How Does Reward Expectation Influence Cognition in the Human Brain?

James B. Rowe[1,2,3], Doris Eckstein[1,2], Todd Braver[1,4], and Adrian M. Owen[2,3]

Abstract

■ The prospect of reward changes how we think and behave. We investigated how this occurs in the brain using a novel continuous performance task in which fluctuating reward expectations biased cognitive processes between competing spatial and verbal tasks. Critically, effects of reward expectancy could be distinguished from induced changes in task-related networks. Behavioral data confirm specific bias toward a reward-relevant modality. Increased reward expectation improves reaction time and accuracy in the relevant dimension while reducing sensitivity to modulations of stimuli characteristics in the irrelevant dimension. Analysis of functional magnetic resonance imaging data shows that the proximity to reward over successive trials is associated with increased activity of the medial frontal cortex regardless of the modality. However, there are modality-specific changes in brain activity in the lateral frontal, parietal, and temporal cortex. Analysis of effective connectivity suggests that reward expectancy enhances coupling in both early visual pathways and within the prefrontal cortex. These distributed changes in task-related cortical networks arise from subjects' representations of future events and likelihood of reward. ■

INTRODUCTION

Money changes the way we think and behave. More generally, the prospect of rewards including money influences cognitive processes such as attention, decision making, and the configuration of mental "rules" that determine our responses to events. This interaction between cognition and reward expectation has an important influence on everyday thought and behavior, even when the reward is predictable or subliminal (Pessiglione et al., 2007). Conversely, abnormal interactions may contribute to the development of gambling and addictive behaviors (Tanabe et al., 2007; Reuter et al., 2005; Goudriaan, Oosterlaan, de Beurs, & Van den Brink, 2004; Bolla et al., 2003).

The neural basis of reward representation is often presented as functionally and anatomically distinct from the neural basis of configuration of cognitive or behavioral rules (sets). For example, the representation of rewards is closely associated with the orbital and medial frontal cortex, the anterior cingulate, and the ventral striatum (Shidara & Richmond, 2002, 2004; O'Doherty, Critchley, Deichmann, & Dolan, 2003; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001), and lesions of these areas impair goal-directed behaviors, especially in regard to changing reward contingencies (Hornak et al., 2004;

Bechara, Damasio, Tranel, & Anderson, 1998). In contrast, cognitive set or "rule" functions are closely associated with the lateral prefrontal cortex (Aron, Monsell, Sahakian, & Robbins, 2004; Sakai & Passingham, 2003; MacDonald, Cohen, Stenger, & Carter, 2000a; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000), and lesions of the lateral prefrontal cortex impair rule-based behaviors (Aron et al., 2004; Manes et al., 2002; Dias, Robbins, & Roberts, 1997).

These lateral and ventral/medial systems must both be active under conditions in which a change in cognitive processes must arise from the representation of goals or reward expectations. However, the nature of the interaction between them is controversial, with two alternative mechanisms. First, it has been proposed that there is a global workspace within which decision-making processes are modulated according to rewards (Dehaene & Changeux, 2000; Dehaene, Kerszberg, & Changeux, 1998). This hypothesis emphasizes an interaction within widely distributed rather than local brain networks. In contrast, it has been proposed that the interactions occur within local regions in the prefrontal cortex. For example, monkey neurons in the lateral prefrontal cortex can encode specific combinations or reward and behavioral response (Matsumoto, Suzuki, & Tanaka, 2003; Wallis & Miller, 2003). In addition, medial frontal neurons (including the anterior cingulate cortex [ACC]) have been found to predict set-switching behavior under conditions of changing rewards (Williams, Bush, Rauch, Cosgrove, & Eskandar, 2004; Shima & Tanji, 1998).

[1]Cambridge University, UK, [2]MRC Cognition and Brain Sciences Unit, Cambridge, UK, [3]MRC Behavioural and Clinical Neurosciences Institute, Cambridge, UK, [4]Washington University

Can these global and local models of the reward–behavior interactions be reconciled? The answer is yes, if there are specific interactions within distinct but widely distributed brain networks. In other words, reward effects may modulate activity in local regions, but they may also alter the nature of interactions (i.e., functional connectivity) between widespread regions that form task-specific processing networks. This type of functional neural circuit model can be tested with functional magnetic resonance imaging (fMRI), provided that the brain activation related to rewards/reward expectations can be distinguished from the activation related to task-specific cognitive processing. Much of the evidence regarding reward–cognition interaction comes from studies with uncertain reward contingencies or variable risk–reward ratios in the context of gambling paradigms (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; De Martino, Kumaran, Seymour, & Dolan, 2006; Paulus & Frank, 2006; O'Doherty et al., 2001; Elliott, Friston, & Dolan, 2000). Less is known about how predictable rewards modulate the control of cognition or behavior. During instrumental learning of rules, for example, the ventral striatum correlates with reward prediction error (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). However, under conditions where the rule is already learned, and thus, reward levels are expected, there should be no prediction error, hence, no activation of the ventral striatum. This raises several questions. How then do we control responses to achieve predictable rewards? Does the magnitude or proximity of expected reward affect these cognitive control processes? How does reward-based cognition control relate to the neural mechanisms of rules and set when these are cued in advance? These three questions were addressed in the current study of healthy adults.

We investigated the effects of reward expectation on rule-based cognitive processes using a novel variant of the continuous performance task (AX-CPT) (Braver, Barch, & Cohen, 1999; Beck, Bransome, Mirsky, Rosvold, & Sarason, 1956). Human brain activation and connectivity was examined during task performance with blood oxygenation level-dependent (BOLD) fMRI. The task required simultaneous monitoring and target detection in both the spatial and verbal dimensions of visually presented stimuli pairs (see Figure 1). In addition, rewards were given whenever three successive targets in a given dimension were correctly identified. Consequently, each successively detected target in a given dimension increased reward expectancy, and potentially increased the cognitive set bias toward that dimension. Critically, the reward expectancy and bias are orthogonalized, enabling a clear separation of the activations related to reward expectation from the activations related to the biased implementation of one or other cognitive sets.

Using Statistical Parametric Mapping (SPM) of regional brain responses, we tested three specific hypotheses. First, that increasing expectation of reward across successive trials would be associated with activation of the medial prefrontal cortex, including the most rostral part of the ACC; no activation was predicted in the ventral striatum due to the predictable nature of the reward contingency. Second, when reward expectancy was linked to a bias toward the spatial dimension, there would be greater activation in a spatial processing network, including the superior frontal sulcus (SF) and the intraparietal cortex (IP) (Sakai & Passingham, 2003; Corbetta, Kincade, & Shulman, 2002). Third, when reward expectancy biased the verbal dimension, there would be greater activation of a verbal processing network including left Brodmann's area 45 (PFv) and the fusiform gyrus (FG) (Sakai & Passingham, 2003; Fujimaki et al., 1999). We also predicted that bias in verbal and spatial dimensions would be associated with specific changes in cortico-cortical connectivity within and between the verbal and spatial processing networks. This last hypothesis was tested using dynamic causal modeling (DCM) within a distributed network of frontal, parietal, and temporal brain regions.

