Resolution of Uncertainty in Prefrontal Cortex

Wako Yoshida[1] and Shin Ishii[1,*]

[1] Graduate School of Information Science
Nara Institute of Science and Technology
8916-5 Takayama, Ikoma, Nara 630-0192
Japan

Summary

Making optimal decisions in the face of uncertain or incomplete information arises as a common problem in everyday behavior, but the neural processes underlying this ability remain poorly understood. A typical case is navigation, in which a subject has to search for a known goal from an unknown location. Navigating under uncertain conditions requires making decisions on the basis of the current belief about location and updating that belief based on incoming information. Here, we use functional magnetic resonance imaging during a maze navigation task to study neural activity relating to the resolution of uncertainty as subjects make sequential decisions to reach a goal. We show that distinct regions of prefrontal cortex are engaged in specific computational functions that are well described by a Bayesian model of decision making. This permits efficient goal-oriented navigation and provides new insights into decision making by humans.

Introduction

Surviving in a complex environment depends critically on the ability to make planned, often sequential decisions to realize future goals. The information needed to make these decisions is often uncertain, however, posing a major challenge to the animal. One of the least understood sources of uncertainty is that associated with the identity or nature of the current state. This occurs commonly in navigation tasks (Stankiewicz et al., 2006). For example, if we are trying to find our way back to our hotel in an unfamiliar city, decisions may be difficult because we don't know with certainty where we are—not because we lack the ability to plan routes. These situations are referred to as partially observable decision-making problems (Cassandra et al., 1994). Despite considerable recent interests in the neural basis of decision making (Sugrue et al., 2005; Wood and Grafman, 2003), the strategy used by the brain to solve this type of problem is both poorly studied and little understood. In navigation tasks, such as that investigated here, an individual must constantly maintain an estimate as to his/her current location as a guide for deciding the next turn, but in the absence of incontrovertible a priori information, this estimate is best represented by the subject's belief. As information is acquired through observation, this belief may become increasingly convincing or alternatively may be discarded in favor of a new one. This is an intuitive way of making estimations that are appropriate for many real-world behaviors, adopted also by a wide variety of intelligent machines, and can be expressed simply in the following diagram:

$$\begin{aligned} \text{(a posteriori information)} \\ = \text{(a priori information)} \\ + \text{(information from observation)}, \end{aligned}$$

which describes a basic form of incremental information processing.

Here, we investigate how different regions of the brain may be responsible for implementing the simple incremental process in the estimation of uncertain states. Recent theoretical studies have suggested that neural processing in sensorimotor (Kording and Wolpert, 2004) and visual (Kersten and Yuille, 2003) systems involves supplementing uncertain observation with a priori belief, and we suggest here that this also holds for higher cognitive systems. Furthermore, recent neuroimaging studies have identified activation in the anterior prefrontal cortex (APF) when subjects perform tasks with competitive rules (Braver and Bongiolatti, 2002; Koechlin et al., 2000; Strange et al., 2001). In such tasks, the correct rule is not explicitly provided, and subjects need to maintain multiple candidate rules and evaluate them using feedback from the currently executed rule; hence, the correct rule as an incomplete (hidden) state has to be estimated using observable stimuli and feedback. These studies led to the hypothesis that APF activation may be involved in hidden-state estimation, which requires the resolution of uncertainty associated with partially observable states, based on the available observation.

To test this hypothesis, we studied brain activity during a navigation task in which subjects were required to make sequential decisions based on incomplete information to reach a goal. In brief, subjects performed repeated goal-search or control visuomotor trials, in separate blocks (Figure 1A), in a "wire-frame" maze whose topography had been robustly learned beforehand. Navigation in the goal-search task depended on observations of only the circum-spaces surrounding their current position; this constitutes a partially observable environment, because these observations are themselves frequently ambiguous (Figure 1B). During testing, subjects had to navigate to a specified goal position from the unknown starting position. To achieve this task efficiently requires estimation of the hidden current position using the restricted observations. Here lies a major methodological challenge when interpreting the corresponding brain activity, in the form of an inverse problem: subjective beliefs held by subjects are themselves hidden from the observer and therefore cannot be established unequivocally. We address this issue using a computational technique based on probabilistic model inference: we statistically estimated the individual subjects' ongoing state estimate (belief) and its associated uncertainty, based on their behavioral decision-making sequence, and then used this model to generate regression functions to predict brain (BOLD) activities.
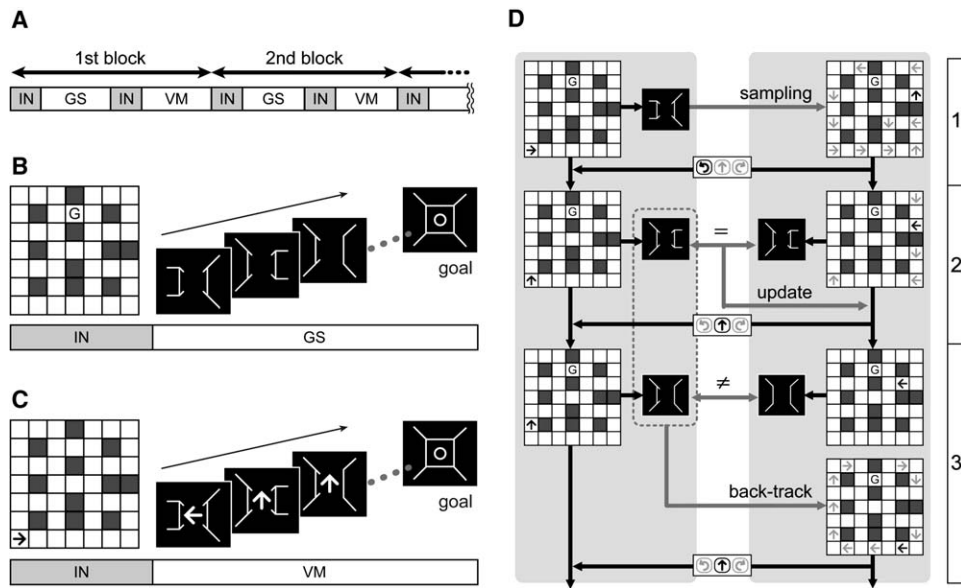
*Correspondence: ishii@is.naist.jp

