Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making

Jan Gläscher, Alan N. Hampton and John P. O'Doherty

Division of Humanities and Social Sciences, California Institute of Technology, MC 114-96, 1200 E. California Boulevard, Pasadena, CA 91125

Considerable evidence has emerged to implicate ventromedial prefrontal cortex in encoding expectations of future reward during value-based decision making. However, the nature of the learned associations upon which such representations depend is much less clear. Here, we aimed to determine whether expected reward representations in this region could be driven by action–outcome associations, rather than being dependent on the associative value assigned to particular discriminative stimuli. Subjects were scanned with functional magnetic resonance imaging while performing 2 variants of a simple reward-related decision task. In one version, subjects made choices between 2 different physical motor responses in the absence of discriminative stimuli, whereas in the other version, subjects chose between 2 different stimuli that were randomly assigned to different responses on a trial-by-trial basis. Using an extension of a reinforcement learning algorithm, we found activity in ventromedial prefrontal cortex tracked expected future reward during the action-based task as well as during the stimulus-based task, indicating that value representations in this region can be driven by action–outcome associations. These findings suggest that ventromedial prefrontal cortex may play a role in encoding the value of chosen actions irrespective of whether those actions denote physical motor responses or more abstract decision options.

Keywords: action value, decision making, reinforcement learning, reversal learning, stimulus value, ventromedial prefrontal cortex

Introduction

A core feature of economic and computational neuroscience models of reward-related decision making is the notion that in order to compute decisions between different available options, it is necessary for the decision agent to encode the expected future reward (expected utility) of those options in order to select the one with the highest expected value (von Neumann and Morgenstern 1944; Montague et al. 1996). Consistent with this notion, there is now considerable evidence from both human neuroimaging studies and animal neurophysiology studies for the existence of expected reward representations in the brain during decision making (O'Doherty 2004; Montague et al. 2006). However, a key outstanding question that has received much less attention to date is the nature of the associative learning mechanisms that drive such representations. In a typical decision problem, subjects make choices between different options that are signified by discriminative stimuli (S) and then must perform an action (A) in order to obtain an outcome (O). In such a decision problem, it is therefore possible that associations between the stimulus and its outcome (S–O), the action and its outcome (A–O), or the stimulus and the associated action (S–A) could in principle be driving expected reward representations and hence contribute toward guiding behavior.

The ventromedial prefrontal cortex (vmPFC) is one region that has received considerable attention for its role in human choice behavior (Bechara et al. 1994) and specifically in encoding expected value representations (Blair et al. 2006; Daw et al. 2006; Hampton et al. 2006). It incorporates the ventral aspects of medial prefrontal cortex and adjacent medial orbitofrontal cortex (Volz et al. 2006). Neuroimaging studies have reported activity in this region scaling approximately linearly with expected value or utility, assessed in a number of ways such as through the application of computational reinforcement learning models to capture trial-by-trial variations in expected reward as a function of experience (Hampton et al. 2006; Kim et al. 2006), through explicit provision of information about expected magnitude or probabilities (Knutson et al. 2005) or by calculating individual subjective preference curves for different decision options (Kable and Glimcher 2007). However, although it is clear on the basis of these past studies that activity in this region relates to expected reward, it is much less clear whether these expectations are stimulus-based, that is, pertaining to the expected value of stimuli that denote a particular decision option, or whether such representations are tied to specific behavioral actions that the subjects must perform in order to choose such options. All the above mentioned studies involved experimental tasks in which subjects had to make a choice between different stimuli. Action–outcome associations were not independently manipulated in these tasks. In a previous functional magnetic resonance imaging (fMRI) study in humans, we and others showed that activity in medial orbitofrontal cortex (OFC) tracks the incentive value of an associated outcome while subjects performed actions in order to obtain reward (Padoa-Schioppa and Assad 2006; Valentin et al. 2007). However, the use of discriminative stimuli in that task means that we could not rule out a contribution for stimulus–outcome associations to the observed activations.

Data from single-unit neurophysiology studies in other primates on the functions of vmPFC is sparse. A few studies have reported action-related value encoding when recording from ventral parts of the medial wall (Matsumoto et al. 2003; Matsumoto et al. 2007). Other studies found only weak evidence for action-related value signals in more dorsal segments of medial prefrontal cortex (Wallis and Miller 2003; Hoshi et al. 2005; Seo and Lee 2007). Neurophysiological studies in primate orbitofrontal cortex typically report expected reward signals corresponding to stimuli but not actions in this area (Thorpe et al. 1983; Tremblay and Schultz

1999, 2000), although such studies have tended to record from central or lateral, but not medial segments of OFC. Thus, the degree of involvement of medial OFC in action-related value encoding is unknown.

A canonical decision problem that has been studied much in both the animal and human reward literature is reversal learning (Clark et al. 2004). In a typical reversal learning task in humans, subjects have to choose between 2 visual stimuli, which at any point are associated with differing probabilities of obtaining reward. One stimulus is designated as the "good" option, in that choice of that stimulus leads to a high probability of a rewarding outcome but only a low probability of a punishing outcome, whereas the other "bad" stimulus is associated with a low probability of reward, but a high probability of a punishing outcome. Through trial and error, subjects must learn to choose the "good" stimulus and avoid the "bad" one. After some time, contingencies reverse so that subjects have to reverse their choice of stimulus. In an effort to control for the

action selection component, the discriminative stimuli are typically presented randomly at different spatial locations, thus requiring different physical actions to be performed to choose the same stimulus depending on its location (Hampton et al. 2006). Reversal learning is therefore often considered to be an assay of stimulus–outcome learning because choices are made between particular stimuli rather than being between particular physical actions. In practice, however, this form of visual discrimination reversal learning is likely to involve both stimulus–outcome and stimulus–action and/or stimulus–action–outcome associations, as the latter 2 could be computed separately for different combinations of stimuli, spatial locations, and actions. Thus, although previous studies have reported expected value signals in vmPFC during reversal learning, it is not clear whether such signals are driven by stimulus–outcome or stimulus–action–outcome relationships.

To address this question in the present study, we devised a different version of reversal learning: action-based reversal, in
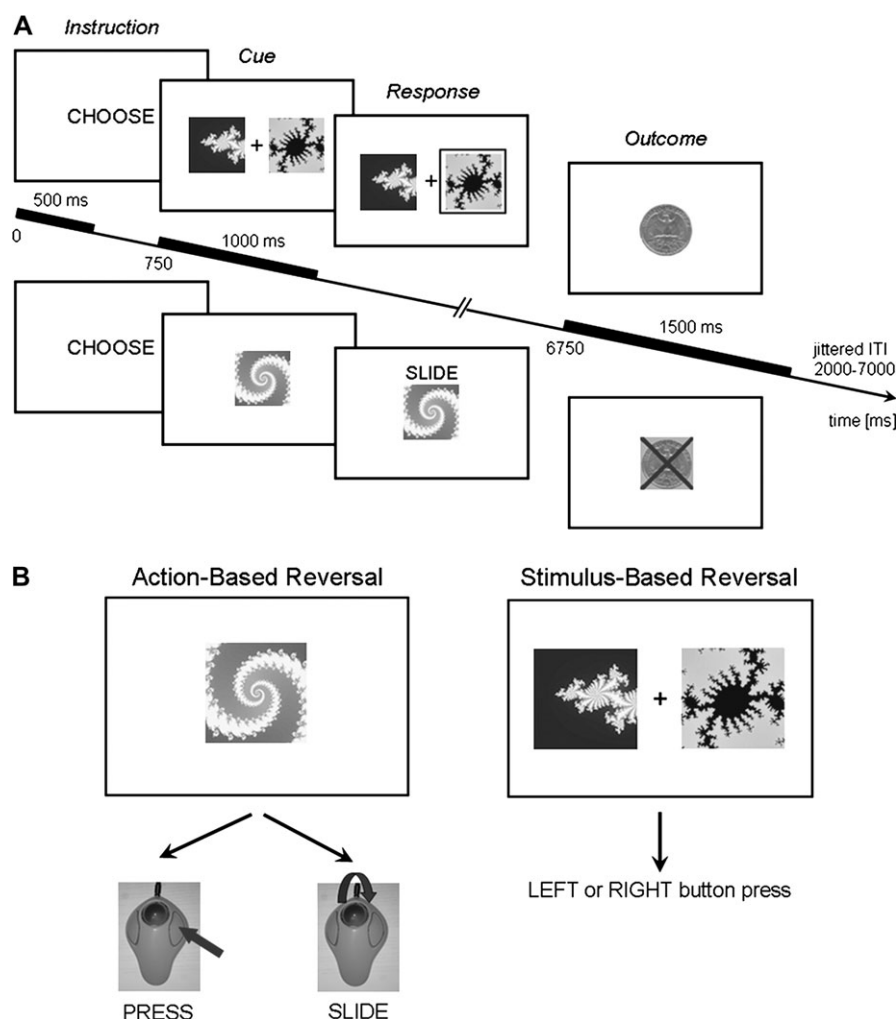
Figure 1. Experimental Task. (*A*) Sequence and timing of events during stimulus-based reversal (top) and action-based reversal (bottom). An instruction signaled the trial type (choice or follow trial). The cue (presented for 1 s) signaled the beginning of the decision period of 1.5 s. Once the subject made their choice response, it was shown to them. The outcome (win or loss) was presented 6 s after the cue for 1.5 s followed by a jittered intertrial interval. (*B*) Contrasting the 2 experimental conditions. During action-based reversal learning, subjects were presented with a single nondiscriminating stimulus and they had to choose between 2 different actions (a slide or a press on a trackball device). During stimulus-based reversal learning, the subjects had to choose between 2 different discriminating stimuli that were presented randomly on both sides of the screen. The subject responded by pressing a button corresponding to the side of the screen that the chosen stimulus was presented.

