same reason, the presence of this null outcome allowed us to investigate the effect of reward probability on brain activity by comparing the two 'slot machines' (A and B) that possess identical potential reward magnitudes ('\$0' or '\$20') but different reward probabilities. Moreover, the use of the 'slot machine' D ($P = 1$ to win nothing) allowed us to control for a non-selective effect of attention/anticipation without motivation to win.

There were a total of six runs, each consisting of 16 trials (four trials for each type of slot machine). Each of the four possible slot machines occurred pseudo-randomly during each run. To prevent fluctuations between runs, the probability of each potential outcome was exact and reached at the end of each run for each slot machine. In addition, to control for the order of the presentation of the slot machines, we designed sequences of slot machines using conditional probabilities at two levels: each slot machine had equal probability of being followed by any of the four slot machines (conditional probability of order 1) and two successive slot machines also had the same probability of occurrence (conditional probability of order 2). The order of the runs was randomized between subjects.

Training Session

Before scanning, subjects performed a training session (lasting ~10 min), in which they performed one run of the task. They were told that, unlike in a casino, the odds of winning for each slot machine were known. They were shown the four types of slot machines with different probability of winning real money indicated by the pie chart (Fig. 1). During this training session, subjects simply had to learn the association between each slot machine and a specific response button on a diamond-shape 4 responses button device (top button for slot A, left button for slot B, right button for slot C and bottom response button for slot D). Subjects were asked to respond with their right thumb by pressing the response button corresponding to the current slot machine presentation, first at the time of the cue S1 and then at the time of the outcome S2 (the same response button press corresponding to the current slot machine was required at S2 regardless of winning or not).

Note that in the electrophysiological study on phasic and sustained dopaminergic neuronal activity, monkeys had to learn the probability between an arbitrary stimulus and a potential reward for ~5 weeks

(100–200 trials/day), and recordings began only after learning, as evidenced by emergence of discriminative conditioned licking responses (Fiorillo *et al.*, 2003). In our design, the explicit presentation of the reward probability and magnitude avoided having the subjects learn the probability between a conditioned stimulus and the outcome. This ensured that subjects knew the exact winning probabilities (since it is difficult to have a behavioral index of this learning effect in humans) and avoided long learning sessions to associate an arbitrary stimulus with a certain reward probability. The explicit display of the reward probability also avoided brain activation that would have arisen at intermediate probabilities if subjects had to learn reward probabilities in the scanner. Even though the probabilities were known, subjects could not predict with certainty the actual reward and continuously experienced the random nature of the outcome, intrinsic to the slot machines A–C.

fMRI Data Acquisition

Imaging was conducted on a General Electric 3 T scanner with a real-time functional imaging upgrade (General Electric Medical Systems, Milwaukee, WI). A sagittal localizer scan was used to orient subsequent scans. Functional imaging scans involved a series of 29 contiguous 3.3 mm axial slices per volume collected over six runs, plus 8 ‘dummy’ volumes at the start of each session. These functional scans used an echo-planar single shot real-time gradient echo T_2^* weighting (EPIRT sequence, repetition time = 2300 ms, echo time = 23 ms, field of view = 24 cm, 64×64 matrix, voxel size = $3.75 \times 3.75 \times 3.3$, flip angle = 90° , ramp sampling on).

Signal dropout in the orbitofrontal cortex due to susceptibility artifact was reduced by using a local high-order z -shimming performed in the axial direction with an oval-shape region that included the orbitofrontal cortex and the basal ganglia. In addition, we tilted the head of the subject with a 30° angle relative to the AC–PC line because this simple head positioning procedure improves the shim in this area (Heberlein and Hu, 1991). Indeed, a region of high field distortion is located above the nasal cavity and a region of low field is located behind the nasal cavity. Since these field effects are dependent on the direction of the main magnetic field relative to the head, it is possible to direct the distortions away from the inferior frontal lobe by using this simple positioning method for acquiring para-axial slices traversing the anterior commissure and the posterior commissure (Heberlein and Hu, 1991). The head position was obtained by slightly increasing the padding on the base of the neck and reducing padding for the back of the head using an air pressure-inflatable pillow. The angle was obtained using a protractor measuring the angle between the vertical and the AC–PC line and was then checked with the sagittal localizer.

High-resolution T_1 -weighted structural scans from each subject used a magnetization prepared gradient echo sequence (MP-RAGE) (180 1.0 mm sagittal slices; FOV = 256 mm, NEX = 1, T_R = 11.4 ms, T_E = 4.4 ms; matrix = 256×256 ; T_1 = 300 ms, bandwidth 130 Hz/pixel = 33 kHz for 256 pixels in-plane resolution = 1 mm^3). Real time reconstruction method and AFNI (<http://afni.nimh.nih.gov/afni/>) were used in the scanner room to monitor subject’s head motion on-line as they were performing the task in the scanner.

Image Analysis

All data were analyzed using Statistical Parametric Mapping (SPM99, <http://www.fil.ion.ucl.ac.uk/spm/spm99.html>; Wellcome Department of Cognitive Neurology, London, UK). Data pre-processing of the functional scans included slice timing and motion correction, co-registration to the anatomical data, alignment to the first volume for each subject and spatial normalization to the Montreal Neurological Institute (MNI) T_1 -weighted template image supplied with SPM99. The images were then smoothed with a 10 mm full width at half maximum Gaussian kernel.

Within-subject time series modeling accounted for the following 15 regressors:

- four regressors at S1: [S1, slot A: $P = 0.25$ \$20, $P = 0.75$ \$0], [S1, slot B: $P = 0.5$ \$20, $P = 0.5$ \$0], [S1, slot C: $P = 0.5$ \$10, $P = 0.5$ \$0], [S1, slot D: $P = 1$ \$0];

- four regressors during the delay: [Delay, slot A], [Delay, slot B], [Delay, slot C], [Delay, slot D];
- seven regressors at S2: [S2_no reward, slot A], [S2_no reward, slot B], [S2_no reward, slot C], [S2_no reward, slot D], [S2_reward, slot A], [S2_reward, slot B], [S2_reward, slot C].

This design allowed us to decompose the three phases of each type of ‘slot machine’ according to reward probability and reward magnitude. The fMRI response to each event type was modeled as a delta function at S1 (1 s) and S2 (2 s) and as a rectangular pulse for the presence of the slot machine on the screen (15 s) convolved with a canonical hemodynamic response function (HRF). The default high-pass filter from SPM99 was applied to the time series.

Condition-specific estimates of neural activity (betas) were computed independently at each voxel for each subject, using the general linear model. Using random-effects analysis, we then entered the relevant contrasts of parameter estimates from the 31 subjects into a series of one-way t -tests. We used a threshold of $P < 0.005$, uncorrected for the whole brain (random effects model).

Contrasts of Interests

First, in order to distinguish the brain regions responding transiently to the error prediction signal and with reward uncertainty during the delay period, we focused on the following contrasts:

Transient Activity (Error Prediction)

Brain regions responding in a transient fashion at the time of the cue S1 with higher reward probability should increase their activity in the comparison: ($S1_{\text{slot B}} > S1_{\text{slot A}}$, i.e. [$S1: P = 0.5$ \$20, $P = 0.5$ \$0] $>$ [$S1: P = 0.25$ \$20, $P = 0.75$ \$0]).

Brain regions responding in a transient fashion at the time of the reward delivery at S2 with lower reward probability should show more activation in the comparison: (rewarded outcome $S2_{\$20 \text{ slot A}} > S2_{\$20 \text{ slot B}}$, i.e. receive \$20 in the context of slot A $>$ receive \$20 in slot B).

Brain regions responding in a transient fashion with the negative prediction error at the time of non-reward delivery at S2 (i.e. when reward is omitted in a more unexpected context) should increase their activity in the comparison: ($S2_{P = 0.5 \$0} > S2_{P = 0.75 \$0}$, i.e. no reward at S2 in the comparison $S2_{\$0 \text{ slot B}} > S2_{\$0 \text{ slot A}}$). This comparison should be viewed with caution because it is unclear whether a neuronal deactivation from a negative predication error response corresponds to a BOLD signal deactivation or activation (as noted by O’Doherty *et al.*, 2003).