METHODS

Behavioral Tasks

The task was based on the AX-continuous performance task (Braver et al., 1999; Beck et al., 1956), but included letter stimuli that were defined both by their location

Figure 1. The task comprised successive letters at different locations, with a pair of letters forming a given trial. Each trial could be a target (called "AX") in either a spatial ($AX_{spatial}$: any letter at 6 o'clock then 3 o'clock) or verbal (AX_{verbal}: letter A then X) dimension, or could be a nontarget ("non-AX"). Nontarget trials might include a target-relevant precue (AY) or a response cue (BX) or neither (BY) in one or other dimension. For a target trial in one dimension, there may nonetheless be target-relevant information in the other dimension. In the example shown, the first trial is a target defined as "$AX_{verbal}BY_{spatial}$."
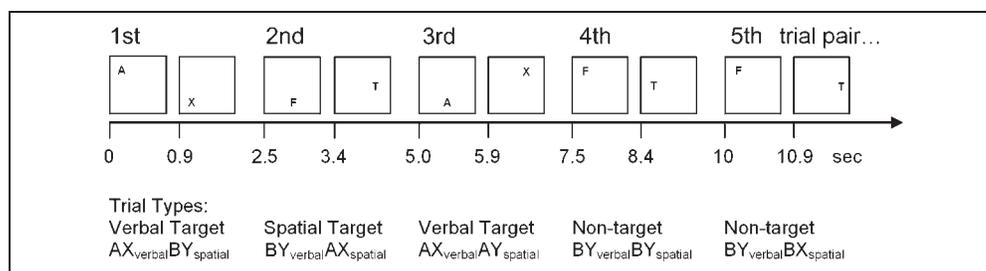
(eight positions, in a circle, radius 1.5°, white on a gray background) and letter identity (one of eight capital letters). Stimuli were presented sequentially in pairs, each pair comprised a precue and a response cue to form a single trial. Stimuli were presented for 300 msec, with an interval of 600 msec. Trial onset asynchrony was 2500 msec.

Trials were either target or nontarget trials. Target trials require that both the precue and the response cue are the specific stimuli that are taught to be associated with targets. There is a target-associated precue called "A," and a target-associated response cue called "X." All other precues are not associated with targets and are collectively known as "B." All other response cues are not associated with targets, and are collectively known as "Y." Therefore, for any stimulus dimension, the pair of stimuli that make up a trial gives rise to four formal trial types: AX, AY, BX, BY. It should be noted that a verbal "AY" trial does not necessarily present the letter "y" as the response cue but any letter that is not the target-associated response cue. This formal notation is applicable to the verbal and spatial dimensions: It should be noted that spatial "AX" trials do not necessarily present the letters "a" and "x" but may present other letters in the spatial positions that are associated with a target in the spatial dimension.

Target trials were defined in the verbal dimension by the letter "A" followed by the letter "X," appearing in any locations. Target trials were defined in the spatial dimension by letters appearing at 6 o'clock then 3 o'clock. A double target (i.e., an A at 6 o'clock followed by an X at 3 o'clock) was never presented, although target trials could include an ambiguous precue (the letter "A" at 6 o'clock) or an ambiguous response cue (the letter "X" at 3 o'clock). Thus, the target dimension and the nontarget dimension can be defined for each trial. All target trials can be denoted "AX" trials. In contrast, the nontarget dimension can be described by one of three forms: AY, BX, BY. B denotes any precue that is not associated with the targets. Y denotes any response cue that is not associated with targets.

We add a suffix to each trial descriptor in order to indicate the dimension for each type of trial (AX, AY, BX, or BY). For example a trial containing an F at 6 o'clock, followed by a T at 3 o'clock is defined as an $AX_{spatial}BY_{verbal}$ trial (see Figure 1 for further examples). Some trials contained no target-relevant information. Such $BY_{spatial}BY_{verbal}$ trials are also termed neutral trials. Trials in which a target pair was not presented (i.e., nontarget trials) may, nonetheless, contain some spatial or verbal cues that are associated with targets (i.e., an "A" *or* "X" or a stimulus presented at 6 *or* 3 o'clock). These trials may also be classified as AY, BX, or BY trials for either dimensions as above.

Overall, 33% were spatial targets ($AX_{spatial}$), 33% were verbal targets (AX_{verbal}), and 33% of trials were nontargets. Subjects indicated whether a trial was a target or a nontarget by pressing their right index or middle finger, respectively. An audible click acknowledged their correct button press. Subjects were instructed that successful detection of three sequential targets within a given dimension would lead to a monetary bonus, and thus, should be a priority in performance. Rewarded trials were each associated with a 10-pence bonus (paid after scanning) and a salient cash register sound ("ka-ching").

A permuted trial sequence was used, such that the probability of a trial being a target or nontarget, in the spatial or verbal dimension, depended on the combination of the previous two trials. This increased the frequency of consecutive target repetitions (21% of trials were a second target in a given dimension and 10% of trials were a third target in a dimension) while preserving the subjective unpredictability of forthcoming trial types among naive participants. A new target trial in one dimension sets the current progress toward reward to Level 1, increasing to 2 then 3 with each successive target trial. Three successive targets in the same dimension defined a rewarded trial. A rewarded trial was always followed by a nontarget trial (to avoid potential postreward refractory effects).

In the main experiment, 28 subjects (mean age = 24 years, range = 19–36, 12 women) were pretrained first with each trial type separately, then with 120 intermixed trials. They then performed 441 trials during fMRI scanning. Short rest blocks, 14 sec, were included after every 14–18 trials. The timing of rest intervals varied to avoid splitting a sequence of target trials.

The presentation of data was controlled using Cogent 2000 software (www.vislab.ucl.ac.uk/Cogent2000) using Matlab 7.1 (www.mathworks.com) in Windows XP (www.microsoft.com). Reaction time (RT) to presentation of the second stimulus and the accuracy of target detection were recorded. RT and arcsin-transformed accuracies were analyzed in SPSS 11.0 (SPSS, Chicago). Repeated measures analyses of variance were performed both on the accuracy and RT data. For the analyses of variance, target trial types (nontarget, first AX, second AX, third AX) and nontarget trial types (BY, BX, AY) for each dimension (spatial and verbal) were within-subject variables.

A separate behavioral experiment was performed prior to scanning (see Supplementary Material). In addition to replicating the main results, this behavioral experiment showed that key behavioral effects are enhanced by explicit financial reward for successful target detection, as used in the magnetic resonance imaging (MRI) experiment. Thus, the behavioral data validate the use of reward expectancy as a means of inducing endogenous (i.e., uncued) sources of cognitive set bias.