which subjects make choices between 2 different physical motor responses (a sliding movement on a trackball vs a button press; Fig. 1) instead of requiring subjects to make choices between 2 distinct visual stimuli. No discriminative stimuli are presented to the subjects signaling such options; instead, a single nondiscriminative stimulus is presented to indicate the onset of each trial. Thus, in order to solve such a task, subjects have to rely on representations of the value of each available physical action, not on the representation of the value of different stimuli. By comparing and contrasting neural activity related to expected value in our action-based reversal task to activity found during stimulus-based reversal, we aimed to determine whether ventromedial prefrontal cortex is involved in maintaining representations of the expected value of particular actions, the expected value of specific stimuli, or both. Subjects were scanned with fMRI while they performed both action-based and stimulus-based reversal tasks in separate sessions that were counterbalanced in their order of presentation across subjects.

Developing an understanding of the nature of the associative representations, guiding choice is an essential step in building a complete picture of the underlying computations involved in solving decision making problems under uncertainty. We hypothesized given the connectivity of vmPFC with other regions of cortex involved in action selection such as anterior cingulate cortex, that activity in vmPFC would scale with expected values during our action-based reversal task as well as during our stimulus-based reversal task, demonstrating the existence in this region of goal-directed action-based value signals. In addition, we hypothesized that a number of other brain regions might encode value signals during action-based reversal but not stimulus-based reversal. These include anterior cingulate cortex that has previously been implicated in action-based decision making (Rushworth et al. 2004) and lateral parietal cortex, parts of which have been found to encode values for specific movements (e.g., eye movements) in monkey neurophysiology studies (Platt and Glimcher 1999; Sugrue et al. 2004).

Materials and Methods

Subjects

Twenty subjects (nine females) with a mean age of 22.3 (4.7 standard deviation) years participated in the study. All were recruited from the Caltech student population, and all subjects were free of neurological or psychiatric diseases and had normal or correct-to-normal vision. Informed consent was obtained from each subject, and the study was approved by the Caltech Institutional Review Board.

Experimental Design and Task

Each subject performed 2 distinct variations of a probabilistic reversal learning task: stimulus-based and action-based reversal. Each task was performed in a separate fMRI session lasting 24.5 min, with approximately 2-3 min break between sessions. Subjects remained in the scanner during both sessions. The order of presentation of these tasks was counterbalanced across subjects. In the scanner, visual input was provided with Restech goggles (Resonance Technologies, Northridge, CA). The assignment of the fractal images used as cues in the stimulus and action reversal conditions was randomized across subjects.

Stimulus-Based Reversal

Each trial commenced with an instruction cue ("CHOOSE" or "FOLLOW") for 500 ms that indicated the trial type (choice or control) to the subject (see timing bar in Fig. 1). On choice trials (which were 75% of the total trial number), after a delay of 250 ms, 2 visual stimuli

were presented to the subject. The positions of the stimuli were randomly assigned to the left and right of a fixation cross. Subjects had to choose one of the stimuli by selecting 1 of 2 buttons on a button press response pad, whereby the left button selected the leftmost stimulus and the right button selected the rightmost stimulus. Given that the stimuli were randomly assigned to the left or right of the screen, choice of a particular stimulus on a given trial could require either motor action depending on where that stimulus had been presented on that trial. One stimulus was designated the correct stimulus in that choice of that stimulus led to a monetary reward (winning 25 cents) on 70% of occasions and a monetary loss (losing 25 cents) 30% of the time. Consequently, choice of this correct stimulus led to accumulating monetary gain. The other stimulus was incorrect in that choice of that stimulus led to a reward 30% of the time and a punishment 70% of the time, thus leading to a cumulative monetary loss. After having chosen the correct stimulus on 4 consecutive occasions, the contingencies reversed with a probability of 0.25 on each successive trial. Once reversal occurred, subjects then needed to choose the new correct stimulus, on 4 consecutive occasions, before reversal could occur again (with 0.25 probability). Subjects were informed that reversals occurred at random intervals throughout the experiment but were not informed of the precise details of how reversals were triggered by the computer (so as to avoid subjects using explicit strategies, such as counting the number of trials to reversal). The subject's task was to accumulate as much money as possible and thus keep track of which stimulus was currently correct and choose it until reversal occurs. Subjects were given a response window of 1500 ms following the cue onset. Once they had made a choice, the selected options would be indicated to them on the screen, by surrounding the chosen option with a black square. If they failed to respond within that time window, a feedback message ("Key press timed out") was presented to them and the trial terminated. The outcome (duration 1500 ms)—a picture of a 25c coin (win) or a crossed out 25c coin (loss)—was presented 6000 ms after the cue onset. A variable intertrial interval (ITI) of 2000–7000 ms concluded the trial.

In order to control for activations due to the visual features of the task, and for activity related to simple motor execution, we also added a control condition on 25% of the trials. During this condition, subjects did not have a choice between the different stimuli; rather they had to follow the choices that were presented to them (i.e., a predetermined choice was shown to them and they simply executed it). In order to avoid confusion with the choice task, subjects were shown a different pair of fractals (visual stimuli). The control trials began with presentation of the word "FOLLOW," which was then followed 250 ms later by presentation of the fractal stimuli. One of these stimuli was highlighted by presenting a black square around it. Subjects then had to choose the highlighted stimulus. If the subject then executed the correct response, the black square around the fractal turned green. If the wrong response was executed, the chosen stimulus was highlighted in a red square. The follow trial concluded with a variable ITI of 2000–7000ms. No reward or punishing outcomes were presented during the control trials. Similarly the trial order (sequence of choice and control trials) and the initial contingency (i.e., which stimulus was the first winning stimulus) were also randomized across subjects. Furthermore, the ITIs—evenly distributed and ranging from 2000–7000 ms—were also randomized across trials and for each subject individually.

Action-Based Reversal

The action-based reversal was in most respects identical to the stimulus-based reversal task except in this case instead of making a choice between 2 different stimuli, subjects had to make a decision between 2 different actions: a slide on a trackball with the right index finger or a press on the right button of a trackball device with the right middle finger. The reason for choosing these particular actions was to alleviate the use of an obvious spatial strategy for the subjects: if we had chosen a left and right button press, then the reward associations could have been formed between the concepts of "left" and "right" rather than between the actions themselves. As with the stimulus-based reversal, one of the actions was arbitrarily designated as the correct choice such that choice of that action led to a high probability of winning money, whereas the other "incorrect" action was associated

with a high probability of losing money, according to the same contingencies applied in the stimulus-based task. Contingencies also reversed according to the same rules described above from the stimulus-based task. As with the stimulus task, 75% of the trials were "choice" trials, and the other 25% were control or "follow" trials. To discriminate the choice and follow trials, subjects first saw the word "CHOICE" or "FOLLOW" for 250 ms. Then, a single fractal stimulus was presented in the center of the screen uniquely signaling whether the trial was a choice trial or a follow trial. However, because on choice trials subjects were always presented with the same stimulus, subjects could not use this stimulus to discriminate between the available actions, in contrast to the stimulus-based reversal task. Thus, subjects had to form an association between the 2 different actions and the outcomes (A–O).

Once subjects had made a choice of action, the selected action was indicated on the screen by presenting the word "SLIDE" or "PRESS" above the fractal image. The control trials ("FOLLOW" trials) were similar to the choice trials: following the instruction cue ("FOLLOW"), either the word "SLIDE" or "PRESS" appeared above the fractal image (in black color), and subjects had to select the appropriate action.

Data Acquisition
Functional imaging was performed on a 3 T Siemens (Erlangen, Germany) Trio scanner. Forty-two contiguous interleaved transversal slices of echo-planar T2*-weighted images were acquired in each volume, with a slice thickness of 3 mm and no gap (repetition time, 2500 ms; echo time, 30 ms; flip angle 80°; field of view, 192mm²; matrix, 64 × 64). Slice orientation was tilted –30° from the line connecting the anterior and posterior commissure to alleviate the signal drop in the orbitofrontal cortex (Deichmann et al. 2003). We discarded the first 3 images before data processing and statistical analysis to compensate for the T1 saturation effects.

Data Processing
Image processing and statistical analyses were performed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm). All volumes from all sessions were corrected for differences in slice acquisition, realigned to the first volume, spatially normalized to a standard echo-planar imaging template included in the SPM software package (Friston et al. 1995) using fourth degree B-spline interpolation, and finally smoothed with an isotropic 9-mm full-width-at-half-maximum Gaussian filter to account for anatomical differences between subjects and to allow for valid statistical inference at the group level.