The comparison at S1 therefore refers to a contrast between two predictions made at cues indicating different reward probabilities, while the comparison made at S2 refers to the difference between two error prediction signals.

Sustained Activity with the Reward Uncertainty Signal

Brain regions responding to the reward uncertainty signal during the delay should be activated in the comparison: (Delay_{slot slot B} $>$ Delay_{Delay slot A}, i.e. [Delay: $P = 0.5$ \$20, $P = 0.5$ \$0] $>$ [Delay: $P = 0.25$ \$20, $P = 0.75$ \$0]).

The results of these analyses aimed at transient and sustained activities are reported in Table 2. Note that since the monkey study shows a symmetric relationship between reward probability and the sustained mode of dopaminergic activity (similar sustained activities were observed for $P = 0.25$ and $P = 0.75$), we reasoned that the comparisons between the two slot machines with reward probabilities of $P = 0.5$ and $P = 0.25$ (slot B versus slot A) would reveal brain regions involved with reward uncertainty during the delay period. The long duration of our trials prevented us from including the full range of reward probabilities (e.g. from 0 to 1 with 0.25 reward probability increment).

Moreover, in the context of our probability range, both the error prediction signal at S1 and the reward uncertainty signal during the delay period should covary with an increase in reward probability (from $P = 0.25$ to $P = 0.5$). However, these two signals can be distinguished by their temporal characteristics since an intrinsic property of the error prediction signal is its transient activity and a major characteristic of the reward uncertainty signal is its sustained activity during the delay.

Table 2

Foci of activations in the different statistical contrasts

Anatomical structure (Brodmann's area)	Error prediction at S1: $P = 0.5 \$20 > P = 0.25 \20		Error prediction at S2: $\$20$ in slot A $> \$20$ in slot B		Reward uncertainty during the delay: $P = 0.5 \$20 > P = 0.25 \20		Conjunction Error prediction at S1 and S2	
	Peak MNI coordinates, $x y z$	Z-value	Peak MNI coordinates, $x y z$	Z-value	Peak MNI coordinates, $x y z$	Z-value	Peak MNI coordinates, $x y z$	Z-value
Dopaminergic midbrain	−4 −23 −15	2.5*	8 −19 −15 −8 −15 −15	3.1 2.97	−4 −19 −15 −4 −8 −11	2.83 2.98	0 −23 −15	2.92
Left DLPFC	−46 19 30	3.3	−34 27 34	3.1			−42 23 34	2.96
Left inferior frontal gyrus (BA 47)	−38 34 0	3.2	−38 27 4	3.0			−38 27 0	3.59
Right orbitofrontal cortex (BA 11)	42 23 −19	3.17	27 27 −15	2.3*			−53 19 −8	3.5
Left superior frontal gyrus	−46 8 57	3.14					42 23 −8	3.62
Right superior frontal gyrus	38 11 53	3.54						
Left anterior cingulate cortex	−11 27 19	2.98						
Right anterior cingulate cortex	4 15 27	2.75						
Superior frontal gyrus (medial)	−4 23 65	3.42	−11 15 61	3.27			−11 8 65	3.65
Left inferior frontal gyrus (BA 44/45)			−53 15 8	2.94				
Left intra-parietal sulcus region	−42 −49 42	2.86	−49 −42 57	3.28			−46 −27 38	2.92
Right intra-parietal region	27 −61 42	2.92					30 −57 57	2.77
Left motor cortex			−39 −20 59	3.16				
Right putamen					23 8 −11	3.52		
Left putamen/caudate					−30 0 −11	3.11		

All areas were significant at $P < 0.005$, uncorrected for multiple comparisons (random effects model). Except * $P < 0.01$, uncorrected for multiple comparisons.

Discrepancy between Reward Magnitudes (Probabilities Matched)

Second, since the sustained dopaminergic activity that occurs with reward uncertainty ($P = 0.5$) is highest with the maximal discrepancy between potential reward magnitudes (the stimulus predicting a small or large reward had higher median change in sustained activity than the stimulus predicting a small or medium reward magnitude) (Fiorillo *et al.*, 2003), we investigated whether a similar effect of maximal discrepancy between potential reward magnitudes would be obtained in humans on the sustained mode of dopaminergic activity during the delay period, using the contrast: $\text{Delay}_{\text{Slot}_B} > \text{Delay}_{\text{Slot}_C}$, i.e. [$\text{Delay}: P = 0.5 \$20, P = 0.5 \0] $>$ [$\text{Delay}: P = 0.5 \$10, P = 0.5 \0].

Expected Reward Value Matched between Slot Machines

Third, we investigated the effect of reward probability/magnitude on the sustained mode of dopaminergic activity by comparing the two slot machines with equal expected reward value in the contrasts: ($\text{Delay}_{\text{Slot}_C} < \text{Delay}_{\text{Slot}_A}$, i.e. [$\text{Delay}: P = 0.5 \$10, P = 0.5 \0] $<$ [$\text{Delay}: P = 0.25 \$20, P = 0.75 \0]) and ($\text{Delay}_{\text{Slot}_C} > \text{Delay}_{\text{Slot}_A}$, i.e. [$\text{Delay}: P = 0.5 \$10, P = 0.5 \0] $>$ [$\text{Delay}: P = 0.25 \$20, P = 0.75 \0]). The motivations underlying these comparisons were: (i) to investigate whether the brain regions activated with maximal reward uncertainty are still activated when comparing two slot machines with equal expected reward value; (ii) the knowledge that the monkey electrophysiological study manipulated reward probability and magnitude in two separate experiments but left open the issue of whether the sustained mode of activity is more responsive to reward probability or magnitude when controlling the expected reward value.

Plots of Event Time-courses

In addition to the random effects analysis previously described, we also performed two fixed-effect analyses in order to plot the time-courses of the events. The first fixed-effect analysis included as covariates the four slot machines at S1, regardless of whether they would be rewarded or not. This fixed-effect analysis was used to plot the time course of activation in the midbrain dopaminergic region that covaried with the error prediction signal at S1 with higher reward probability (Fig. 3*b*, left portion) and for plotting the time-course of activation in the ventral striatum with reward uncertainty (Fig. 6*c*).

The second fixed analysis only included as covariates an impulse response function at S1 that included the three conditions in which the 'slot machines' are rewarded and the control slot machine ($P = 1$ to receive $\$0$). This fixed effect analysis was used to plot the time-course of activation in the midbrain dopaminergic region that covaried with

lower reward probability at the time of the outcome S2 for rewarded trials (Fig. 3*b*, right portion).

Results

Behavioral Data

Before scanning, subjects underwent a training session in which they learned the association between each slot machine and a specific response button (see Materials and Methods and Fig. 1). During scanning, as expected from the simplicity of the task, subjects performed at ceiling, both at the time of the cue S1 and at the time of the outcome S2. The percentages of correct responses made at S1 were 98.6% (SD.02), 99.5% (SD.01), 98.4% (SD.02) and 99.3% (SD.02) for the slot machines A–D respectively. The percentages of correct responses made at the time of the outcome S2 were: 99.2% (SD.02) for the no reward delivery and 98.6% (SD.05) for the reward delivery of the slot A, 99.7% (SD.1) for the no-reward and 100% for the reward of the slot B, 99.4% (SD.02) for the no-reward and 100% for the reward of the slot C and 99.6% (SD.02) for the slot D.

Response times (RTs) at the time of S1 and S2, analyzed for correct response only, are depicted in Figure 2. Repeated-measures ANOVA with the four types of slot machines at S1 revealed no significant RT difference between conditions [$F(3,30) = 1.7, P = 0.2$]. Response times at S2 were analyzed by performing a 3×2 repeated-measures ANOVA that included the slot machines A, B and C and their two possible outcomes (reward/no-reward). There was a main effect of outcome [$F(1,30) = 5.3, P < 0.05$], subjects responding faster when they were rewarded than when they were not. There was no difference between types of slot machines [$F(2,60) = 1.6, P = 0.2$] and no interaction between factors [$F(2,60) = 0.7, P = 0.5$].