MRI Data Acquisition and Analysis

A Siemens Tim Trio 3-Tesla scanner was used at the Medical Research Council's Cognition and Brain Sciences Unit, Cambridge. fMRI used BOLD-sensitive T2*-weighted

EPI images (TR = 2000 msec, TE = 30 msec, FA = 78°) with 32 slices, 3.0 mm thick, in-plane resolution 3 × 3 mm, with slice separation 0.75 mm, in sequential descending order. A total of 780–810 images were acquired for each subject, the first six of which were discarded to allow for steady-state magnetization. Subjects also underwent high-resolution magnetization prepared rapid gradient echo (MP-RAGE) scanning (TR = 2250 msec, TE = 2.99 msec, FA = 9°, IT = 900 msec, 256 × 256 × 192 isotropic 1 mm voxels) and single volume TSE (TR = 5060 msec, TE = 102 msec, FA = 140, 28 × 4 mm slices) for the purposes of normalization of images, localization of activations on individual and group brains, and assurance of structural normality. Temporary technical problems introduced frequent artifacts in the fMRI data for eight consecutive subjects: These subjects' MRI data were therefore excluded prior to preprocessing.

Data analysis used SPM5 (www.fil.ion.ucl.ac.uk/spm) in Matlab 7 environment (R14, Mathworks, CA). fMRI data were converted from DICOM to NIFTII format, spatially realigned to the first image, and sinc interpolated in time to the middle slice to correct acquisition delay. The mean fMRI volume and MP-RAGE were coregistered using mutual information, and the MP-RAGE segmented and normalized to the Montreal Neurological Institute T1 template in SPM by linear and nonlinear deformations. The normalization parameters were then applied to all spatio-temporally realigned functional images, the mean and structural images, prior to spatial smoothing of fMRI data with an isotropic Gaussian kernel full-width half-maximum 10 mm.

First-level statistical parametric modeling for each subject used a general linear model with one regressor representing the presentation of a trial (of any type). This was subject to parametric modulation according to the degree of reward expectation (0, 1, 2, or 3); bias (0 for neutral; 1, 2, or 3 for increasing bias toward spatial dimension; −1, −2, or −3 for increasing bias toward verbal dimension). Nontarget trials reset the reward and bias covariates to zero. The bias and expectancy values were also reset by a new target trial in a different dimension, thus overwriting any opposite bias established by targets in the other dimension in previous trials. Therefore, there was no permitted overlap in bias to the competing dimensions. We did not model the differences in trial type in the nontarget dimension or differentiate the different types of nontarget trials.

In addition, covariates were included to indicate rewarded trials. In this way, the neural correlates of reward expectation can be distinguished from the neural correlates of rewarded trials. We also included separate regressors for error trials whether errors of omission or commission; and regressors expressing the "target reset" trials described above for which there was reversal of the biased dimension, and thus, an extradimensional shift of attention (e.g., a first spatial target AX-spatial followed by a first verbal target AX-verbal). The first-level model also included the RT for the trials as a regressor (msec), enabling the estimation of effects of expectation and bias without their being confounded by the potentially different RTs between trials of high expectation or bias. These covariates constitute "effects of interest" in the SPM.

The first-level model also included regressors representing occasional movement events and radio-frequency artifacts if they were detected by in-house image quality control software (typically 0–5 such events per subject). These regressors were considered "effects of no interest." Such single-scan regressors effectively remove these images from the estimation of parameters for the covariates of interest. The model also used a high-pass filter with a cutoff of 128 sec, and AR(1) modeling of temporal autocorrelations. Contrast images for each effect of interest were made for entry to second-level analyses.

Second-level models (random effects) for each effect of interest were made using one-sample *t* tests of the contrast images from each subject's analysis at the first level. Given the similar design of first-level analyses, this two-step approach is equivalent to a mixed effects analysis incorporating within- and between-subjects variance. SPM(t) maps were generated for each effect of interest, thresholded such that the false discovery rate (FDR) was 0.05 for whole-brain comparisons. In view of our hypothesis regarding bias toward the verbal dimension, the contrast of verbal bias at the second level was subject to a second threshold, correcting for multiple comparisons within an anatomically defined prespecified region of interest. This region included the left Brodmann's area 45 maximal probability map (Amunts et al., 2004) combined with the left fusiform gyrus, drawn according to anatomical landmarks using MRIcro (www.mricro.com).

Analysis of Effective Connectivity

To understand the changing pattern of connectivity across the brain that mediated the behavioral and regional activation effects, we used DCM of fMRI data (Friston, Harrison, & Penny, 2003). The rapid event-related design of this experiment makes it well suited to DCM (analysis of interactions between the neural processes estimated from the deconvolution of the BOLD MRI signal) but unsuitable for structural equation modeling or psychophysiological interactions (PPIs) expressed in general linear models. DCM models include all data points spanning all trial types to optimize the simultaneous solution for all coupling parameters, modulatory bilinear parameters, and hyperparameters determining the BOLD response to neural activity.

DCM permits analysis of changes in brain network connectivity, in terms of bilinear effects or moderator terms that express the psychological context within which physiological interactions mediate task performance. It is a hypothesis-driven approach, testing specific effects within

defined neuroanatomical networks. To construct a DCM, one must determine the regions of the putative network; the permitted connections between them; the region(s) through which the network is perturbed during trials; and the paths in which there is differential coupling according to the context.

The Regions

For our models, fMRI time series from seven regions of interest were extracted for each subject. Specific coordinates were selected from activation foci from second-level SPM analyses of reward expectancy, verbal bias, spatial bias, and all-tasks-versus-rest. Regions were centered on: −50, 38, 8 (ventral lateral prefrontal cortex [PFv] within maximal probability map for area 45); −46, 28, 22 (dorsal lateral prefrontal cortex [PFd], area 46); −8, 44, 4 (medial frontal cortex [MF], area 24); −20, −2, 54 (superior frontal sulcus [SF], area 8); −30, −90, 0 (prestriate [PS], area 19); −38, −62, −20 (fusiform gyrus [FG]); and −14, −52, 54 (intraparietal cortex [IP], area 7). Time series were based on the mean of voxels from spheres 7 mm radius which exceeded $p < .05$ for an F test of all effects of interest. Time series were adjusted for all experimental effects of interest, preserving noise but removing the effects of specified spikes or movement artifacts. Two subjects showed no significant activation in one or more specific regions and were therefore excluded from further network modeling.

The Connections

There is insufficient evidence of direct and reciprocal connections from human neuropathology alone to validate the anatomical connectivity of most models. However, there is evidence from the macaque, if one accepts functional and connectionist homologies associated with cytoarchitectonic homology. With recent MRI evidence of such functional homology in many, if not all, parietal and prefrontal regions (Nieder & Miller, 2004; Orban, Van Essen, & Vanduffel, 2004), we used macaque-based data to support our anatomically constrained dynamic causal model (www.cocomac.org).

The Driving Input to the Network

Within DCM models, brain regions usually change their activity only in response to inputs from other brain regions, not as a direct result of external events. Exceptions include sensory regions that are subject to external influences. Task-related activation propagates throughout the network from such sensory regions, following visual stimulation, perturbing a network which would otherwise return gradually to a steady state. This initial perturbation of the network was modeled by task event stimulations of the prestriate cortex using a covariate expressing all trial types' events (DCM external influ-

ences matrix **C**: sometimes termed "injection" of perturbation into the network).

The Bilinear Effects

Differential propagation of task related activation may occur under different trial types or contexts. Such differential propagation is specified within a DCM as pathways for which the connectivity time-constants were permitted to vary according to spatial or verbal bias (known as bilinear effects, defined by DCM matrix **B**, expressing a PPI; see Figure 4). Two bilinear effects expressed verbal bias and spatial bias, respectively.