Statistical Analysis

Computational Learning Model
The computational model used in this study updates the values of both cues using the following rule (Hampton et al. 2007):

$$V_{t+1}^c = V_t^c + \eta \times (R_t - V_t^c)$$
$$V_{t+1}^{nc} = V_t^{nc} + \eta \times (-R_t - V_t^{nc})$$

Thus, the new value at trial $t + 1$ (V_{t+1}^c) for the currently chosen cue is based on the sum of the observed prediction (V_t^c) and the prediction error ($R_t - V_t^c$), whereas the new value at trial $t + 1$ for the unchosen option (V_t^{nc}) is based on a fictitious prediction error ($-R_t - V_t^{nc}$) that takes the counterfactual outcome of the current trial ($-R_t$) into account. In this model, η is the learning rate which controls the speed of learning, i.e. the influence of both prediction errors on the value update. The inclusion of an update rule for the unchosen option captures an important feature of the task structure, namely that the choice values are anti-correlated. This update rule approximates the optimal Bayesian solution to the reversal learning problem as described in a previous paper (Hampton et al. 2006), by incorporating the knowledge the subjects have that when the action they are choosing increases or decreases in value, the value of the action they are not choosing does the opposite. The similarities between the fictitious error model used here and the full Bayesian model are discussed in more detail in Hampton et al. (2007).

These values are then translated into choice probabilities $p(A)$ and $p(B)$ using softmax action selection on the value differences of the two options A and B:

$$p(A) = \frac{1}{1 + e^{\beta \times (\alpha - (V_A - V_B))}}, p(B) = 1 - p(A)$$

In this selection rule, α is the indecision point, i.e. both option are equally likely to be selected. Mathematically, this corresponds to the mid-point on the sigmoid function. β is the inverse temperature controlling the stochasticity of the choices, i.e. the slope of the sigmoid.

Behavioral Data Analysis
One of the critical elements for successful task performance during reversal learning is the ability to detect contingency changes and to adapt choice behavior accordingly. Thus, we converted the choice probabilities of the model above to probabilities for switching and staying. The probability to switch on a given trial is equal to the probability of choosing the other option on the next trial. The model above has 3 free parameters (α, η, and β) and was fitted to the behavioral data using a variant of the simulated annealing procedure (Kirkpatrick et al. 1983: see Hampton et al. 2006) with an average maximum log likelihood fitting criterion for both probabilities of switching and staying:

$$L = mean\left(\frac{\sum \log(P_{switch})}{N_{switch}} + \frac{\sum \log(P_{stay})}{N_{stay}}\right)$$

It should be noted that use of a fitting criterion based on the switch and stay probabilities is equivalent to one based on the actual choice probabilities because the former are directly derived from the latter. As an independent test of behavioral model fit we regressed actual switch and stay trials of each subject onto the model-predictions of switch and stay using logistic regression. Model fits of this logistic model were computed as R^2 and statistically assessed by a global F-Test.

Imaging Data Analysis
Using the model-parameters from the behavioral optimization procedure, we constructed a subject-specific (1st level) model for the imaging data that incorporated the following predictors: (a) an event encoding the value of the chosen option at the time of cue presentation, (b) an event encoding the prediction error of the chosen option at the time of outcome presentation, and (c) an event encoding whether the subject switched or stayed on the following trial at 2 s after the outcome. All these predictors were entered as parametric modulations to regressors modeling a blood oxygen level–dependent (BOLD) response at the onset of the respective event. Trials with missing responses were modeled as separate nuisance regressors. In addition, 6 regressors modeling the head motion as derived from the affine part of the realignment procedure were included in the model. Serial auto-correlation was modeled as a 1st order autoregressive model and the data were high-pass filtered at a cutoff of 120 sec. For the second level analysis, we constructed 3 paired t-tests and included the images of the parameter estimates (betas) from the 3 parametric modulations (chosen value, prediction error, and switch/stay). These mixed effects models allow testing for areas that commonly covary with the model-based predictors (conjunction) as well as for areas that exhibit a differential effect for action and stimulus reversal. We present all our statistical maps (SPMs) at a threshold of $p < 0.001$ for displaying purposes. Small volume correction was used to assess the reliability of the observed effects; for consistency we always chose a spherical search volume of 8mm radius. The coordinates for centering the spheres were derived from anatomical masks (Tzourio-Mazoyer et al. 2002; AAL template for amygdata only) and from previous imaging studies that have investigated reversal learning and value-related signals derived from computational learning theory (Daw et al. 2006; Hampton et al. 2007; Hampton et al. 2006; Kim et al. 2006; O'Doherty et al. 2003; Seymour et al. 2004; Seymour et al. 2005) (see Table 1 for details).

Results

Behavioral Results
In order to compute trial-by-trial estimates of expected value for each available decision option, we used an extension of a standard reinforcement learning (RL) model, which on each

Table 1

Coordinates and significance levels for contrasts

Contrast

Region	BA	Left Hemisphere						Right Hemisphere					
		x	y	z	Z	p	ref.	x	y	z	Z	p	ref.
Conjunction of Vchosen for AR and SR													
mOFC	11	−6	24	−21	4.06								
vmPFC	11							6	30	−9	3.89	**	1,4
vmPFC	11							9	27	−12	3.30	**	1,4
vmPFC	10							3	54	−3	3.67	**	1,2
Amyg-hippo Junction	20	−27	−18	−18	4.21	**	1,2	24	−9	−18	4.66	**	1,2
Post. cing. cortex								9	−30	51	4.52		
SMA	6	−3	−9	51	4.24			3	−6	51	4.19		
Mid. occip. gyrus	39	−36	−60	24	5.43								
Cerebellum	19	−21	−63	−18	4.05			21	−66	−18	4.47		
Sens.-mot. cortex	4	−51	−18	54	4.58			51	−15	54	4.89		
Mid. temp. sulcus	21	−57	−3	−15	4.35			63	−15	3	4.23		
AR Vchosen > SR Vchosen													
SMG/inf parietal	2	−57	−30	39	4.16			63	−30	48	3.60		
Mid-cing. cortex	24	−6	6	45	3.21			12	6	36	3.59		
SMA	6	−18	−6	57	4.49			0	3	57	3.69		
Premotor cortex	6	−33	−6	42	4.13			48	−3	39	3.64		
Postcentral sulcus	3	−42	−21	42	3.52								
Post. insula	48	−39	−15	15	3.23								
Front. operculum	48							57	3	12	3.66		
SR Vchosen > AR Vchosen													
No significant clusters at $p < 0.001$ uncorrected													
Conjunction of Prediction Error for AR and SR													
Ventral Striatum		−15	3	−15	5.53	***	5	9	9	−12	5.77	***	5
Dorsal Striatum		−24	6	6	4.45	***	7	24	3	15	4.62	***	7
								30	−6	9	4.58		
vmPFC	11							0	33	−12	5.16	***	5
Rostral ACC	24							3	36	21	3.99	**	6
Lat. OFC	47	−42	33	−9	4.33	***	6						
Amygdala	34	−18	3	−15	5.07	***	8	18	6	−15	5.01	***	8
AR Prediction Error > SR Prediction Error													
SMG								57	−33	33	3.33		
SR Prediction Error > AR Prediction Error													
No significant clusters at $p < 0.001$ uncorrected													
Conjunction of AR Switch>Stay and SR Switch>Stay													
Ant. insula	47							27	27	−3	4.11	**	3
Lat. OFC	11							24	48	−6	3.63		
dlPFC	46	−33	48	21	3.38			36	45	30	4.22		
Conjunction of AR Stay>Switch and SR Stay>Switch													
vmPFC	11	−3	39	−12	4.75	***	3						
Post. cing. cortex	23	3	−15	42	3.75	**	3						
(AR Switch>Stay) − (SR Switch>Stay)													
Intraparietal sulcus	3	−33	−30	45	4.37	**	4						
(SR Switch>Stay) − (AR Switch>Stay)													
No significant clusters at $p < 0.001$ uncorrected													

Abbreviations: BA = Brodmann Area, AR = action-based reversal learning, SR = stimulus-based reversal learning, mOFC/lat. OFC = medial/lateral orbitofrontal cortex, vmPFC = ventromedial prefrontal cortex, SMG = supramarginal gyrus, SMA = supplementary motor area, dlPFC = dorsolateral prefrontal cortex, post. cing. = posterior cingulate

Significance levels: * = $p < 0.05$ corrected, ** = $p < 0.01$ corrected, *** $p < 0.001$ corrected (all corrections are based on a reduced spherical search volume of 8mm radius), all other (non-marked) regions without apriori hypotheses $p < 0.001$ uncorrected

References for center coordinates of spherical search volumes: [1] Hampton et al., 2006, [2] Hampton et al., 2007, [3] O'Doherty et al., 2003, [4] Daw et al., 2006, [5] Kim et al., 2006, [6] Seymour et al., 2005, [7] McClure et al., 2003, [8] AAL template Tzourio-Mazoyer et al., 2002. In case of multiple references, the average coordinate was chosen as the center of the search volume.

trial updates the value of the option that is chosen and the option that is not chosen. This variant of the standard RL model (Sutton and Barto 1998) thereby accounts for the structure in the reversal learning task manifested by the anticorrelation in the reward probabilities available on each decision option (for more details see "Materials and Methods"). To determine how well this model accounted for subjects' actual choice behavior on both the action-based and stimulus-based reversal task, we fitted this model to the choice behavior of our subjects for each task separately. Across all subjects, our optimization procedure resulted in a learning rate of $\eta = 0.51$, an indecision parameter $\alpha = 0.02$, and an inverse temperature of $\beta = 1.73$ for the action-based reversal and $\eta = 0.45$, $\alpha = 0.04$, $\beta = 2.17$ for the stimulus-based reversal. These were later used to derive model predictions in order to construct regressors for the imaging analysis. No significant differences were found in the individual parameter fits between tasks indicating that subjects learned both tasks with comparable rates and had similar decision criteria during action selection for both tasks.