Functional Imaging Data

Specific predictions can be made on the basis of electrophysiological recordings from midbrain dopaminergic neurons in monkeys. In those single-cell recordings, after learning, higher

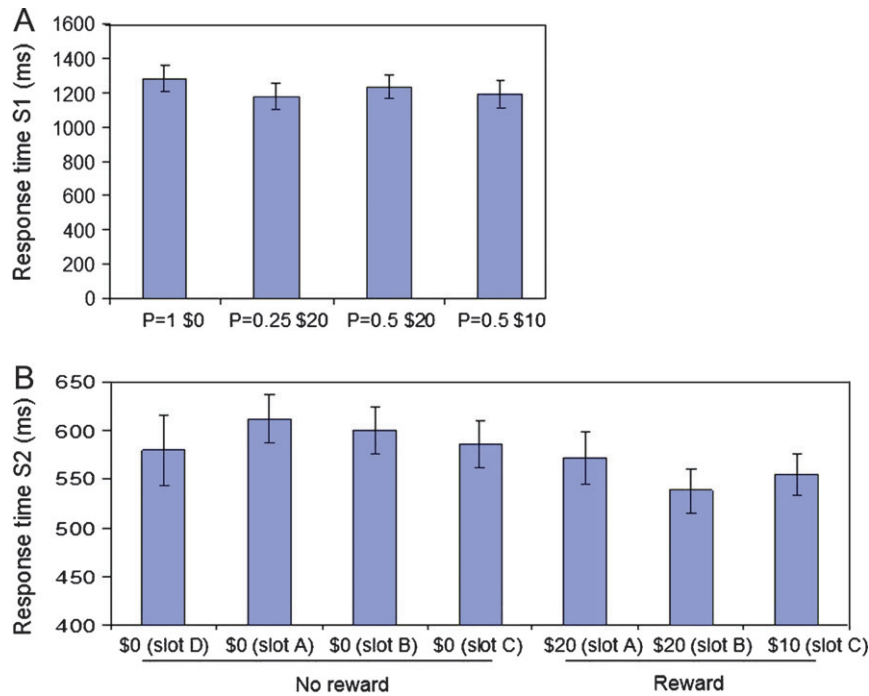


Figure 2. Behavioral results. (A) Mean response time (RT) across subjects at the time of the cue S1 in the different 'slot machine' conditions. Error bars represent \pm SEM. There was no significant difference in RT between the four types of slot machines. (B) Mean response time at the time of the outcome S2 (rewarded and non-rewarded). There was a significant effect of outcome, i.e. subjects responded faster for the rewarded slot machines than for the non-rewarded ones.

reward probability induces larger phasic response of midbrain neurons at the time of the reward-predicting stimulus and less activity at the time of the outcome. Moreover, the sustained mode of activity increases with reward uncertainty during the delay period between the conditioned stimulus and the outcome (Fiorillo *et al.*, 2003). As predicted by these electrophysiological findings, our fMRI results showed that the midbrain region responded transiently to both higher reward probability at the time of the cue (comparison: $S1_{\text{Slot}_B} > S1_{\text{Slot}_A}$, extent cluster $k = 3$) and to lower reward probability at the time of the rewarded outcome (comparison: rewarded outcome $S2_{\$20_Slot_A} > S2_{\20_Slot_B} , extent cluster $k = 19$), and in a sustained fashion to reward uncertainty during the delay period ($P < 0.005$, uncorrected; comparison: $\text{Delay}_{\text{Slot}_B} > \text{Delay}_{\text{Delay_Slot}_A}$, extent cluster $k = 20$; Fig. 3 and Table 2). As also predicted, the sustained activity observed in the midbrain region with maximal reward uncertainty was higher with greater discrepancy between potential reward magnitudes ($P < 0.05$ uncorrected; comparison: $\text{Delay}_{\text{Slot}_B} > \text{Delay}_{\text{Slot}_C}$, MNI coordinates: $x, y, z = -4, -19, -19$, $Z = 2.1$, extent cluster $k = 9$). Note that a liberal threshold was used for this specific comparison because it was motivated by an *a priori* hypothesis based on the monkey study.

It could be proposed that the midbrain sustained activity only reflects an increase in expected reward value. This interpretation cannot be ruled out from the monkey electrophysiological data because in that study reward probability and reward magnitude were varied in two separate experiments without controlling the expected reward value (thus expected reward value and reward probability were confounded in the first experiment and expected reward value and reward magnitude were confounded in the second experiment). Our experiment allowed us to reject this interpretation because when comparing two conditions with equal expected reward value but

variable probability and magnitude, the midbrain region was more robustly activated in anticipation of an uncertain reward (50% chance) with low magnitude than in anticipation of a reward with known low probability (25% chance) but higher magnitude ($P < 0.005$ uncorrected; comparison: $\text{Delay}_{\text{Slot}_C} > \text{Delay}_{\text{Slot}_A}$; $x, y, z = 0, -8, -19$, $Z = 3.1$, extent cluster $k = 18$). Thus, the sustained midbrain activity was not due to an increase in expected reward value alone. It should be noted that the spatial resolution of fMRI does not allow us to specify which midbrain nuclei are involved. However, the MNI coordinates of the peaks of our midbrain activity fall precisely in the substantia nigra according to the atlas of Lucerna *et al.* (2002).

The same statistical comparison of functional activity maps that revealed activity in the midbrain region with the error prediction signal and the reward uncertainty signal also allowed us to investigate whether activity in distinct post-synaptic midbrain dopaminergic projection sites covaries with the two types of signals broadcast by midbrain neurons. At the time of the cue, higher reward probability transiently activated a frontal network that included the left inferior frontal gyrus, left dorsolateral prefrontal cortex (DLPFC), medial part of the superior frontal gyrus, as well as the right orbitofrontal cortex and anterior cingulate cortex (ACC) ($P < 0.005$, uncorrected; comparison: $S1_{\text{Slot}_B} > S1_{\text{Slot}_A}$; Fig. 4a and Table 2). Activation was also observed to a lesser extent in the intra-parietal region bilaterally. At the time of the rewarded outcome, a similar frontal network, consisting of the left inferior frontal gyrus, left DLPFC and medial part of the superior frontal gyrus covaried with lower reward probability ($P < 0.005$, uncorrected; comparison: rewarded outcome $S2_{\$20_Slot_A} > S2_{\20_Slot_B} ; Fig. 4b and Table 2). Moreover, all these frontal brain regions were specifically involved with the reward error prediction signal because they were not significantly activated by reward uncertainty during the delay ($P > 0.05$, uncorrected) and were