To study the group connectivity, we employed a second-level analysis of the dynamic causal models. This used the Bayesian model "averaging" approach in SPM5 to compute the posterior mean and precision from the whole group for all connectivity parameters. The posterior mean of a connectivity parameter from previous subjects is used as the prior for the next subject. The posterior mean is a weighted combination of the prior mean (following previous subjects) and the likelihood (datum from current subject), where the weighting is given by the relative precisions (inverse variance/covariances) of the prior and the data. This is not necessarily the arithmetic mean of parameters across subjects because of possible covariance among parameters, but does represent the most likely value given the data.

The results of the group DCM are presented as connection weights (given by the time constants in Hz) for intrinsic connections and bilinear moderators for which there was a high confidence (>90%). Positive effects were tested post hoc by a one-sample two-way t test of single-subject values for a bilinear time constant with the null hypothesis that the mean was zero. A positive value for an intrinsic connection indicates that the source region causes an increase in activity in the target region (per unit source activity per unit time). A positive value for a bilinear connection indicates that the psychological variable, say spatial bias, is associated with an increase in connectivity between two regions, such that a unit of activation in the source regions causes a greater increase in activity in the target region per unit time.

It would be impossible to conduct a comparison of all ~10^{15} models that could be constructed from our seven regions and two moderator variables, with feedforward and/or feedback connections. Therefore, we instead assessed four nested plausible models based on the same anatomical connectivity, but with different bilinear effects representing different patterns of PPI. The models included the same anatomical regions and time series, and the same intrinsic connectivity. However, the models differed in terms of the possible ways in which connectivity changed according to the context of each trial (defined by the bias induced by reward expectation).

The different models were computed for which: (a) Verbal bias may modulate connections between PS, FG, PFv, and PFd, and between PFv and IP, whereas spatial bias may modulate connections between PS, IP, SF, and PFd, and between PFv and IP, illustrated in Figure 4 and Supplementary Figure S1; This was the model most supported by the evidence of the data. Of note, the connections of the MF are of constant strength, although it should be recalled that the activity in this region varies with reward expectation enabling a variable influence on target regions even with fixed connectivity. (b) All connections, except those to or from the MF, are modifiable by verbal and/or spatial bias; This model is more flexible by allowing independent effects of each bias on any connection. However, it requires more degrees of freedom to do so. (c) Like model a, but with additional permitted modulation of all connections of the MF by spatial and/or verbal bias. This third model allows a second means by which the MF could alter activity in its target regions: by increased activity within the region and/or increased effective connectivity to target regions. (d) Like model b with independent effects of each bias type on each connection, including permitted modulation of all connections of MF by spatial and/or verbal bias. This model is most flexible, but may be less likely if it is overspecified for the data.

To summarize the four models, models c and d allow variable connections of the MF cortex, whereas models a and b do not. Models b and d allow both spatial and verbal bias to modulate the strength of all task-processing connections, whereas models a and c differentiate those connections that can be modified by verbal bias from those that can be modified by spatial bias.

Reward expectation entered the DCM vicariously, as permitted joint bilinear effects of spatial and verbal biases, and as regional activation of the MF. Akaike's Information Criterion and Bayesian Information Criterion values were used to compute Bayes Factors (BF) for the relative evidence in favor of one model versus another (Penny, Stephan, Mechelli, & Friston, 2004b), assuming each model had equal a priori probability.

RESULTS

Behavioral Results

There was clear behavioral evidence for graded cognitive set shift toward reward-relevant dimensions (Figure 2). For both verbal and spatial dimensions, performance was improved when the current trial was a target in a biased dimension, and was further improved according to the number of previous targets in that modality, in terms of both faster RTs [RT-spatial: $F(3, 25) = 37$, $p < .001$; verbal: $F(3, 31) = 30$, $p < .001$; see Figure 2] and greater accuracy [Acc-spatial: $F(3, 25) = 39$, $p < .001$; verbal: $F(3, 25) = 43$, $p < .001$]. Overall accuracy across scanned subjects was 97.5% ($SD = 0.02$; mean arcsin

Figure 2. The top panel shows reaction times (mean RT ± *SE*) to spatial and verbal trials according to whether they were nontarget trials (neutral BY trials, with neither first nor second letter/location associated with a target) or contained a first, second, or third successive target (AX) in the relevant dimension. The lower panels show RTs to each type of stimulus pair, according the position in a succession of targets (BY in target dimension, first target AX, second target AX, third target AX) separated according to the modulation of trial types in the nontarget dimension (BY, BX, AY). Subjects became faster with increasing reward expectation, and also insensitive to modulation of the nontarget dimension.

accuracy = 1.35, $SD = 0.07$). Accuracy increased with target repetition. For first, second, and third targets, accuracy rose from 96.5% to 99.2% and 99.2%, respectively for spatial targets, and from 95.0% to 98.6% to 99.2%, respectively, for verbal targets.

Importantly, the subjects were not merely more generally attentive to the stimuli, but their attention became selective for one dimension with diminishing sensitivity to trial-type differences in the alternate modality. For example, the effect of expectation the trial precue led to

an expectation of a target in the other dimension (AY), interacted with the type of trial in the *nontarget* dimension (AY, BX, BY) [RT-spatial: $F(6, 22) = 9.4$, $p < .01$; RT-verbal: $F(6, 22) = 3.4$, $p < .01$]. Thus, when a target trial had not been preceded by previous targets, performance was slower when compared to when the precue did not lead to such expectations (BX, BY). However, as the number of previous targets in a given dimension increased, the effect of the precue in the other dimension diminished, as indicated by a significant interaction between the number of previous targets and the trial type of the nontarget dimension for both spatial and verbal targets. These behavioral effects were diminished when the task was performed under non-rewarded conditions, suggesting that reward expectancy modulated attentional control in task-specific processing networks (see Supplementary Results).

Neuroimaging Results: Regional Activations

Analysis of regional activations revealed by fMRI confirmed that increasing expectation of reward across successive targets was associated with increased activation of the medial frontal lobe including the cingulate cortex (peak $-8, 44, 4$, $t = 5.82$, FDR < 0.01, cluster 488 voxels) and bilaterally in the superior temporal gyrus (peak 48, $-18, 8$, $t = 6.73$, FDR < 0.01, cluster 1204 voxels; peak $-50, -20, 4$, $t = 6.27$, FDR < 0.01, cluster 974 voxels) (see Figure 3A). The activation of this anterior cingulate

region is comparable to the single-cell neuronal activation observed with increasing reward expectancy in monkeys (Shidara & Richmond, 2002) and correlation with reward value in humans (O'Doherty et al., 2001). It has been suggested that in the context of task switching paradigms, rostral cingulate activation is related to processing response conflict (Swainson et al., 2003). However, in our paradigm, reward expectation is associated with reduced conflict between spatial and verbal processes.

When the reward expectancy leads to a bias toward the spatial dimension, we predicted greater activation in a dorsal network including the superior frontal sulcus (SF) and the intraparietal cortex (IP), and when the verbal dimension was biased by reward expectancy, we predicted greater activation in a ventral network. These predictions were confirmed as shown in Figure 3B and C (details in Supplementary Table S1).