In order to confirm a significant model fit, we performed logistic regression in each subject by using the model-derived action probabilities as an explanatory variable for the subject's actual choices. The overall *F*-test in a logistic regression confirmed a significant fit of the model to subjects' choices in each participant in both the action-based and stimulus-based tasks at a minimum of $P < .0001$ in each subject and condition,
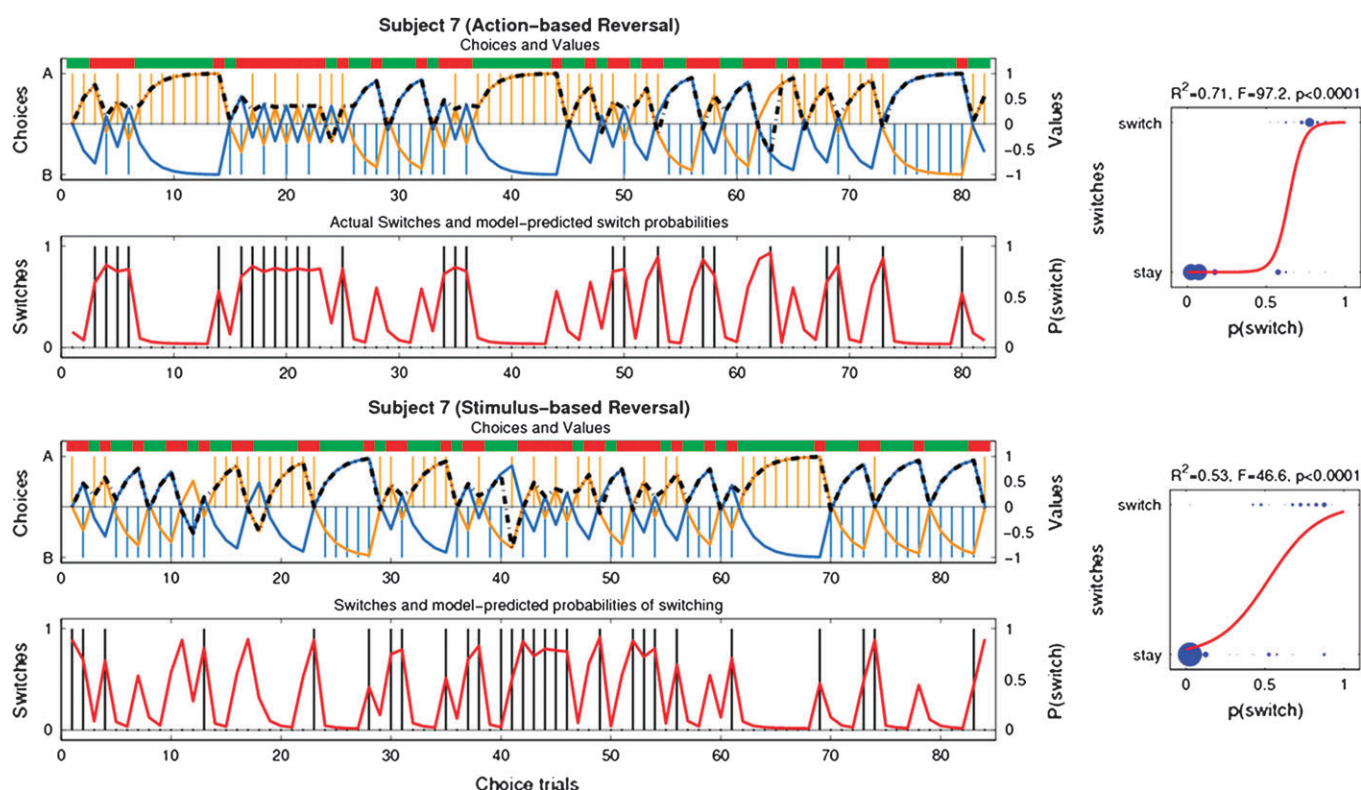
Figure 2. Behavioral data (single subject analysis). Trial-by-trial behavioral model fit of our fictitious update model for an example subject. The top row shows actual choices of the different decision options (sticks) and the model-predicted expected value for each option (lines). The colored bar on top show win trials in green and loss trials in red. We converted these choice probabilities into probabilities of switching and staying and plotted in the second row the actual switches (black stick) and the model-predicted switch probabilities (red lines). An independent logistic regression model was fitted to these switch data (small graphs to the right) to confirm that the switch probabilities of our fictitious update model significantly explained the actual switch data. The logistic fit is plotted in red, the data in blue circles with bigger diameters for overlaying data points.
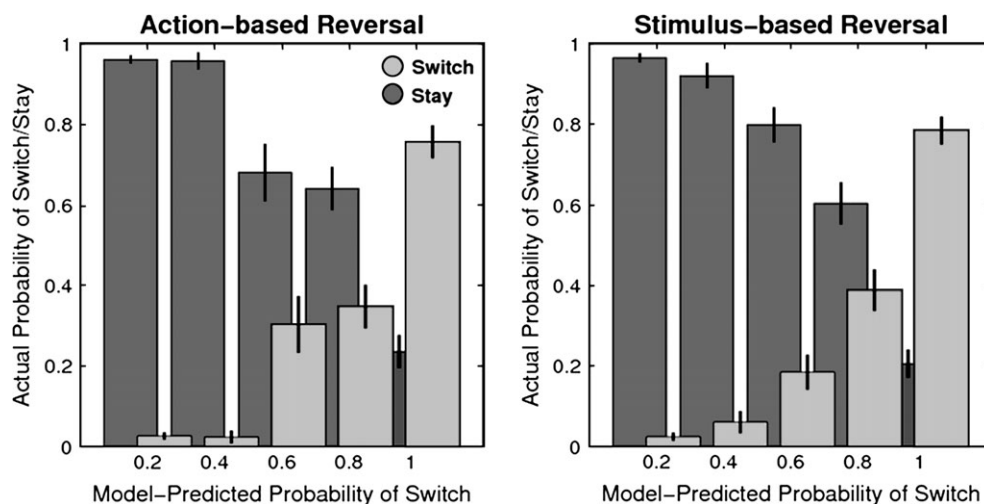
Figure 3. Behavioral data (group analysis). Switch probabilities as predicted by our computational model were grouped into 5 bins (x-axis) and were plotted against the actual switch (light gray) and stay probabilities (dark gray, y-axis) in both experimental conditions (action-based and stimulus-based reversal).

suggesting a good overall fit of the model to subjects' actual choices in both cases. Figure 2 shows the behavioral model fit for an example subject.

To provide a further assay of our model fit, we grouped the model-predicted switch probabilities into 5 bins and plotted the actual mean switch and stay probabilities across subjects for each of these bins (Fig. 3). Switch probabilities are shown in light gray

and the stay probabilities in dark gray. Switch and stay probabilities are complementary in this study because the value functions for both options are anticorrelated. Thus, model-predicted and actual switch probabilities show a positive correlation, whereas model-predicted switch and actual stay probability show a negative correlation. An independent linear regression for the actual vs model-predicted switch probabilities

confirmed that the slope across the different bins was significantly larger than zero in both experimental conditions (stimulus-based reversal: slope = 0.17 (±0.12 95% confidence interval [CI]), R^2 = 0.88, F = 22.06, $P < 0.05$; action-based reversal: slope = 0.18 (±0.13 95% CI), R^2 = 0.88, F = 21.80, $P < 0.05$), suggesting that the model-derived probabilities of switching predicted the actual probability of switching across the entire sample.

fMRI Results

Using the model parameters described above, we took the trial-by-trial predictions of our computational model for expected value and prediction error and entered these into a regression analysis against the fMRI data separately for the action-based and stimulus-based tasks. We were interested in identifying brain areas responding commonly to expected value in the action-based and stimulus-based reversal conditions as well as areas that show differential correlations with value in the action-based compared with stimulus-based task and similarly for prediction errors.

Common Responses to the Value of the Chosen Option in Action- and Stimulus-Based Reversal

To test for regions showing common responses to value in both reversal conditions, we performed a conjunction analysis. This analysis revealed significant responses to value in ventromedial prefrontal cortex (vmPFC), the medial orbitofrontal cortex (mOFC), and the amygdalo–hippocampal junction (see Fig. 4, left panels for statistical maps and plots of BOLD time courses

and Table 1 for coordinates, z values and significance levels in all statistical contrasts). The time course plots reveal elevated BOLD responses in vmPFC and posterior amygdalo–hippocampal junction when the value of the chosen option was high and decreased responses when this value was low. This effect is present in both action- and stimulus-based reversal conditions.