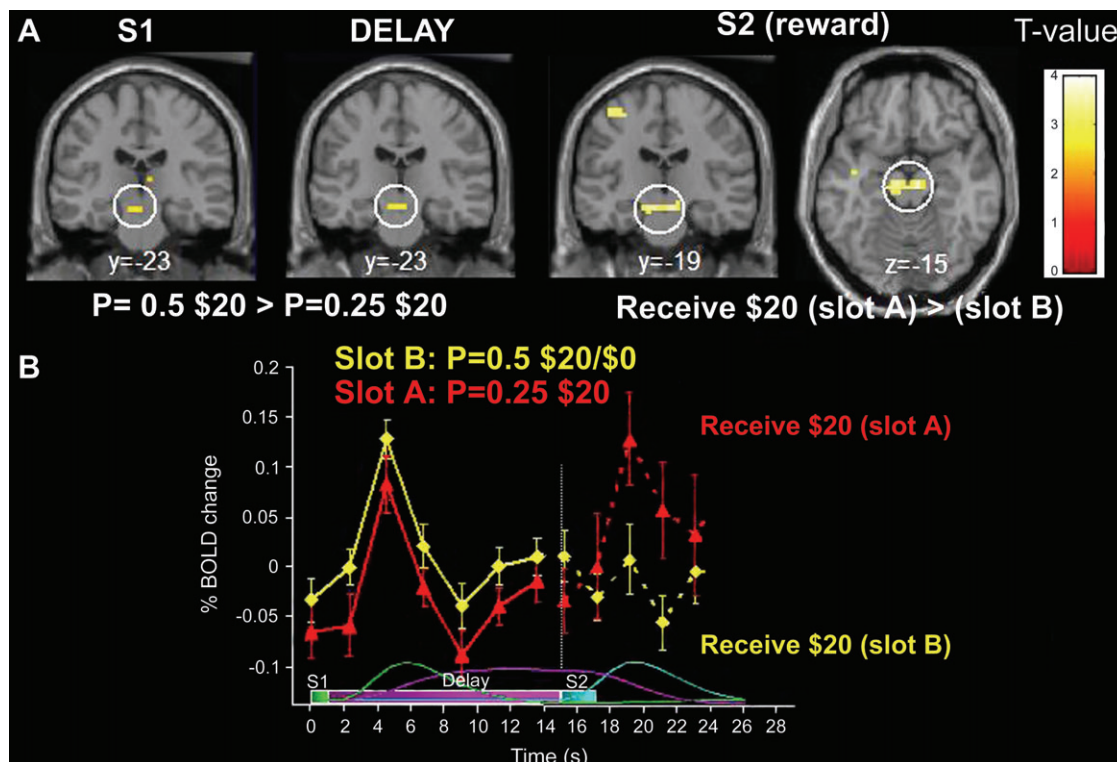


Figure 3. Transient and sustained midbrain activities. (A) Left: Location of transient midbrain responses covarying with the error prediction signal at the cue S1 and (right) the rewarded outcome S2; middle: location of sustained midbrain activity covarying with the reward uncertainty signal during the delay, each superimposed on a structural scan indexed by MNI coordinates (random-effects model, $P < 0.005$ uncorrected). Consistent with electrophysiological recordings (Fiorillo *et al.*, 2003), the human midbrain region was transiently activated with higher reward probability at the cue S1 ($S1_{\text{Slot B}} > S1_{\text{Slot A}}$) and with lower reward probability at the rewarded outcome S2 ($S2_{\$20, \text{Slot A}} > S2_{\$20, \text{Slot B}}$). Moreover, the midbrain region showed higher sustained activity with reward uncertainty during the delay ($\text{Delay}_{\text{Slot B}} > \text{Delay}_{\text{Slot A}}$). (B) Left portion: Time-course of activity (\pm SEM) of the midbrain for slot A ($P = 0.25 \$20$, red) and slot B ($P = 0.5 \20, yellow) for a spherical ($r = 5\text{mm}$) region of interest centered at the voxel maximally activated at the cue ($S1_{\text{Slot B}} > S1_{\text{Slot A}}$) obtained from a fixed-effect analysis that differentiated only between the four slot machines at S1, whether or not rewarded at S2. Right portion: Time-course of activity for rewarded trials only for slot A (dashed red) and slot B (dashed yellow) at the peak of activation at the rewarded outcome ($S2_{\$20, \text{Slot A}} > S2_{\$20, \text{Slot B}}$). The color bar at the bottom shows the actual timing of the three phases of a trial: S1, Delay and S2. Separate regressors were used to independently model each of these phases in the random effects model used for the statistical maps shown in (a). Green, pink and blue lines show hemodynamic functions convolved with the boxcar of each phase assuming a 4–6 s lag in the response.

significantly more activated in association with these phasically modeled responses than in association with a sustained modeled response related to reward uncertainty during the delay period ($P < 0.001$, uncorrected) (Fig 4c). In order to investigate which brain regions are commonly activated with the error prediction signal at the cue and the rewarded outcome, we also performed a conjunction analysis of the brain regions commonly activated in a transient fashion with higher reward probability at the time of the cue and with lower reward probability at the time of the rewarded outcome S2. This additional analysis activated a network that included the mid-brain dopaminergic region, the left lateral prefrontal cortex, the right orbitofrontal cortex and the intra-parietal cortex bilaterally ($P < 0.005$, uncorrected; Table 2).

In addition, at the time of a more unexpected reward omission that lead to a negative prediction error response ($P < 0.005$ uncorrected; comparison: non-rewarded outcome $S2_{\$0, \text{Slot B}} > S2_{\$0, \text{Slot A}}$), activation was found in a cortical network similar to the one observed with the positive prediction error response. This network was composed of the inferior frontal gyri bilaterally, the left dorsolateral prefrontal cortex as well as the right orbitofrontal cortex/fronto-polar cortex and the inferior parietal lobule bilaterally (Fig. 5; Table 3).

Finally, we investigated which post-synaptic regions respond to the reward uncertainty signal during the delay (comparison: $\text{Delay}_{\text{Slot B}} > \text{Delay}_{\text{Slot A}}$). We found that the ventral

striatum (putamen) showed sustained activation that covaried with maximal reward uncertainty during reward anticipation ($P < 0.005$, uncorrected; Fig. 6). Additional comparisons demonstrated that this brain region was specifically activated with the reward uncertainty signal during the delay period because it did not show significant transient response to higher reward probability at the time of the cue ($P > 0.1$, uncorrected) or to lower reward probability at the time of the rewarded outcome ($P > 0.1$, uncorrected) and was significantly more activated with reward uncertainty during the delay period than with the error prediction signal at the time of the cue ($P < 0.05$, small volume correction for a sphere of 20 mm centered at the maximum of activity). Moreover, the sustained ventral striatum activity cannot be attributed to increased expected reward value alone because it was still present when comparing two slot machines with equal expected reward value (comparison $\text{Delay}_{\text{Slot C}} > \text{Delay}_{\text{Slot A}}$, $P < 0.05$ corrected for small volume in a 20 mm sphere centered at the peak of ventral striatum activity). Taken together, these results indicate that the ventral striatum is specifically activated with the reward uncertainty signal during the delay period.

Discussion

Our fMRI findings showed that midbrain region responded both to reward uncertainty in a sustained fashion during the delay