A possible explanation of the activations associated with spatial bias is that there were differential eye movements between conditions. Although we were unable to record eye movements, we feel that this explanation is inadequate. The caudal prefrontal activation of spatial bias lies anterior to the usual location of the frontal center for control of saccades (Petit, Clark, Ingeholm, & Haxby, 1997). The cortex adjacent to this region has a reported role in the control of target selection (Schall, 2002) or spatial working memory (Curtis, Sun, Miller, & D'Esposito, 2005), consistent with an attentional set bias

Figure 3. SPM(t) maps of regional activations associated with different conditions: (A) increasing reward expectancy in the left medial frontal cortex including the rostral ACC (BA 24/32); (B) increasing verbal bias in the ventral lateral frontal cortex (BA 45); and (C) increasing spatial bias in the superior frontal sulci (BA 8) and the intraparietal cortex. SPM(t)s are thresholded at FDR $p < .05$ and are superimposed on the group-average MP-RAGE structural image, with contrast-specific color bars corresponding to t values. (D) Single-subject (#17) BOLD data from three voxels of peak activation in the medial frontal cortex, the intraparietal cortex, and the ventral prefrontal cortex

peaks (solid lines), on trials for which there was verbal or spatial bias to a low (1), middle (2), or high (3) degree. Trials of low and medium bias were included for these plots if they were not followed by a subsequent trial of greater bias. Data shown are averaged BOLD signals 6 sec after the onset of trials (expected peak the delayed BOLD response), adjusted only for spike and movement artifacts. Linear regression estimates are shown in dotted lines, split for the medial frontal data according to verbal and spatial bias. (i) The medial frontal cortex is more active for trials in which there is higher bias to either modality, indicating greater reward expectancy. (ii) Ventral prefrontal activity is higher with verbal bias. (iii) Parietal cortical activity is higher for trials with spatial bias than verbal bias.
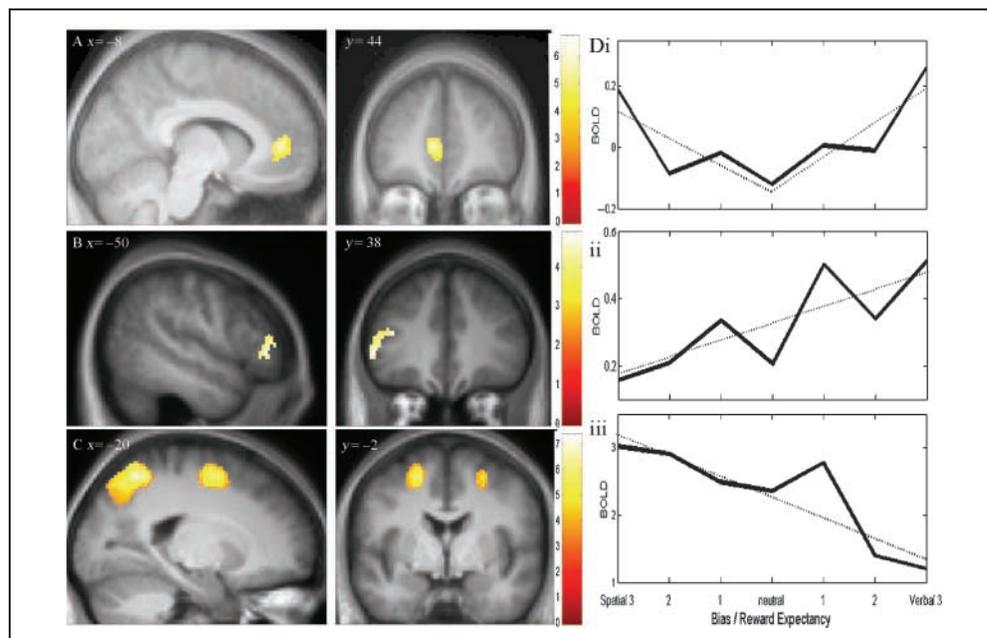
without additional eye movements. Bias toward the verbal dimension increased activation associated within the cytoarchitectonically defined Brodmann's area 45 (Amunts et al., 2004). Verbal bias was also associated with activation of the fusiform gyrus, which may include a ''word form area'' (Price, Wise, & Frackowiak, 1996) that responds to letters and letter strings more than digits, nonletter patterns, or symbols.

Neuroimaging Results: Changes in Effective Connectivity

The most likely one of the tested DCM models included the dorsolateral prefrontal cortex (PFd, area 46). This region was proposed because of its previous association with cognitive set formation (Dalley, Cardinal, & Robbins, 2004; Sakai & Passingham, 2003; MacDonald et al., 2000a; Rogers et al., 2000), despite the lack of activation in the analysis of regional activation above. Intrinsic connections were modulated by graded spatial or verbal bias, expressed by the bilinear components of the model (DCM matrix **B**). Figure 4 presents the model which had the highest posterior probability in all 18 subjects by Bayes Information Criterion and the highest probability in 16 subjects by Akaike's Information crite-
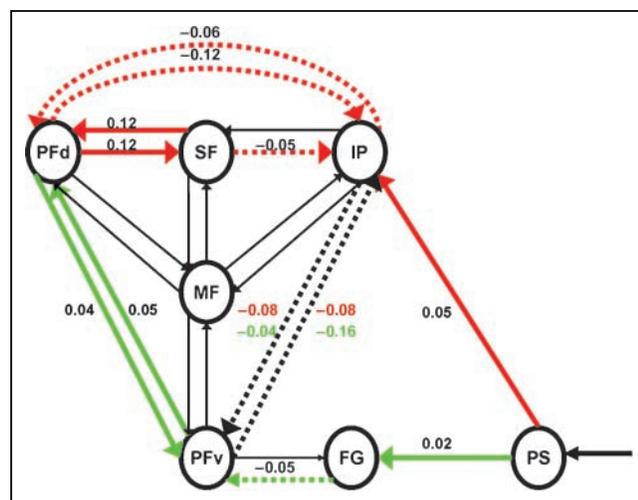
Figure 4. The network shows bilinear effects representing psychophysiological interactions in the preferred dynamic causal model that included the medial frontal (MF) cortex, the dorsal (PFd) and ventral (PFv) lateral prefrontal cortex, the superior frontal sulcus (SF), the intraparietal cortex (IP), the fusiform gyrus (FG), and the prestriate cortex (PS). Values given are time constants (Hz) for the bilinear influences for which the group posterior mean was positive (solid lines) or negative (dashed lines) for verbal bias (thick green), spatial bias (thick red), or both (thick black). These bilinear effects have strong evidence that they are nonzero, corresponding to Bayes factor >20 [p(model|data) > 95%]. Post hoc t tests confirmed the group mean bilinear effect was greater than zero for the connections between PS and FG ($t = 2.3$, $df = 17$, $p < .05$) and between PFd and SF ($t = 2.3$, $df = 17$, $p < .05$).

rion. This suggests that it is the most likely model given the data. All intrinsic connections (matrix **A**) were significant (Supplementary Figure S1) and mostly positive. Visual presentation of trial stimuli modulated prestriate cortex connectivity in a positive direction, as expected (matrix $\mathbf{C}_{\text{trials to PS}} = 0.06$, $p = 1.000$).