Other areas showing significant correlations with the value of the chosen option in both tasks include the posterior cingulate cortex, supplementary motor area, middle occipital gyrus, sensorimotor cortex, and an area extending over the middle and superior temporal sulcus (see Table 1).

We also tested whether these effects for the value of the chosen option during action-based reversal would change if we aligned this value signal to the time of response, rather than to the time of cue. The results from the response-aligned model did not change substantively from those in the cue-aligned model, as all the same areas were found to survive our statistical threshold. This finding likely reflects the fact that the interval between the cue and response may be too short in the present design (~670 ms) to allow discrimination between signals responding to the cue or the response itself, due to limits in the temporal resolution of the BOLD signal.

Differential Effects for Value of the Chosen Option in Action Compared with Stimulus Reversal

Differential effects were established by setting up a linear contrast that weights the parametric regressors encoding the
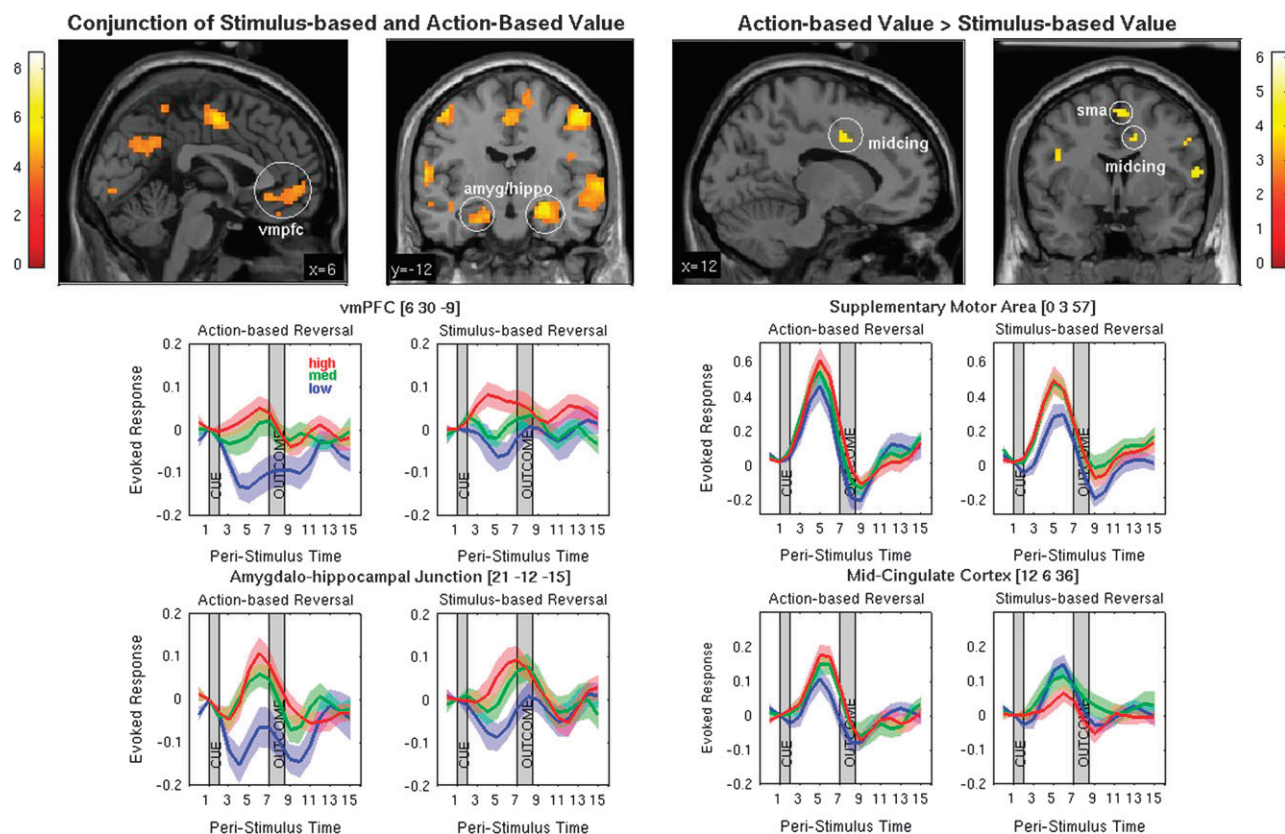
Figure 4. Value-related activations in both experimental conditions. Left panels: results of a conjunction analysis testing for common effect for value-related signals in action-based and stimulus-based reversal. Significant effects were found in the ventromedial prefrontal cortex (vmpfc) and in the amygdalo–hippocampal junction (amyg/hippo) and corresponding blood oxygen level–dependent (BOLD) time course are plotted in the panels below. Right panels: results of a contrast testing for greater value-related effect during action-based than stimulus-based reversal. Significant effects were found in the supplementary motor area (SMA) and in the midcingulate cortex (midcing) with corresponding BOLD time courses plotted in the panels below.

value of the chosen option (Vchosen) in both task conditions. An area in the supplementary motor area (SMA) and in the midcingulate cortex showed significantly greater effects for the value of the chosen option (Vchosen) during action reversal than during stimulus reversal (see Fig. 4, right panels). The plots of the BOLD time course show elevated evoked responses to trials with higher Vchosen only during action-based reversal but not during stimulus-based reversal. In the midcingulate cortex, trials with lower Vchosen actually elicit a higher response. In addition, we also observed similar effects bilaterally in the supramarginal gyrus reaching into the intraparietal sulcus, premotor cortex, left postcentral gyrus, and right frontal operculum (see Table 1).

No voxels survived our statistical threshold for the reverse contrast testing for areas showing significantly stronger correlations with value for the stimulus-based compared with the action-based reversal task.

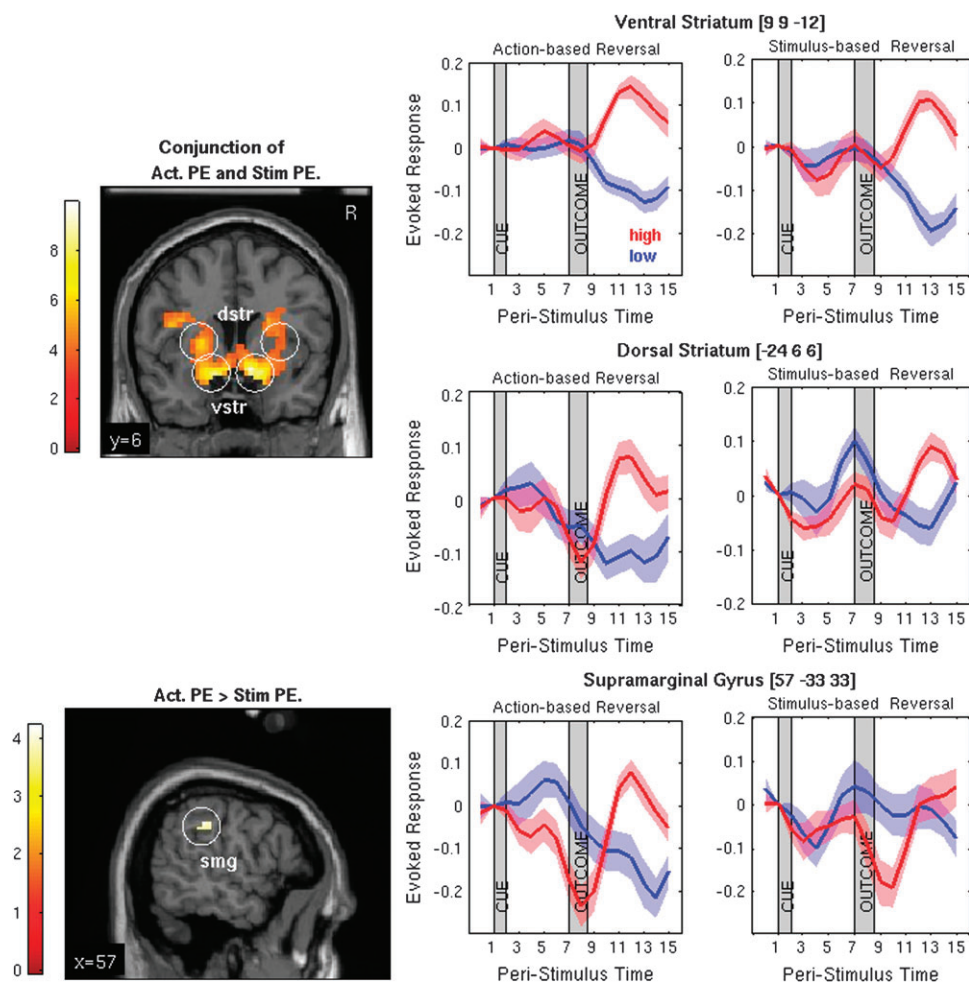### Common Responses to the Prediction Error in Action- And Stimulus-Based Reversal

Significant effects for prediction errors in both reversal conditions were observed bilaterally in the ventral and dorsal striatum (see Fig. 5, top and middle panels, and Table 1). Plots of the time courses in these areas show that the BOLD responses following the presentation of the outcome are stronger on trial with a high compared with a low prediction error. Other areas responding to prediction errors in both conditions include the vmPFC, the rostral anterior cingulate cortex, left lateral OFC, and bilateral amygdala.