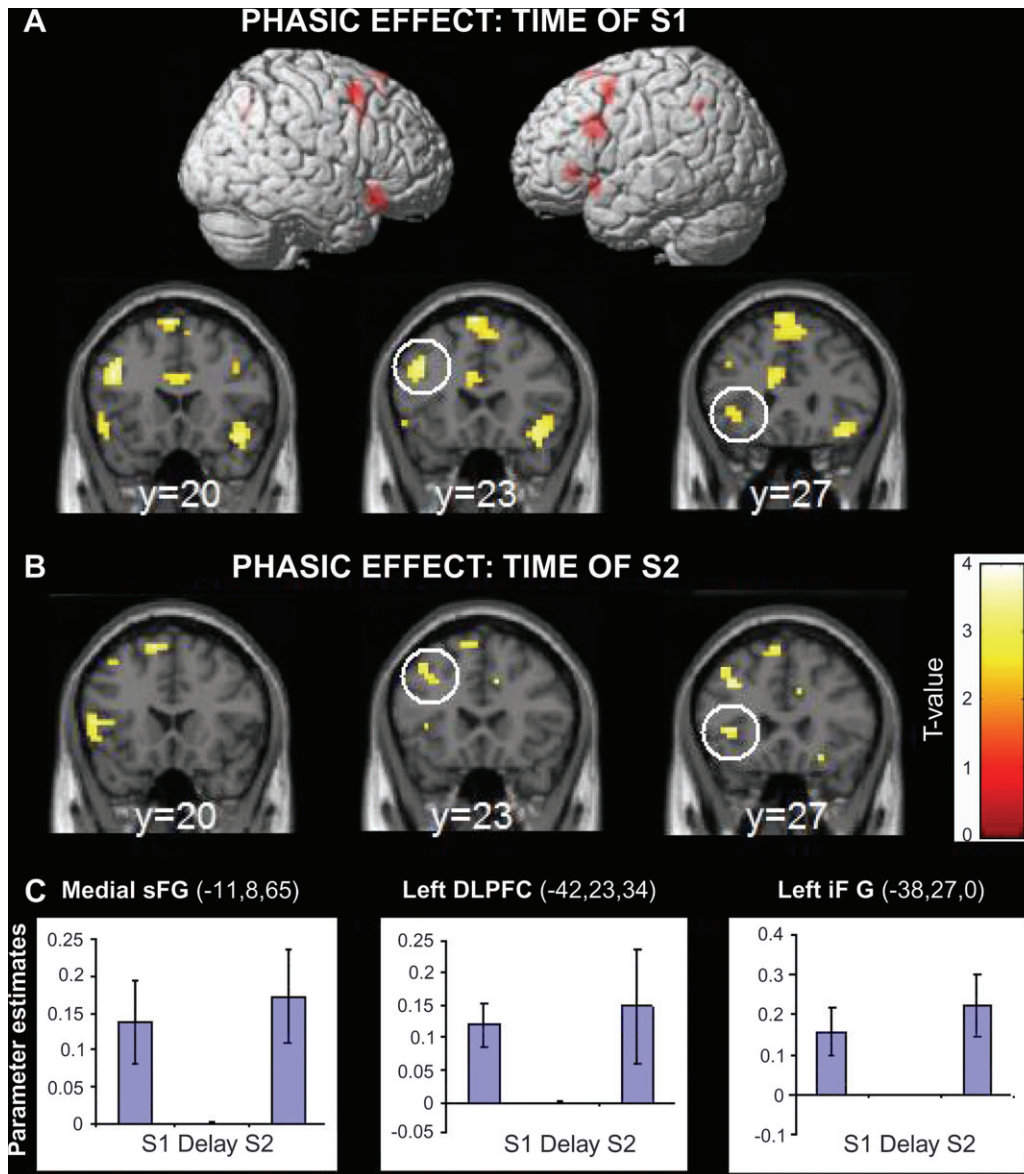


Figure 4. Brain regions covarying with the transient positive error prediction signal at S1 and S2. (A) Location of transient activation in left inferior frontal gyrus (iFG), left DLPFC, medial superior frontal gyrus (sFG), right orbitofrontal and anterior cingulate cortices covarying with higher reward probability at the time of the cue S1 ($S1_{Slot_B} > S1_{Slot_A}$) overlaid onto a 3D-rendered brain (top) and on a coronal section (bottom). (B) Location of transient left iFG, left DLPFC and medial sFG activations covarying with lower reward probability at the time of the rewarded outcome S2 ($S2_{Slot_A} > S2_{Slot_B}$). White circles in (a) and (b) indicate common iFG and DLPFC activities at S1 and S2. (C) Parameter estimates at the activity peak of left DLPFC, left iFG and medial sFG commonly activated in a conjunction analysis of brain regions responding transiently to higher reward probability at S1 and to lower reward probability at S2. Parameter estimates plotted at S1, during the delay and at S2 respectively concern the comparisons: $S1_{Slot_B} > S1_{Slot_A}$, $Delay_{Slot_B} > Delay_{Slot_A}$ and $S2_{Slot_A} > S2_{Slot_B}$.

and transiently to the error prediction signal (Fig. 3), extending in humans the distinction between sustained and transient dopamine activity dynamics observed in monkeys in similar experimental settings. More importantly, we found activities in different midbrain projection sites that covary with the two types of signals broadcast by dopaminergic neurons: a frontal network showed a transient activation that covaried with the reward error prediction signal both at the time of the cue and at the time of the outcome (Figs 4 and 5), while the ventral striatum (putamen) showed sustained activation that covaried with maximal reward uncertainty during reward anticipation (Fig. 6). These results have important functional consequences because they indicate distinct roles for two dopamine-related

networks in processing different reward information signals. In the following two sections, we further discuss the meanings of these findings.

Both Transient and Sustained Activities Are Observed in Human Midbrain

The fact that the human midbrain responded transiently to both higher reward probability at the cue and to lower reward probability at the rewarded outcome, and in a sustained fashion to reward uncertainty during the delay period (Fig. 3), supports a unified cross-species view in which midbrain dopaminergic neurons obey common basic principles of neural computation.

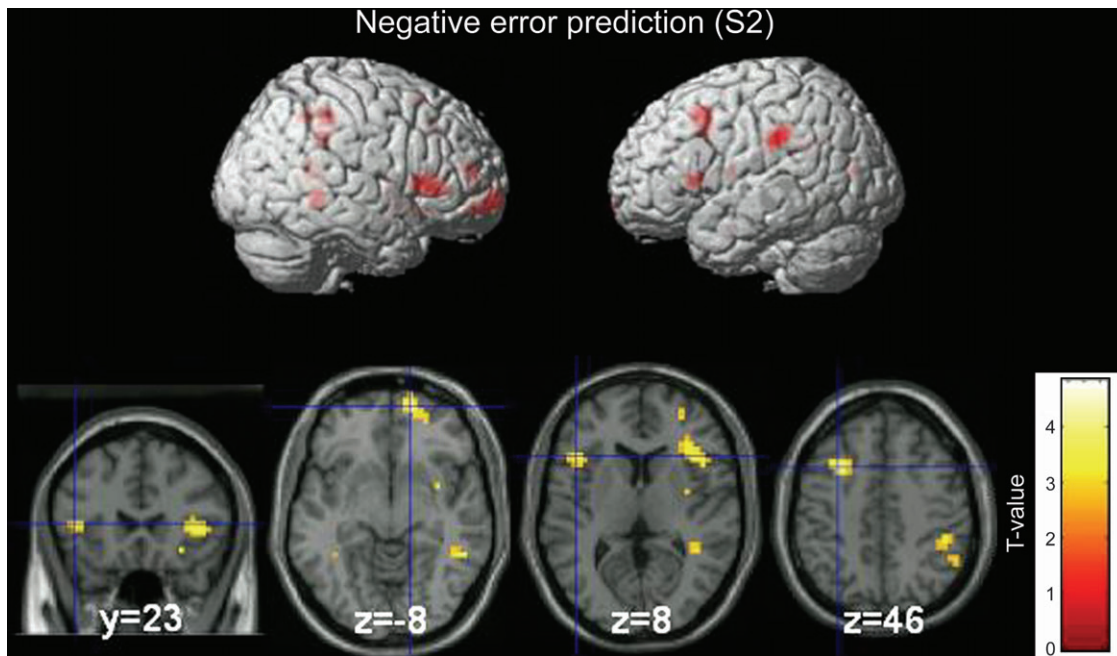


Figure 5. Brain regions covarying with the transient negative error prediction signal at S2. Location of transient activation in inferior frontal gyri, left dorsolateral prefrontal cortex, right orbitofrontal cortex/fronto-polar cortex and bilateral inferior parietal cortex covarying with the negative error prediction signal at the time of non-reward delivery at S2 ($S2_{\$0_Slot_B} > S2_{\0_Slot_A}) overlaid onto a 3D-rendered brain (top) and on a coronal and axial sections (bottom).