The network model revealed a clear pattern of changes in connectivity resulting from bias induced by reward expectancy. Bias toward the verbal dimension was associated with stronger connectivity from the prestriate cortex to the fusiform gyrus (Figure 4). Although a feedforward connection, the facilitation of its connectivity is likely to be due to top–down mechanisms of cognitive control, in view of the constant nature of the stimulus properties. There was also enhanced reciprocal connectivity between the PFv (area 45) and the PFd (area 46) but reduced reciprocal connectivity with the intraparietal cortex (area 7). It is interesting to note that despite stronger connectivity in the early visual pathway, and between frontal cortical regions, there was reduced connectivity in the intermediate stage between the fusiform gyrus and the PFv. Bias toward the spatial dimension was associated with an analogous effect of stronger local connectivity within the dorsal stream, with both feedforward facilitation of the connection from the prestriate cortex (area 19) to the intraparietal cortex (area 7), and greater reciprocal connectivity between the superior frontal cortex (area 8) and the PFd (area 46) as illustrated in Figure 4. There was a reduction in connectivity between the parietal cortex and its connections in the PFv and the SF.

In the preferred model, there was no change in the modulatory connections between the ACC and the targets of its projections. This does not imply that there is no change in influence of the anterior cingulate with reward expectations. This highlights the complementary insights gained by connectivity models and standard analyses of regional effects. DCM seeks to explain regional effects in terms of changing patterns of connectivity in a network that is perturbed by an experimental event. Recall that there is increased activation of the anterior cingulate with increasing reward expectation. With fixed connection strengths, greater activity at the source will lead to greater activity at the targets. In our case, increasing activity of the rostral cingulate would lead to greater activation of the ventral prefrontal cortex and the intraparietal cortex. It may help to think of an electrical circuit analogy, in which a higher voltage leads to a stronger current across a connection even if the conductance/resistance of the connection is unchanged.

A second difference is illustrated by the role of the dorsal lateral prefrontal cortex. This was not identified by the main SPM contrast of ''all trials versus rest'' but it does have changes in connectivity with cognitive bias in the DCM analysis. This is because a region with no net activation following the network perturbation (by SPM contrasts of differential activation between conditions

in a group) may, nonetheless, have significant afferent connection. Differential connectivity without differential activity has been noted in other connectivity analyses using DCM, structural equation modeling, and PPIs within general linear models (Sonty et al., 2007; Stephan et al., 2003; Rowe et al., 2002). In our DCM model, the areas 45 and 8 had opposing influences on area 46, which may have reduced the net regional activation of area 46.

DISCUSSION

The novel AX-CPT task enabled us to study how the proximity to reward over successive trials led to transitions between verbal and spatial cognitive sets. Subjects were able to accurately monitor the stream of visual stimuli for targets determined by two concurrent rules. We propose that a key aspect to proximity to reward in the current design is the expectation of reward (cf. Shidara & Richmond, 2002). The behavioral data confirm that such reward expectation over three successive targets enhanced speed and accuracy of responses due to a specific enhancement of the relevant cognitive set. This cognitive set bias was not caused by a separate rule cue prior to the trial stimuli, but was due to information implicit in the trial stimuli themselves. Moreover, the bias was not dichotomous, but represented a continuum between two contrasting rules.

The key feature of this task, in contrast to previous studies, is that it successfully dissociated the correlates of expectation of rewarded trials from the resulting changes in cognitive set. As predicted, the two cognitive sets had distinct neural substrates characterized by specific changes in regional activation and connectivity. The balance between these two spatial and verbal sets was governed by the reward expectation, that is to say, the proximity to reward across successive trials. Within the frontal lobe, this goal expectation was encoded by the medial frontal cortex, including the most anterior part of the cingulate cortex.

Changes in Brain Activation

These fMRI results explain how subjects continuously adapt the bias between cognitive sets according to the expectancy of reward as represented in the medial frontal cortex. The effects of reward expectation were seen in interconnected regions of the dorsal and ventral lateral prefrontal cortex. These changes are similar to those that follow explicit advanced categorical task cues, but which convey no information about reward (Sakai & Passingham, 2003, 2006). It is probable that these changes include activity of the class of prefrontal neurons that has been shown to encode specific combinations or reward and behavioral response (Matsumoto et al., 2003; Wallis & Miller, 2003).

There were additional changes in a wide set of lateral brain regions beyond the frontal lobes. Although these

were distributed changes, our results do not suggest a ''global workspace'' (Dehaene & Changeux, 2000; Dehaene et al., 1998) for the integration of expected rewards with the cognitive strategies necessary to achieve them. Rather, specific changes in activation occurred in distinct dorsal and ventral networks according to the appropriate task bias. Such sensitivity to reward has also been found in the extrastriate cortex in the context of working memory (Krawczyk, Gazzaley, & D'Esposito, 2007).

Our paradigm differs from previous studies of rule or set based behaviors. First, the set bias induced by reward expectation need not have improved performance on subsequent trials. For example, approximately 10% of trials were rewarded (three targets in a row in one dimension), but there were also 10% to 14% reversal events in which a target in the unbiased dimension was presented (brain activation related to reversal events is described in the supplementary material). Failure to respond correctly on this reversal trial could prevent one receiving the next reward. Thus, the strategy of increasing attentional bias on the basis of proximity to reward is only one possibility that participants may have adopted to optimize reward accumulation. Nevertheless, it is likely that this strategy may have been the default because target presentation in one dimension automatically increases the salience of that dimension. This increased saliency may have been further reinforced by the parallel increase in proximity to reward.

A second important difference is the graded or parametric modulation of bias toward one or other set, rather than the simple dichotomy invoked by cued set paradigms (Sakai & Passingham, 2003; MacDonald et al., 2000a). This is evident in raw data plots from individual subjects following each trial type (Figure 2D). A third difference is that the cognitive bias toward a rule is established without separate rule cues. In previous studies, a specific cognitive set is often established by cueing of the rule (Crone, Wendelken, Donohue, & Bunge, 2006; Bunge, 2004; Sakai & Passingham, 2003; MacDonald, Cohen, Stenger, & Carter, 2000b). Such cueing may occur in daily life (imagine sitting at the traffic lights when the amber light comes on) but it is often not the case.

These earlier studies left open the question of how the different regions of the prefrontal cortex interact to produce rule-guided behaviors in the absence of separate or advanced rule cues. One possibility is that the same regions encode goals and rules. For example, in the lateral prefrontal area 46, there are neurons that encode abstract rules (Wallis, Anderson, & Miller, 2001) and also neurons for which spatial working memory related activity is enhanced by reward expectations (Tsujimoto & Sawaguchi, 2004; Leon & Shadlen, 1999). Alternatively, a functional link between medial/ventral prefrontal reward systems and lateral rule/set systems could arise if both systems contribute to, and draw from, a shared distributed ''global workspace'' (Dehaene &

Changeux, 2000; Dehaene et al., 1998), which may include an activated cingulate cortex during goal-oriented rule-based tasks (Landmann et al., 2006). A third possibility is that there exist distributed changes in the specific brain networks mediating different aspects of the task. For example, in a gambling-task context, the coupling between the orbital and lateral prefrontal cortex increases on trials that have increased risk and reward (Cohen, Heller, & Ranganath, 2005). Our results support the latter possibility, but such a connectionist hypothesis cannot properly be tested by classical imaging analyses of regional specialization. Instead, a connectionist hypothesis calls for a formal analysis of brain connectivity.