### Differential Effects for Prediction Error

An area in the right supramarginal gyrus exhibited a significantly larger effect for prediction error during action-based reversal (see Fig. 5, bottom panels, and Table 1). The plots of the evoked BOLD responses show that only during action-based reversal did trials with a large prediction error elicit stronger activations than those with a low prediction error; during stimulus-based reversal, this pattern was reverse following the presentation of the outcome.

No voxels survived our significance threshold for the reverse contrast, testing for areas showing greater prediction error responses during the stimulus-based compared with the action-based task.

Figure 5. Activations correlating with prediction error in both experimental conditions. Top and middle panels: results of a conjunction analysis testing for common activation to the prediction error derived from our computational model in action-based and stimulus-based reversal. The blood oxygen level–dependent (BOLD) time course in the right panels reveal that trials with a high prediction error elicit strong evoked responses in ventral (vstr) and dorsal striatum (dstr) following the outcome. Bottom panels: results of a differential contrast comparing prediction error signals in action-based and stimulus-based reversal, revealing a region of supramarginal gyrus (smg) exhibiting stronger correlations with prediction error in the action compared to the stimulus-based task.
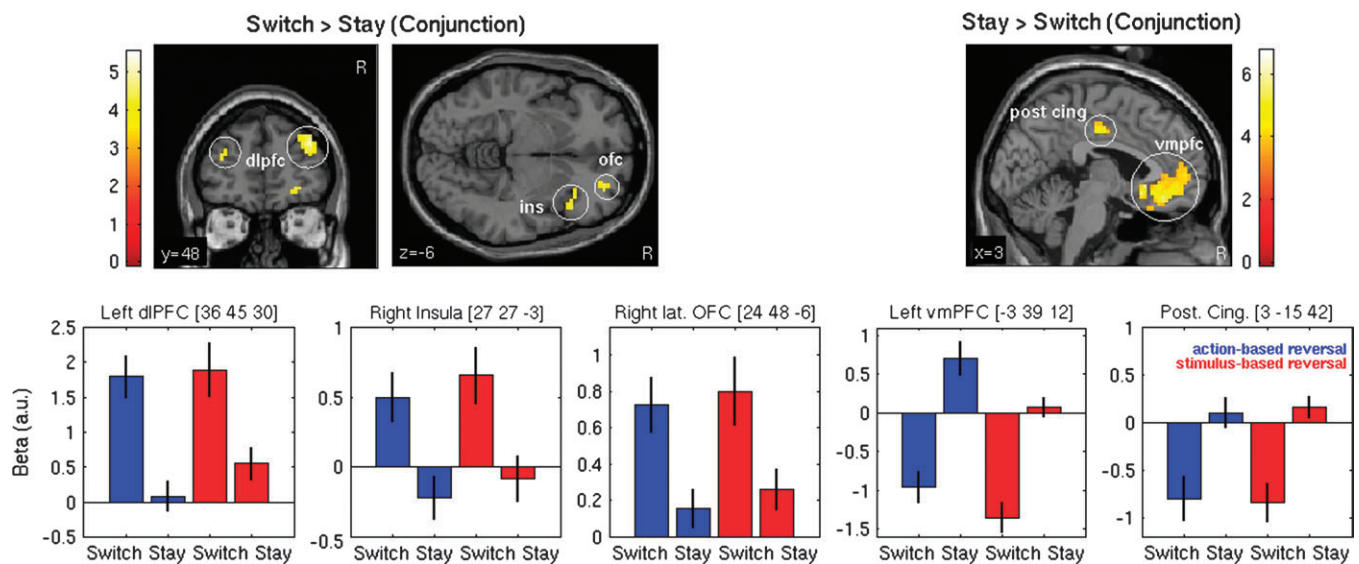
Figure 6. Activations related to switching or staying in both experimental conditions. Left panels: results from a conjunction analysis testing for switch-related activations common to both action-based and stimulus-based reversal. Common effects were found in the dorsolateral prefrontal cortex (dlpfc), the lateral orbitofrontal cortex (ofc), and in the anterior insula (ins). Right panels: results from a conjunction analysis testing for stay vs switch responses in both action-based and stimulus-based reversal. Effects were found in the ventromedial prefrontal cortex (vmpfc) and in the posterior cingulate cortex (post cing) (error bar = SEM across subjects).

Areas Involved in Signaling Changes in Behavior: Common Responses in Action and Stimulus-Based Reversal

In addition to testing for regions correlating with signals derived from our computational model, we also tested for areas responding on trials immediately preceding a change in behavior on the task, that is, when subjects switch their choice of decision option, compared with when subjects maintain their current choice of decision option (stay). Using a conjunction analysis, we found significant effects for switch > stay in both reversal conditions in the right anterior insula, lateral OFC, and bilateral dorsolateral prefrontal cortex (dlPFC) as shown in Figure 6 (left and middle panels). Plot of the parameter estimates reveal that the size of the effect is comparable in both experimental conditions, but strongest in the dlPFC. A conjunction analysis for the reverse contrast (stay > switch) shows activations related to maintaining the current choice of decision option in the vmPFC and in the posterior cingulate cortex close to the midline (see Fig. 6, right panels and Table 1).

Differential Effects in Areas Signaling Behavioral Change

We found significantly greater responses for switch > stay during action-based than during stimulus-based reversal in the intraparietal sulcus (IPS) reaching anteriorly into the postcentral sulcus (see Fig. 7 and Table 1). No voxels survived our statistical threshold for the reverse contrast testing for areas showing significantly stronger correlations with switch > stay for the stimulus-based compared with the action-based reversal task.

Value of the Chosen Action Versus Pavlovian Stimulus Values during Action-Based Reversal

We also tested an alternative explanation for our results in the action-based reversal task which is that signals observed in vmPFC during this task could reflect the value of the discriminative stimulus presented at trial onset, irrespective of what action is ultimately selected. In essence, this would reflect a pure Pavlovian association, as it would incorporate no information about specific actions at all. To address this alternative possibility, we ran an additional analysis in which we estimated what in reinforcement learning terms is a pure "state value" signal at the time of choice—that is, the expected future reward that follows from observing the single discriminative stimulus in the action reversal condition. This signal is learned by simply on each trial updating the value of the stimulus in proportion to a prediction error generated by the difference between that stimulus value and the outcome obtained (irrespective of what action is taken). We took this signal and entered this into the fMRI analysis alongside the value of chosen action signal that we argue is being computed in vmPFC for the action reversal condition (thereby reflecting action–outcome and not stimulus–outcome associations). When comparing these two signals in the same analysis, we found no evidence that the activation in vmPFC reflects a Pavlovian state value. In the action-based reversal task, the regressor for the state value did not show any significant effects, even at lenient thresholds of $P < .01$. In contrast, the value of the chosen option showed a robust and significant effect at $P < .001$, in the same regions of vmPFC identified above. Therefore, this analysis confirmed our original findings.

Comparison of Follow Versus Choice Trials

We also compared the activation elicited by the force choice condition (FOLLOW trial) in comparison to the free choice condition. This analysis revealed an activation in the left premotor cortex, contralateral to the hand that the subjects responded with, and in the left occipito–parietal cortex (see Supplementary Fig. 1).

Discussion

A fundamental question in the area of decision neuroscience concerns the nature of the associative representations being employed in the brain while subjects' are making choices
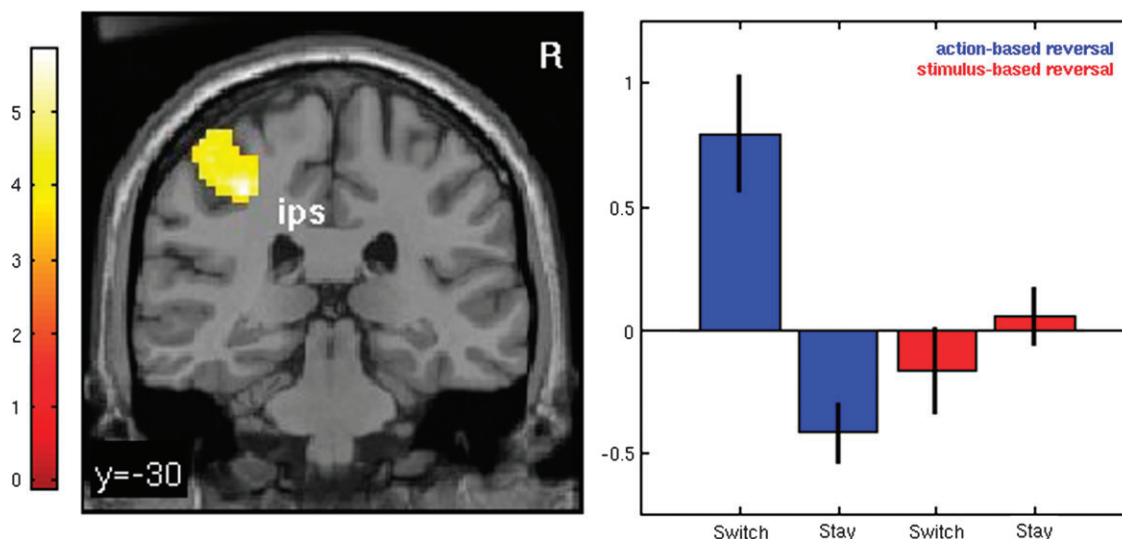
Figure 7. Activation pattern related to stronger switch > stay response in the action-based compared with the stimulus-based task [interaction contrast: (AR switch > stay) − (SR switch > stay)]. Top panel. SPM showing a stronger switch > stay effect in the action-based reversal in the left intraparietal sulcus (ips). Bottom panel. Parameter estimates indicating the average activation to switch and stay trials in both task conditions (error bar = SEM across subjects).