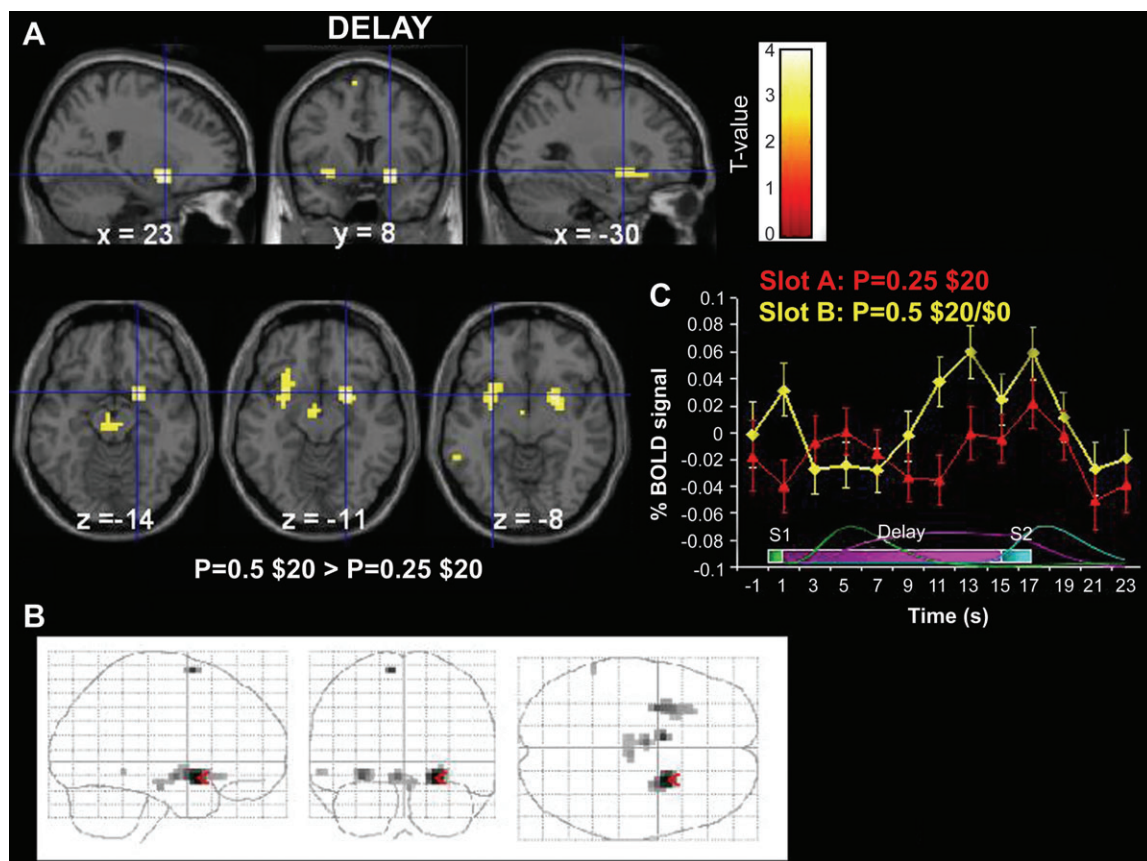


Figure 6. Brain regions covarying with the sustained reward uncertainty signal. (A) Location of sustained bilateral ventral striatum activations covarying with reward uncertainty during the delay period ($Delay_{Slot_B} > Delay_{Slot_A}$). (B) Glass brain views illustrating that this comparison only activated focal sub-cortical regions (midbrain and ventral striatum). (C) Time-course of activation in the same comparison at the peak of right ventral striatum activity for slot machines A (red) and B (yellow).

Table 3

Foci of activations with the negative error prediction signal at the time of non-reward delivery at S2 ($P < 0.005$, uncorrected)

Anatomical structure (Brodmann's area)	Error prediction at S2 (no reward): $P = 0.75 \$0 > P = 0.5 \0		Error prediction at S2 (no reward): $P = 0.5 \$0 > P = 0.75 \0	
	Peak MNI coordinates, x y z	Z-value	Peak MNI coordinates, x y z	Z-value
Left middle FG (BA 6)			−30 15 46	4.53
Left inferior frontal gyrus (BA 45)			−46 23 8	3.25
Right inferior frontal gyrus (BA 45)			42 23 4	3.51
Right middle frontal gyrus (BA 8)			23 30 38	2.78
Right OFC/fronto-polar cortex			15 61 −8	4.11
Right inferior temporal gyrus (BA 37)			46 −42 −4	3.51
Left inferior temporal gyrus (BA 37)			−38 −46 −8	2.58
Left inferior parietal lobule (BA 40)			−57 −30 30	3.15
Right inferior parietal lobule (BA 40)			42 −38 30	2.90
Thalamus	−8 15 8	3.99		
Posterior cingulate cortex	−11 −46 11	3.26		
Left premotor cortex	−30 −15 68	3.22		
Anterior cingulate cortex	−4 −8 46	3.56		

These results are particularly striking because they were obtained with a more extended time-course than the one used in the monkey electrophysiological experiment. In addition, the sustained activity of the midbrain region, observed when reward uncertainty was maximal, was higher with the greatest discrepancy between potential reward magnitudes. Moreover, this sustained midbrain activity was not due to an increase in expected reward value alone (i.e. reward probability \times magnitude). Indeed, when comparing two conditions with equal expected reward value but variable probability and magnitude, the midbrain was more robustly activated in anticipation of a reward with maximal uncertainty and low magnitude than in anticipation of a reward with lower probability but higher magnitude. The result of this latter comparison also indicates that the sustained mode of midbrain activity is more sensitive to reward uncertainty than to reward magnitude. This suggests that, in the context of the range of our reward probabilities, the sustained midbrain activity may code the expectation of reward information in terms of the information theory (Shannon, 1948). According to this theory, the more uncertain the outcome (reward or no reward), the more information the outcome contains (and information is available in the outcome only in the presence of uncertainty). This parallels findings that the central nervous system uses probabilistic models in domains such as object perception (Kersten *et al.*, 2004), sensorimotor learning (Kording and Wolpert, 2004) and cognitive control (Dreher *et al.*, 2002b; Koechlin *et al.*, 2003), and suggests that the neural computations regarding reward information may share common mathematical principles of information theory with these domains.

Thus, our findings shed new light on the function of the sustained mode of neuronal activity in the midbrain region observed with reward uncertainty by suggesting that it reflects the expectation of reward information in terms of information theory. The sustained reward uncertainty signal may be useful for promoting exploratory behavior in the balance that an

organism must strike between exploring the environment to gain new information and exploiting existing knowledge in order to obtain a reward (the latter involving the phasic error prediction signal) (Dayan and Balleine, 2002). A computational theory explaining the two modes of midbrain dopaminergic activities remains to be developed and may benefit from models proposing that dopamine acts at different time-scales for selecting, maintaining or suppressing currently active representations (Dehaene and Changeux, 2000; Dreher and Burnod, 2002; Dreher *et al.*, 2002a). It is clear that actual temporal difference models of dopamine discharge, relying on the assumption of the maximization of total future reward, cannot account for the motivational properties of dopaminergic neurons (Dayan and Balleine, 2002). Future models of dopaminergic discharge will need to include a motivational component, taking into account the cost and benefit of performing an action (Sato *et al.*, 2003; Salamone *et al.*, 2005) and will need to clarify the nature of the sustained dopaminergic signal (Daw *et al.*, 2005; Niv *et al.*, 2005).

Other potential interpretations of the sustained mode of midbrain activity that we observed, such as motivation or attention alone, are unlikely because electrophysiological results indicate that when monkeys are more motivated/attentive to obtain a reward when the reward probability is $P = 0.75$ relative to $P = 0.5$ (as evidenced by an increase in their licking behavior in anticipation of the reward) there is an associated decrease of the sustained mode of activity (Fiorillo *et al.*, 2003). Furthermore, these interpretations do not explain the increased sustained midbrain activity observed when comparing the two slot machines with equal expected reward values.

It should be noted that we did not observe midbrain activation with the negative prediction error signal when no reward was delivered, which is consistent with the majority of previous reward fMRI studies that investigated brain regions responding with this signal (Fletcher *et al.*, 2001; Pagnoni *et al.*, 2002; McClure *et al.*, 2003; O'Doherty *et al.*, 2003; Ramnani *et al.*, 2004). To our knowledge, only one study (Aron *et al.*, 2004) reported midbrain activation with a form of negative prediction error (when comparing negative with positive feedback). This fMRI study investigated brain regions involved during a probabilistic classification learning task in which subjects presented with a stimulus had to decide between two responses followed by one of two potential feedback images presented after a delay. The midbrain region was also found activated, relative to a cross fixation, both at the times of stimulus presentation and of feedback, but not during the delay period. Moreover, a ROI analysis showed an activation of this brain region that correlated with the degree of uncertainty during the delay. Although interesting, comparisons with our study are difficult because the Aron *et al.* study concerned a learning paradigm and was designed neither to reveal brain regions activated in a sustained fashion nor to distinguish between the transient and sustained components of the reward system (since the short delay duration of mean = 2 s prevented such distinction). In addition, brain activation observed at the time of feedback might have been influenced by the pseudo-random timing of feedback occurrence since dopaminergic neurons are sensitive to reward timing uncertainty. Finally, the decision required at the time of stimulus presentation and the perception of a contingency between this presentation and feedback to the response probably influences the reward

brain system (Elliott *et al.*, 2004; O'Doherty *et al.*, 2004; Tricomi *et al.*, 2004).