Changes in Brain Network Connectivity

We therefore went on to determine the changes in brain network connectivity associated with reward expectancy. Our analysis approach utilized DCM (Friston et al., 2003) to determine the presence of task-induced changes in effective connectivity. There were clear alterations in effective connectivity in the dorsal and ventral streams. These were not restricted to the prefrontal cortex but included the early visual pathways reflecting changes in the processing of reward-relevant dimensions of visual stimuli. It is interesting to note that the reward expectation led to changes in connectivity among regions that were not themselves directly modulated by reward expectancy. This is quite different from previous reports that have understandably focused on changes in connectivity among the medial and orbital frontal regions that are also individually correlated with reward (Cohen et al., 2005).

It is likely that the modulation of connectivity in early feedforward connections represent top–down modulation of connectivity. The dorsolateral prefrontal cortex is a candidate region for the source of such top–down control. Despite the finding that increased bias or expectancy did not modulate task-related activity in this region, the network model makes clear that it does play an important role in the distributed network governing reward-based changes in cognition. The phenomenon of cognitive control though changing connectivity rather than increased focal activation has been reported previously in cued rule-based tasks (Stephan et al., 2003) and may have been overlooked in other standard analyses of imaging data.

The validity of inferences on the basis of network models can only be as good as the models they come from. We used DCM as the most appropriate analytic method for our rapid event-related design. Alternative connectivity analysis of brain activation during reward-based behavior has been done using functional connectivity analysis on the basis of interregional correlations. This avoids the need for hypothesis-driven model specification but does not enable one to test causal relations. In contrast, DCM relies on anatomical and other prior information to construct specific network models within which one can test causal relations between regions. One can then draw reliable inferences about contextual changes in task-related connectivity (Lee, Friston, & Horwitz, 2006), which can be conceptualized as PPIs (Friston et al., 2003). The DCM model's flexibility can also accommodate a wide range of hemodynamic response functions that can vary between regions and across subjects. This, together with its dynamic rather than stationary modeling of an entire time series, made it ideal for the analysis of rapid event-related paradigms which are otherwise impossible for alternative methods such as PPIs in general linear models (Friston et al., 1997) or structural equation modeling (Penny, Stephan, Mechelli, & Friston, 2004a). The method also allows for model comparison, and a quantification of the degree of confidence associated with a preferred model, using the BIC and AIC criteria (Penny et al., 2004b). It is not practical to test all models against each other with multiple regions and cognitive moderator terms but, fortunately, the preferred model identified in this study was strongly preferred by both measures across the group.

Limitations of the Study

The design of the SPM analysis is such that the identified effects of reward expectation are independent of the events that coincide with actual reward (identification of three successive targets followed by ka-ching and money). These latter events include reward itself and the possible cessation of cognitive set on completion of the goal: We cannot separate these factors. However, the aim of the study was to understand the effect of reward expectation as one moved through successive target trials (cf. the reinforcement schedule of Shidara & Richmond, 2002). There are, nonetheless, two important caveats. First, we used a compound reward, including money to be paid after the experiment; a salient ka-ching cash register sound; and the personal reward inherent in successful completion of an arbitrary task. We cannot be certain which aspect of reward was most relevant for the modulation of cognitive set. This contrasts with the monkey studies that used primary reinforcers such as juice to record reward-related neuronal firing in the ACC (Shidara & Richmond, 2002, 2004), but is similar to other human imaging studies (O'Doherty et al., 2001, 2003). Different reward types might have different neural mechanisms, even though there is an apparent consensus in the literature on the importance of the ACC.

A second caveat is the correlation of reward expectation across successive targets with top–down attentional control. There are three possible effects here. First, that the ACC exerts top–down control itself, as a function of reward representation, like the "motivational map" proposed by Mesulam (1981). This possibility has been largely superseded by the second possibility, that the

ACC influences top–down control by interaction with other areas such as the dorsolateral prefrontal cortex (Kerns et al., 2004). Third, that ACC activation we see in association with reward expectation is due entirely to the correlated changes in top–down control, and is not directly related to reward representation (Johnston, Levin, Koval, & Everling, 2007). Given the anatomical and connectionist heterogeneity of the ACC, these possibilities may coexist, with direct attentional control exerted by more dorsal–caudal regions of the ACC (Kerns et al., 2004; Duncan & Owen, 2000; MacDonald et al., 2000b; Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999) and reward representation rostrally (O'Doherty et al., 2001, 2003). Unfortunately, our data cannot definitively exclude the third possible interpretation. However, in view of the rostral location of our cluster, we suggest that reward expectation represented in the ACC region leads to top–down bias through its interactions with other regions including the lateral prefrontal cortex.

Conclusion

The medial frontal cortex was found to be sensitive to the increasing level of reward expectancy over successive trials. This led to specific but distributed changes in terms of brain activation and connectivity, in both the dorsal and ventral streams, that mediated the relevant spatial and verbal tasks. Further work is required to confirm the generalization of our results to other goal-oriented cognitive processes, and the many different types of reward that influence human behavior. For example, there may be differential effects of monetary, salutary, or sexual rewards. However, we suggest that the combination of localizationist and connectionist approaches is especially informative to understand the behavioral effects of reward expectation through specific distributed changes in the brain.

REFERENCES

Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., et al. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space—The roles of Brodmann areas 44 and 45. *Neuroimage, 22,* 42–56.

Aron, A. R., Monsell, S., Sahakian, B. J., & Robbins, T. W. (2004). A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain, 127,* 1561–1573.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience, 18,* 428–437.

Beck, L. H., Bransome, E. D., Jr., Mirsky, A. F., Rosvold, H. E., & Sarason, I. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350.

Bolla, K. I., Eldreth, D. A., London, E. D., Kiehl, K. A., Mouratidis, M., Contoreggi, C., et al. (2003). Orbitofrontal cortex dysfunction in abstinent cocaine abusers performing a decision-making task. *Neuroimage, 19,* 1085–1094.

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402,* 179–181.

Braver, T. S., Barch, D. M., & Cohen, J. D. (1999). Cognition and control in schizophrenia: A computational model of dopamine and prefrontal function. *Biological Psychiatry, 46,* 312–328.

Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective & Behavioral Neuroscience, 4,* 564–579.

Cohen, M. X., Heller, A. S., & Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain Research, Cognitive Brain Research, 23,* 61–70.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *Journal of Cognitive Neuroscience, 14,* 508–523.

Crone, E. A., Wendelken, C., Donohue, S. E., & Bunge, S. A. (2006). Neural evidence for dissociable components of task-switching. *Cerebral Cortex, 16,* 475–486.