between different decision options. Although a number of studies have reported expected reward signals in diverse brain regions including ventromedial prefrontal cortex during decision making, the specific associations underlying such signals in the human brain have hitherto remained unclear. Here, we present evidence from a human fMRI study to indicate that ventromedial prefrontal cortex is involved in tracking expectations of future reward attributable to particular motor actions. More specifically, activity in this region was found to scale with value signals derived from a variant of a computational reinforcement learning model while subjects performed a reversal learning task involving a choice between 2 motor actions, in the absence of specific discriminative visual stimuli to denote those choices.

We also compared and contrasted activity in the action-based reversal task with that elicited during the stimulus-based reversal task. In the latter condition, decision options are denoted via presentation of specific discriminative stimuli; however, the 2 physical actions denoting the different choice options are randomly assigned (depending on random spatial position of the 2 discriminative stimuli). In common with the action-based reversal task, we also observed expected reward signals in vmPFC while subjects performed the stimulus-based task, consistent with a number of previous reports (Daw et al. 2006; Hampton et al. 2006; Kim et al. 2006; Valentin et al. 2007). Stimulus-based reversal is often assumed to depend on stimulus–outcome associations, and indeed neuronal activity in orbitofrontal cortex recorded during performance of such a task typically reveals stimulus-related neuronal activity, but not response-related signals (Hoshi et al. 2005; Schoenbaum et al. 1998; Seo and Lee 2007; Thorpe et al. 1983; Wallis and Miller 2003). Consequently, activity we observe in medial orbitofrontal cortex and adjacent medial PFC during the stimulus-based task may pertain to encoding of stimulus–outcome associations. According to this interpretation, expected reward signals in vmPFC could be driven by both stimulus–outcome and action–outcome associations depending on the task context. However, accumulating evidence from rodent lesion studies suggests that at least in the rodent brain,

regions of prefrontal cortex involved in mediating stimulus–outcome (or Pavlovian) learning are distinct and dissociable from those regions involved in mediating action-outcome (or goal-directed learning) (Balleine et al. 2007), with rat orbitofrontal cortex implicated in the former (Ostlund and Balleine 2007) and prelimbic cortex implicated in the latter (Ostlund and Balleine 2005). If these findings can be extrapolated to the primate brain, then this would rule out the interpretation that vmPFC (part of which may be homologous to prelimbic cortex in the rodent brain) is involved in both goal-directed and Pavlovian learning.

An alternative possibility compatible with the rodent data is that activity in vmPFC during the stimulus-based reversal task is in common with that in the action–outcome task, also being driven by goal-directed action–outcome associations. Although in the stimulus-based task the particular physical motor response required to implement a specific decision varies on a trial-by-trial basis (depending on where the stimuli are presented), it is possible for associations to be learned between a combination of visual stimuli locations, responses, and outcomes. Thus, the common involvement of vmPFC in both the action- and stimulus-based reversal could be attributable to the possibility that this region is generally involved in encoding values of chosen actions but that those action–outcome relationships are encoded in a more abstract and flexible manner than concretely mapping specific physical motor responses to outcomes. The more flexible encoding of "actions" that this framework would entail, may have parallels with computational theories of goal-directed learning in which action selection is proposed to occur via a flexible forward model system, which explicitly encodes the states of the world, the transition probabilities between those states, and the outcomes obtained in those states. In this context, an "action" is any behavior by the subject causing a particular path to be implemented through those states, for example, an action could be "when in state x choose the action that leads me to state y," irrespective of the specific physical motor act required to implement that action (Daw et al. 2005). The fact that value-related activity in vmPFC is best captured by an extension of reinforcement learning that encodes the

structure of the reversal learning problem (and hence the appropriate transition probabilities between states), as shown in Hampton et al. (2006, 2007), and used in the present study, is also consistent with the notion that vmPFC is involved in model- or state-based inference of this sort.

We also found an area of posterior amygdala extending into anterior hippocampus showing correlations with the value of the chosen option in both action- and stimulus-based reversal. Previous studies in both animals and humans have emphasized an important role for amygdala in encoding expected rewards (Hampton et al. 2006; Paton et al. 2006; Schoenbaum et al. 1998). Moreover, interactions between this region and orbitofrontal cortex has been shown to be necessary for establishing expected reward representations in prefrontal cortex in both rodents and humans (Hampton et al. 2007; Schoenbaum et al. 1998, 2003). The results of the present study indicate that the amygdala may not be involved exclusively in encoding stimulus-reward value but might also be involved in encoding the value of chosen actions.

We found evidence for distinct encoding of value signals for specific physical motor actions (i.e., during the action-based task) compared with the stimulus-based task in more dorsal parts of the brain that unlike vmPFC are directly connected to primary motor cortex, such as the supplementary motor area (SMA) and midcingulate cortex. These areas are known to be involved in reward-based motor selection tasks in monkeys (Shima and Tanji 1998) and more generally in response selection (Picard and Strick 2001). These findings are compatible with the results of a recent monkey recording study in SMA which reported in increase in neuronal spike rates in this area as the animal approached a rewarding target (Sohn and Lee 2007). A specific role for dorsal anterior cingulate cortex in reward-based action selection has been proposed by Rushworth et al. (2007), on the basis of a series of both monkey lesion and human fMRI studies. In one such fMRI study, activity in anterior cingulate was observed under situations where subjects actively chose what action to take in a reward-related response task compared with a situation in which subjects were instructed to take a specific action, in which case anterior cingulate was not involved (Walton et al. 2004). Furthermore, lesions of monkey anterior cingulate cortex were found to impair action selection based on the monkey's history of past reinforcement, but not adjustments in behavior following errors (Kennerley et al. 2006). Similarly, a recent single-unit recording study in monkeys reported neurons in dorsal anterior cingulate cortex that were modulated by the reward history in accordance with value signals derived from a reinforcement learning model (Seo and Lee 2007). Although in the present study, we found evidence for midcingulate involvement in both stimulus- and action-based reversal tasks, a part of this area was significantly more active during the action-based condition compared with the stimulus-based condition. Thus, our findings are broadly consistent with the possibility that when choices need to be made between different physical motor responses, additional circuitry in supplementary motor cortex and dorsal mid-cingulate cortex are recruited.

While midcingulate cortex was correlated with expected reward, we found a more anterior region of pericingulate cortex to be correlated negatively with expected reward during the stimulus-based reversal task (see Supplementary Fig. 2). In other words, the less rewarding a particular chosen option was (according to the model prediction), the greater the activity in this region. Previously we have shown using a multivariate classification approach that activity in a similar region of anterior cingulate cortex is highly predictive of subjects' subsequent behavioral decisions while subjects are performing stimulus-based reversal learning (Hampton and O'Doherty 2007), with increasing activity in this area predictive of subsequent changes in behavior. Complicating matters even further, we found extensive correlations with expected reward during both action- and stimulus-based reversal in posterior cingulate cortex. These findings therefore suggest that different regions of cingulate cortex (anterior, middle, and posterior) may mediate quite distinct functions during reward-based decision making, both in terms of the nature of the signals being encoded (whether they are positively or negatively correlated with expected reward), and the type of decision task in which they are involved (action based, stimulus based, or both).

In addition to testing for regions involved in tracking the value of particular actions, we also tested for regions correlating with subjects' actual choice behavior. In reversal learning, the subject can implement one of 2 types of behavior: either maintaining their choice of the current decision option or switching their choice to the alternate option. We tested for regions showing activation after subjects receive an outcome on a given trial but before subjects' make a choice on the subsequent trial, that is, activity that differs depending on whether subjects maintain their current choice (stay) or switch to the alternate option (switch). We found increased activity in anterior insula (frontoinsular cortex) extending into caudolateral OFC and bilaterally in the dorsolateral prefrontal cortex, when on subsequent trials subjects switched their behavioral choice compared with when they maintained the current choice, replicating a number of previous findings that have implicated these areas in signaling behavioral switches during reversal (Cools et al. 2002; O'Doherty et al. 2003). Switch-related activity was present in these regions during both action-based and stimulus-based reversal, suggesting that this region is involved in signaling changes in behavior irrespective of whether this behavioral change involves switching between specific motor actions or more abstractly, switching between different decision options.