Different Networks Are Involved with Transient versus Sustained Reward Information Signals

Frontocortical Regions Respond Transiently to the Reward Error Prediction

We found transient activations in a frontal network that covaried with the reward error prediction signal both at the time of the cue and at the time of the outcome (Figs 4*a,b* and 5). This frontal network was specifically involved with the reward error prediction signal because it was not significantly activated by reward uncertainty during the delay and was significantly more activated in association with these phasically modeled responses than in association with a sustained modeled response related to reward uncertainty during the delay period (Fig. 4*c*). These results extend previous fMRI reports that DLPFC, inferior frontal gyrus and orbitofrontal cortex activity correlates with an error prediction signal related to abstract stimulus-response associations or taste reward, although some of these studies focused more on ventral striatum activity (Berns *et al.*, 2001; Fletcher *et al.*, 2001; O'Doherty *et al.*, 2003; Corlett *et al.*, 2004; Paulus *et al.*, 2004). The lateral prefrontal cortex may generate the reward prediction because neurons from this brain region represent predictions about expected rewards according to the context (Kobayashi *et al.*, 2002; Watanabe *et al.*, 2002). Alternative, and non-exclusive, interpretations of the lateral prefrontal cortex activation include an increase in attention (at the cue and the rewarded outcome), the cognitive representation of possible outcomes (at the cue) and the hedonic consequences of receiving a reward (at the time of the reward) (Kringelbach *et al.*, 2004). Our lateral prefrontal cortex activation is unlikely to result from selection of response or maintaining stimulus-response association in working memory because these two processes were equally present for each slot machine, and were therefore subtracted in the comparison revealing the error prediction.

The ACC activity also provides fMRI evidence that converges with monkey electrophysiological findings (Schall *et al.*, 2002) and human electrophysiological scalp recordings of the involvement of this brain region with a reward prediction error signal (Holroyd *et al.*, 2002). A recent theory proposed that the ACC uses the reward prediction error signal sent from the midbrain dopaminergic neurons to reinforce adaptive behaviors (Holroyd *et al.*, 2002). This theory holds that the ACC is activated both by internal (from an efference copy of the response command) and external (from feedback in the outside environment) sources of unexpected error information. Confirming this theory, when subjects learn to select between two responses by trial-and-error using feedback stimuli that indicate monetary gains and losses, both error responses and error feedback activate the same ACC region (Holroyd *et al.*, 2004).

It is worth noting that, when subjects have to learn an association between a conditioned stimulus and sweet juice, some of the fMRI studies mentioned above (Fletcher *et al.*, 2001; McClure *et al.*, 2003; O'Doherty *et al.*, 2003; Paulus *et al.*, 2004) have also reported activation of the ventral striatum/putamen with the reward error prediction. These learning studies were neither designed to dissociate sustained from transient neural

activation nor to identify sustained activity correlating with reward uncertainty. Thus, their results may include sustained activity dynamics correlating with reward uncertainty. Moreover, the fact that nothing had to be learned in our task might explain why the ventral striatum/putamen activation did not covary significantly with the error prediction signal.

We also found that the intra-parietal region was activated with the error prediction signal, which was also reported—but not highlighted—in a previous study (O'Doherty *et al.*, 2003). This result is interesting because lateral intraparietal (LIP) neurons may code a map of the relative value of each movement relative to the available alternative when monkeys perform decision-making tasks (LIP neurons would represent both the magnitude of the reward expected from an eye-movement and the probability that a particular response will be rewarded) (Platt and Glimcher, 1999). Here, we show in humans that the intra-parietal region is activated with the error prediction *independently of any decision*.

The Ventral Striatum Covaries with Reward Uncertainty During Reward Anticipation

The ventral striatum (putamen) showed sustained activation that covaried with maximal reward uncertainty during reward anticipation (Fig. 6). This brain region was specifically activated with the reward uncertainty signal during the delay period because it did not show significant transient response to higher reward probability at the time of the cue or to lower reward probability at the time of the rewarded outcome and was significantly more activated with reward uncertainty during the delay period than with the error prediction signal at the time of the cue. Moreover, the sustained ventral striatum activity cannot be attributed to increased expected reward value alone because it was still present when comparing two slot machines with equal expected reward value. These findings are consistent with the ventral striatum playing a role in the expectation of reward information and motivation (Apicella *et al.*, 1992; Berridge and Robinson, 1998; Shidara *et al.*, 1998). For example, electrophysiological studies report that when monkeys perform sequences of behaviors that must be completed successfully before reward delivery, the increased reward information and motivation as the rewarded trial approaches are reflected in ventral striatal neuronal activity (Shidara *et al.*, 1998). More importantly, our results may elucidate controversies regarding the precise function of distinct components of the dopaminergic system in reward information processing. Indeed, one important alternative functional hypothesis to the error prediction theory (Schultz *et al.*, 1997) proposes that the dopaminergic system (especially the neostriatum and nucleus accumbens) mediates the incentive value of reward (Berridge and Robinson, 1998). Our results offer a means to reconcile these two theories by suggesting that the ventral striatum reflects the expectation of reward information or the incentive value of reward while the network of prefrontal cortex regions may generate the reward prediction.

It has been proposed that gambling, with its intrinsic reward uncertainty characteristic, has reinforcing properties that may share common mechanisms with addictive drugs (Fiorillo *et al.*, 2003). This proposal is especially interesting in light of recent reports indicating a specific role of the dopaminergic innervation to the striatum in well-established drug-seeking behavior (Bradberry, 2000; Ito *et al.*, 2002). Our finding also offers an account for previous reports of human ventral striatum

activation during anticipation of monetary and taste reward as coding, at least in part, the expectation of reward information (Breiter *et al.*, 2001; Knutson *et al.*, 2001; O'Doherty *et al.*, 2002). This signal could gain access to striatal neurons through ascending dopamine fibers as well as structures implicated in the evaluation of the motivational significance of stimuli, especially the amygdala and the orbitofrontal cortex (Nishijo *et al.*, 1988; Schoenbaum *et al.*, 1998). Alternative interpretations of our putamen activation, such as preparation of the motor response or duration perception, are unlikely because these processes were equally present in each condition and were therefore subtracted in the comparisons. Furthermore, the absence of conditioning between cues and reward in our paradigm precludes attribution of our putamen activation to a learning effect.

Our finding of two networks covarying with different reward information signals may indicate that dopaminergic projection sites can distinguish the two signals. It is also possible that these targets show independent transient (prefrontal cortex) and sustained (ventral striatum) activities related to the two signals and/or that they help to shape dopaminergic neuronal activity by differentially modulating their phasic and sustained modes of firing, which occur independently in individual neurons (Fiorillo *et al.*, 2003). This latter hypothesis is supported by anatomical observations that different populations of dopaminergic neurons are innervated predominantly by the target areas to which they project, or by the regions that, in functional terms, are the most closely linked to the target area (Sesack *et al.*, 2003). For example, in rodents, dopaminergic neurons projecting to the prefrontal cortex receive direct reciprocal inputs from this brain region, but not from the striatum, while dopaminergic neurons projecting to the striatum receive afferents from that brain region, but not from the prefrontal cortex, thereby forming two projection systems (Sesack *et al.*, 2003). This suggests a general principle for midbrain dopaminergic neuronal afferent regulation: namely that the prefrontal cortex and the striatum are responsible for regulating and controlling different modes of dopaminergic neuronal firing.