Curtis, C. E., Sun, F. T., Miller, L. M., & D'Esposito, M. (2005). Coherence between fMRI time-series distinguishes two spatial working memory networks. *Neuroimage, 26,* 177–183.

Dalley, J. W., Cardinal, R. N., & Robbins, T. W. (2004). Prefrontal executive and cognitive functions in rodents: Neural and neurochemical substrates. *Neuroscience and Biobehavioral Reviews, 28,* 771–784.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441,* 876–879.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision-making in the human brain. *Science, 313,* 684–687.

Dehaene, S., & Changeux, J. P. (2000). Reward-dependent learning in neuronal networks for planning and decision making. *Progress in Brain Research, 126,* 217–229.

Dehaene, S., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 14529–14534.

Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: Restriction to novel situations and independence from "on-line" processing. *Journal of Neuroscience, 17,* 9285–9297.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475–483.

Elliott, R., Friston, K. J., & Dolan, R. J. (2000). Dissociable neural responses in human reward systems. *Journal of Neuroscience, 20,* 6159–6165.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6,* 218–229.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage, 19,* 1273–1302.

Fujimaki, N., Miyauchi, S., Putz, B., Sasaki, Y., Takino, R., Sakai, K., et al. (1999). Functional magnetic resonance imaging of neural activity related to orthographic, phonological, and lexico-semantic judgments of visually presented characters and words. *Human Brain Mapping, 8,* 44–59.

Goudriaan, A. E., Oosterlaan, J., de Beurs, E., & Van den Brink, W. (2004). Pathological gambling: A comprehensive review of biobehavioral findings. *Neuroscience and Biobehavioral Reviews, 28,* 123–141.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience, 16,* 463–478.

Johnston, K., Levin, H. M., Koval, M. J., & Everling, S. (2007). Top–down control-signal dynamics in anterior cingulate and prefrontal cortex neurons following task switching. *Neuron, 53,* 453–462.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303,* 1023–1026.

Krawczyk, D. C., Gazzaley, A., & D'Esposito, M. (2007). Reward modulation of prefrontal and visual association cortex during an incentive working memory task. *Brain Research, 1141,* 168–177.

Landmann, C., Dehaene, S., Pappata, S., Jobert, A., Bottlaender, M., Roumenov, D., et al. (2006). Dynamics of prefrontal and cingulate activity during a reward-based logical deduction task. *Cerebral Cortex, 17,* 749–759.

Lee, L., Friston, K., & Horwitz, B. (2006). Large-scale neural models and dynamic causal modelling. *Neuroimage, 30,* 1243–1254.

Leon, M. I., & Shadlen, M. N. (1999). Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron, 24,* 415–425.

MacDonald, A. W., III, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000a). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838.

MacDonald, A. W., III, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000b). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control [In Process Citation]. *Science, 288,* 1835–1838.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain, 125,* 624–639.

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

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology, 10,* 309–325.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences, U.S.A., 101,* 7457–7462.

O'Doherty, J., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *Journal of Neuroscience, 23,* 7931–7939.

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4,* 95–102.

Orban, G. A., Van Essen, D., & Vanduffel, W. (2004). Comparative mapping of higher visual areas in monkeys and humans. *Trends in Cognitive Sciences, 8,* 315–324.

Paulus, M. P., & Frank, L. R. (2006). Anterior cingulate activity modulates nonlinear decision weight function of uncertain prospects. *Neuroimage, 30,* 668–677.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004a). Modelling functional integration: A comparison of structural equation and dynamic causal models. *Neuroimage, 23(Suppl. 1),* S264–S274.

Penny, W. D., Stephan, K. E., Mechelli, A., & Friston, K. J. (2004b). Comparing dynamic causal models. *Neuroimage, 22,* 1157–1172.

Pessiglione, M., Schmidt, L., Draganski, B., Kalisch, R., Lau, H., Dolan, R. J., et al. (2007). How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science, 316,* 904–906.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442,* 1042–1045.

Petit, L., Clark, V. P., Ingeholm, J., & Haxby, J. V. (1997). Dissociation of saccade-related and pursuit-related activation in human frontal eye fields as revealed by fMRI. *Journal of Neurophysiology, 77,* 3386–3390.

Price, C. J., Wise, R. J., & Frackowiak, R. S. (1996). Demonstrating the implicit processing of visually presented words and pseudowords. *Cerebral Cortex, 6,* 62–70.

Reuter, J., Raedler, T., Rose, M., Hand, I., Glascher, J., & Buchel, C. (2005). Pathological gambling is linked to reduced activation of the mesolimbic reward system. *Nature Neuroscience, 8,* 147–148.

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience, 12,* 142–162.

Rowe, J., Stephan, K. E., Friston, K., Frackowiak, R., Lees, A., & Passingham, R. (2002). Attention to action in Parkinson's disease: Impaired effective connectivity among frontal cortical regions. *Brain, 125,* 276–289.

Sakai, K., & Passingham, R. E. (2003). Prefrontal interactions reflect future task operations. *Nature Neuroscience, 6,* 75–81.

Sakai, K., & Passingham, R. E. (2006). Prefrontal set activity predicts rule-specific neural processing during subsequent cognitive performance. *Journal of Neuroscience, 26,* 1211–1218.

Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 357,* 1073–1082.

Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science, 296,* 1709–1711.

Shidara, M., & Richmond, B. J. (2004). Differential encoding of information about progress through multi-trial reward

schedules by three groups of ventral striatal neurons. *Neuroscience Research, 49,* 307–314.

Shima, K., & Tanji, J. (1998). Role for cingulate motor area cells in voluntary movement selection based on reward. *Science, 282,* 1335–1338.

Sonty, S. P., Mesulam, M. M., Weintraub, S., Johnson, N. A., Parrish, T. B., & Gitelman, D. R. (2007). Altered effective connectivity within the language network in primary progressive aphasia. *Journal of Neuroscience, 27,* 1334–1345.

Stephan, K. E., Marshall, J. C., Friston, K. J., Rowe, J. B., Ritzl, A., Zilles, K., et al. (2003). Lateralized cognitive processes and lateralized task control in the human brain. *Science, 301,* 384–386.

Swainson, R., Cunnington, R., Jackson, G. M., Rorden, C., Peters, A. M., Morris, P. G., et al. (2003). Cognitive control mechanisms revealed by ERP and fMRI: Evidence from repeated task-switching. *Journal of Cognitive Neuroscience, 15,* 785–799.

Tanabe, J., Thompson, L., Claus, E., Dalwani, M., Hutchison, K., & Banich, M. T. (2007). Prefrontal cortex activity is reduced in gambling and nongambling substance users during decision-making. *Human Brain Mapping, 28,* 1276–1286.

Tsujimoto, S., & Sawaguchi, T. (2004). Neuronal representation of response-outcome in the primate prefrontal cortex. *Cerebral Cortex, 14,* 47–55.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001). Single neurons in prefrontal cortex encode abstract rules. *Nature, 411,* 953–956.

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience, 18,* 2069–2081.

Williams, Z. M., Bush, G., Rauch, S. L., Cosgrove, G. R., & Eskandar, E. N. (2004). Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nature Neuroscience, 7,* 1370–1375.