While switch-related activity in the above regions was common to both action and stimulus-based reversal tasks, differential switch-related activity between the 2 tasks was found in left intraparietal sulcus (IPS). This area was selectively involved in signaling a switch in subjects' behavioral choices during action-based but not the stimulus-based reversal. Neurons in this area have previously been implicated in processes related to action-based decision making in non-human primates (Dorris and Glimcher 2004; Platt and Glimcher 1999; Sugrue et al. 2004). A recent human fMRI study reported a change in activity in this region when subjects switched between exploratory and exploitative decision modes, such that activity in this region was higher when subjects decided to explore actions considered to have lower value than the best available option in order to gain more information about the rewards available on those actions (Daw et al. 2006). Here, activity in this region appeared to be related to subjects' switching their choices, but only when those choices involved physical actions, not abstract options. Taken together, these findings support an important role for this brain region in action-based decisions.

It should be emphasized that the value signals we report in vmPFC and elsewhere correspond to the expected reward of the chosen option, whether it is the chosen physical action or the chosen discriminative stimulus. Such signals likely reflect the consequence of the decision process in the sense that the chosen option can only be encoded once the decision of what action to choose has been made. However, in order to make the decision itself, a different type of signal needs to be encoded, namely, the value of each individual option in the choice set, be it the value of specific actions or specific stimuli. On account of the anticorrelation between the action reward probabilities or stimulus reward probabilities in the reversal task used here, we cannot separately measure these prechoice action values and so cannot establish whether vmPFC also plays role in encoding such signals. Nevertheless, the chosen value signals we do report in the action-based task likely depend on retrieval of learned action–outcome associations, suggesting that this region does process action–outcome information, even if it is only corresponding to the value of the action ultimately chosen. It is notable that in addition to chosen values, vmPFC also contains signals related to the behavioral choice itself, with activity increasing in this area on trials where subjects decide to continue their current choice strategy as opposed to switching. The presence of behavioral choice signals in vmPFC alongside the value of chosen actions does appear to be consistent with an important role for this region in decision making, either by contributing directly to the decision or at the very least in reporting the consequences of the decision. Further studies will be needed to disambiguate these possibilities.

In conclusion, the main finding of this study is that we found a role for ventromedial prefrontal cortex in a reward-related decision making task during which subjects are required to make a choice between different physical actions (button press vs tracker ball slide), in addition to the previously reported role for this region in decision making tasks in which decision options are denoted by different discriminative stimuli. In both cases, activity in this region correlated significantly with expected future reward derived from a computational model. The finding that vmPFC is involved in action-based reversal in which no discriminative stimulus is present to signal the different decision options suggests that vmPFC is involved in action–outcome learning by encoding the expected future reward attributable to particular physical actions. The present findings therefore demonstrate that human vmPFC is not only merely involved in encoding the values assigned to particular discriminative stimuli but is also involved in encoding values assigned to particular physical motor responses. A parsimonious explanation for the present results is that vmPFC plays a general role in goal-directed learning, encoding action–outcome relationships irrespective of whether those actions correspond to specific physical motor actions or denote implementation of a decision option on a more abstract level.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

Notes

Address correspondence to Jan Gläscher, Division of Humanities and Social Sciences, California Institute of Technology, MC 114-96, Caltech, 1200 E. California Boulevard, Pasadena, CA 91125. Email: glascher@hss.caltech.edu.

References

Balleine BW, Delgado MR, Hikosaka O. 2007. The role of the dorsal striatum in reward and decision-making. J Neurosci. 27:8161-8165.

Bechara A, Damasio AR, Damasio H, Anderson SW. 1994. Insensitivity to future consequences following damage to human prefrontal cortex. Cognition. 50:7-15.

Blair K, Marsh AA, Morton J, Vythilingam M, Jones M, Mondillo K, Pine DC, Drevets WC, Blair JR. 2006. Choosing the lesser of two evils, the better of two goods: specifying the roles of ventromedial prefrontal cortex and dorsal anterior cingulate in object choice. J Neurosci. 26:11379-11386.

Clark L, Cools R, Robbins TW. 2004. The neuropsychology of ventral prefrontal cortex: decision-making and reversal learning. Brain Cognit. 55:41-53.

Cools R, Clark L, Owen AM, Robbins TW. 2002. Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci. 22:4563-4567.

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 8:1704-1711.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates for exploratory decisions in humans. Nature. 441:876-879.

Deichmann R, Gottfried JA, Hutton C, Turner R. 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage. 19:430-441.

Dorris MC, Glimcher PW. 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron. 44:365-378.

Friston KJ, Ashburner J, Frith CD, Poline JB, Heather JD, Frackowiak RS. 1995. Spatial registration and normalization of images. Hum Brain Mapp. 2:165-189.

Hampton AN, Adolphs R, Tyszka MJ, O'Doherty JP. 2007. Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. Neuron. 55:545-555.

Hampton AN, Bossaerts P, O'Doherty JP. 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci. 26:8360-8367.

Hampton AN, O'Doherty JP. 2007. Decoding the neural substrates of reward-related decision making with functional MRI. Proc Natl Acad Sci U S A. 104:1377-1382.

Hoshi E, Sawamura H, Tanji J. 2005. Neurons in the rostral cingulate motor area mon0itor multiple phases of visuomotor behavior with modest parametric selectivity. J Neurophysiol. 94:640-656.

Kable JW, Glimcher PW. 2007. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 10:1625-1633.

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF. 2006. Optimal decision making and the anterior cingulate cortex. Nat Neurosci. 9:940-947.

Kim H, Shimojo S, O'Doherty JP. 2006. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol. 4:e233.

Kirkpatrick S, Gelatt CD, Jr, Vecchi MP. 1983. Optimization by Simulated Annealing. Science. 220:671-680.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G. 2005. Distributed neural representation of expected value. J Neurosci. 25:4806-4812.

Matsumoto K, Suzuki W, Tanaka K. 2003. Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science. 301:229-232.

Matsumoto M, Matsumoto K, Abe H, Tanaka K. 2007. Medial prefrontal cell activity signaling prediction errors of action values. Nat Neurosci. 10:647-656.

Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci. 16:1936-1947.

Montague PR, King-Casas B, Cohen JD. 2006. Imaging valuation models in human choice. Annu Rev Neurosci. 29:417-448.

O'Doherty J, Critchley H, Deichmann R, Dolan RJ. 2003. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci. 23:7931-7939.

O'Doherty JP. 2004. Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr Opin Neurobiol. 14:769-776.

Ostlund SB, Balleine BW. 2005. Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. J Neurosci. 25:7763-7770.

Ostlund SB, Balleine BW. 2007. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. J Neurosci. 27:4819-4825.

Padoa-Schioppa C, Assad JA. 2006. Neurons in the orbitofrontal cortex encode economic value. Nature. 441:223-226.

Paton JJ, Belova MA, Morrison SE, Salzman CD. 2006. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature. 439:865-870.

Picard N, Strick PL. 2001. Imaging the premotor areas. Curr Opin Neurobiol. 11:663-672.

Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. Nature. 400:233-238.

Rushworth MF, Buckley MJ, Behrens TE, Walton ME, Bannerman DM. 2007. Functional organization of the medial frontal cortex. Curr Opin Neurobiol. 17:220-227.

Rushworth MF, Walton ME, Kennerley SW, Bannerman DM. 2004. Action sets and decisions in the medial frontal cortex. Trends Cogn Sci. 8:410-417.

Schoenbaum G, Chiba AA, Gallagher M. 1998. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat Neurosci. 1:155-159.

Schoenbaum G, Setlow B, Saddoris MP, Gallagher M. 2003. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron. 39:855-867.

Seo H, Lee D. 2007. Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. J Neurosci. 27:8366-8377.

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. 2004. Temporal difference models describe higher-order learning in humans. Nature. 429:664-667.

Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R. 2005. Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat Neurosci. 8:1234-1240.

Shima K, Tanji J. 1998. Role for cingulate motor area cells in voluntary movement selection based on reward. Science. 282:1335-1338.

Sohn JW, Lee D. 2007. Order-dependent modulation of directional signals in the supplementary and presupplementary motor areas. J Neurosci. 27:13655-13666.

Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the representation of value in the parietal cortex. Science. 304:1782-1787.

Sutton RS, Barto AG. 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge (MA): Massachusetts Institute of Technology Press.

Thorpe SJ, Rolls ET, Maddison S. 1983. The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp Brain Res. 49: 93-115.

Tremblay L, Schultz W. 1999. Relative reward preference in primate orbitofrontal cortex. Nature. 398:704-708.

Tremblay L, Schultz W. 2000. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. J Neurophysiol. 83:1864-1876.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273-289.

Valentin VV, Dickinson A, O'Doherty JP. 2007. Determining the neural substrates of goal-directed learning in the human brain. J Neurosci. 27:4019-4026.

Volz KG, Schubotz RI, von Cramon DY. 2006. Decision-making and the frontal lobes. Curr Opin Neurol. 19:401-406.

von Neumann J, Morgenstern O. 1944. Theories of Games and economic behavior. Princeton (NJ): Princeton University Press.

Wallis JD, Miller EK. 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur J Neurosci. 18:2069-2081.

Walton ME, Devlin JT, Rushworth MF. 2004. Interactions between decision making and performance monitoring within prefrontal cortex. Nat Neurosci. 7:1259-1265.