It should be noted that the blood oxygenation level-dependent (BOLD) signal observed with fMRI is not sensitive to dopamine release *per se* but reflects the input and intracortical processing of a given area rather than its spiking output (Logothetis *et al.*, 2001). Thus, detection of differential BOLD responses in midbrain dopaminergic region may reflect afferent inputs and local neural activity rather than corresponding directly to dopaminergic output. For the same reason, the fact that dopamine is released in the striatum and the prefrontal cortex following firing of dopaminergic neurons does not necessarily imply that these post-synaptic dopaminergic sites should show a differential BOLD response with fMRI. This may explain why prefrontal activation did not covary significantly with anticipation of maximal reward uncertainty and why striatal activation did not covary with the error prediction signal, even though it is known from single unit monkey recordings that both transient and sustained signals are observed throughout the brain, and in particular in the prefrontal cortex and the striatum.

Conclusion

We have reported a number of findings that shed new light on the functional properties of various components of the reward

system in humans. We demonstrated that the midbrain region exhibits both transient and sustained activities and that the lateral prefrontal cortex and the ventral striatum respectively are engaged in a specific fashion with two signals related to distinct statistical properties of reward information. Since the mesolimbic/nigrostriatal dopaminergic pathways develop earlier than the mesocortical pathway in evolution, it may be proposed that the expectation of reward information, which may promote exploratory behavior, occurred phylogenetically earlier than the ability to learn from reward prediction error, which may have appeared more progressively. Our approach may be fruitfully extended towards an understanding of diseases involving dysfunctions of the dopaminergic system, such as schizophrenia and Parkinson's disease as well as pathologies of reward processing (such as gambling and addiction).

Notes

We thank Paul Koch, Rosanna Olsen and Deepak Sarpal for help scheduling and scanning the subjects and Dr Daniel Weinberger and Dr Etienne Koechlin for comments on early version of the manuscript.

Address correspondence to J.-C. Dreher, Institut des Sciences Cognitives, Centre National de la Recherche Scientifique (CNRS), 67 Bd Pinel, 69675 Bron, France. Email address: dreher@isc.cnrs.fr.

References

- Apicella P, Scarnati E, Ljungberg T, Schultz W (1992) Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J Neurophysiol* 68:945-960.
- Aron AR, Shohamy D, Clark J, Myers C, Gluck MA, Poldrack RA (2004) Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *J Neurophysiol* 92:1144-1152.
- Berns GS, McClure SM, Pagnoni G, Montague PR (2001) Predictability modulates human brain response to reward. *J Neurosci* 21:2793-2798.
- Berridge KC, Robinson TE (1998) What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Brain Res Rev* 28:309-369.
- Bradberry CW (2000) Acute and chronic dopamine dynamics in a non-human primate model of recreational cocaine use. *J Neurosci* 20:7109-7115.
- Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses. *Neuron* 30:619-639.
- Corlett PR, Aitken MR, Dickinson A, Shanks DR, Honey GD, Honey RA, *et al.* (2004) Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* 44:877-888.
- Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and striatal systems for behavioural control. *COSYNE 2005, Computational and Systems Neuroscience*. Salt Lake City, UT.
- Dayan P, Balleine BW (2002) Reward, motivation, and reinforcement learning. *Neuron* 36:285-298.
- Dehaene S, Changeux JP (2000) Reward-dependent learning in neuronal networks for planning and decision making. *Prog Brain Res* 126:217-229.
- Dreher JC, Burnod Y (2002) An integrative theory of the phasic and tonic modes of dopamine modulation in the prefrontal cortex. *Neural Netw* 15:583-602.
- Dreher JC, Guigon E, Burnod Y (2002a) A model of prefrontal cortex dopaminergic modulation during the delayed alternation task. *J Cogn Neurosci* 14:853-865.
- Dreher JC, Koechlin E, Ali SO, Grafman J (2002b) The roles of timing and task order during task switching. *Neuroimage* 17:95-109.
- Elliott R, Newman JL, Longe OA, William Deakin JF (2004) Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *Neuroimage* 21:984-990.

- Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–1902.
- Fletcher PC, Anderson JM, Shanks DR, Honey R, Carpenter TA, Donovan T, *et al.* (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat Neurosci* 4:1043–1048.
- Glimcher PW (2003) The neurobiology of visual-saccadic decision making. *Annu Rev Neurosci* 26:133–179.
- Heberlein K, Hu X (1991) Improved shim by subject head positioning. ISMRM (International Society of Magnetic Resonance in Medicine), 9th Annual Meeting, Glasgow, p.1157.
- Holroyd CB, Coles MG, Nieuwenhuis S (2002) Medial prefrontal cortex and error potentials. *Science* 296:1610–1611.
- Holroyd CB, Nieuwenhuis S, Yeung N, Nystrom L, Mars RB, Coles MG, *et al.* (2004) Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nat Neurosci* 7:497–498.
- Ito R, Dalley JW, Robbins TW, Everitt BJ (2002) Dopamine release in the dorsal striatum during cocaine-seeking behavior under the control of a drug-associated cue. *J Neurosci* 22:6247–6253.
- Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. *Annu Rev Psychol* 55:271–304.
- Knutson B, Adams CM, Fong GW, Hommer D (2001) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci* 21:RC159.
- Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O (2002) Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J Neurophysiol* 87:1488–1498.
- Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. *Science* 302:1181–1185.
- Kording KP, Wolpert DM (2004) Bayesian integration in sensorimotor learning. *Nature* 427:244–247.
- Kringelbach ML, de Araujo IE, Rolls ET (2004) Taste-related activity in the human dorsolateral prefrontal cortex. *Neuroimage* 21:781–788.
- Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412:150–157.
- Lucerna S, Salpietro FM, Alafaci C, Tomasello F (2002) In vivo atlas of deep brain structures. Berlin, Heidelberg, New York: Springer-Verlag.
- McClure SM, Berns GS, Montague PR (2003) Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38:339–346.
- Mellers BA (2000) Choice and the relative pleasure of consequences. *Psychol Bull* 126:910–924.
- Nishijo H, Ono T, Nishino H (1988) Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J Neurosci* 8:3570–3583.
- Niv Y, Duff MO, Dayan P (2005) Dopamine, uncertainty and TD learning. *Behav Brain Func* 1:6.
- O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. *Neuron* 33:815–826.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. *Neuron* 38:329–337.
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
- Pagnoni G, Zink CF, Montague PR, Berns GS (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci* 5:97–98.
- Paulus MP, Feinstein JS, Tapert SF, Liu TT (2004) Trend detection via temporal difference model predicts inferior prefrontal cortex activation during acquisition of advantageous action selection. *Neuroimage* 21:733–743.
- Pearce JM, Hall G (1980) A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87:532–552.
- Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
- Ramnanani N, Elliott R, Athwal BS, Passingham RE (2004) Prediction error for free monetary reward in the human prefrontal cortex. *Neuroimage* 23:777–786.
- Salamone JD, Correa M, Mingote SM, Weber SM (2005) Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Curr Opin Pharmacol* 5:34–41.
- Satoh T, Nakai S, Sato T, Kimura M (2003) Correlated coding of motivation and outcome of decision by dopamine neurons. *J Neurosci* 23:9913–9923.
- Schall JD, Stuphorn V, Brown JW (2002) Monitoring and control of action by the frontal lobes. *Neuron* 36:309–322.
- Schoenbaum G, Chiba AA, Gallagher M (1998) Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci* 1:155–159.
- Schultz W (2000) Multiple reward signals in the brain. *Nat Rev Neurosci* 1:199–207.
- Schultz W (2002) Getting formal with dopamine and reward. *Neuron* 36:241–263.
- Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
- Sesack SR, Carr DB, Omelchenko N, Pinto A (2003) Anatomical substrates for glutamate-dopamine interactions: evidence for specificity of connections and extrasynaptic actions. *Ann NY Acad Sci* 1003:36–52.
- Shannon CE (1948) A mathematical theory of communication. *Bell Sys Tech J* 27:379–423.
- Shidara M, Aigner TG, Richmond BJ (1998) Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J Neurosci* 18:2613–2625.
- Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. *Neuron* 41:281–292.
- Waelti P, Dickinson A, Schultz W (2001) Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48.
- Watanabe M, Hikosaka K, Sakagami M, Shirakawa S (2002) Coding and monitoring of motivational context in the primate prefrontal cortex. *J Neurosci* 22:2391–2400.