Figure 1. Experimental Design

(A) Block design for fMRI imaging. IN, instruction period (4 s); GS, goal-search task; VM, visuomotor task. Every subject underwent two successive 10 min sessions separated by a 5 min interval. The length of each task block was varied with subjects' performance of the GS task (average, 1 min 38 s).

(B) An example of stimulus sequence in a GS task. During the IN period, a goal position, "G," was shown on the 2D maze map consisting of a white path and black wall grids. In each trial (2 s) in a GS period, a 3D scene was represented as wire frames of the wall status of six grids; current, left, right, forward, forward-left, and forward-right surrounding the subject's current position, and renewed dependent on the subject's action. The 3D scenes in this figure exhibit states that allow a forward or left move (first scene), a forward or left move (second scene), and a forward-and-then-right move (third scene). Goal achievement was indicated by a white circle, terminating the GS task.

(C) In a VM task, the stimulus sequence corresponded to the GS task in (B). During the IN period, an arrow indicating the start position and body orientation was shown. For each 3D scene, an up, left, or right arrow representing the action (up, forward; left, left turn; right, right turn) performed during the preceding GS task was presented; the subjects were required to press the corresponding action button.

(D) An example hidden current position estimation processes hypothesized in the model, performed along the sequences of the 3D scenes shown in (B). The left side exhibits the time series of the true positions (black arrow) and their 3D scenes, while the right side shows the position estimation process that are estimated to occur, according to our model, in the subject's brain, consisting of position candidates (gray arrows), position estimate (black arrow), and the predicted scenes. The black arrow on the central action block shows an action taken by the subject. For the initial 3D scene, the subject enumerated possible position candidates, chose and maintained one as an estimate, selected the left turn action, and then temporarily altered the estimate based on the selected action. Next, he/she validated the altered estimate by matching the predicted and actual 3D scenes. In a matched trial, as seen in the second line, the subject confirmed the estimate (update). In a mismatched trial, as diagramed in the third line, the subject enumerated position candidates again to satisfy the current scene, the previous scene, and the previous action (cf. one-step back-track), then selected any one of them as a new estimate and took a forward action.

Results

Estimation of Subject's Hidden Current Position

We first constructed a hidden Markov model (HMM) (Macdonald and Zucchini, 1997) of each subject's cognitive state. We inferred two cognitive states: the first was a belief about where the subject was in the maze (estimate states), and the second was a cognitive or operant set (operant states). We used a Bayesian belief update procedure to compute the probability distribution over these states for each trial, given the subject's prior performances. We then used these inferred cognitive states to identify regionally specific neurophysiological correlates in the brain using a conventional voxel-based analysis. Full details of the HMM and Bayesian inversion are found in the Supplemental Data.

In the goal-search task, to select the optimal actions for goal achievement, it is essential to estimate the hidden current position, based on the memorized map and the history of three-dimensional (3D) scenes (Figure 1D). In addition, there were two cognitive operant states, corresponding to a "proceed or update mode," and a

"re-evaluate or back-track mode" (Figure 1D), which could be determined by whether a predicted 3D scene was the same as an actual one or not. In brief, the HMM comprised a series of estimate states, each of which corresponds to a possible position in the maze and could be switched with back-track operations (Figure 2A). The conditional probability over these states was computed using incremental Bayes, which enabled us to identify the most likely cognitive state at each trial and the uncertainty about various estimate states, which was quantified in terms of entropy. There were multiple paths whose probabilities were different, in particular for a poor profile following a tortuous path; for example, on the middle and the right maps of Figure 2A, there are two reproduced paths for one actual path. We used the probability distribution over the estimate states to compute the conditional uncertainty about where the subject thought they were on each trial, and we used the probability distribution over the operant states to assess how likely the subject was to be in back-track mode. These probabilistic measures, called hidden current position (HCP) entropy (Figure 2B) and back-track
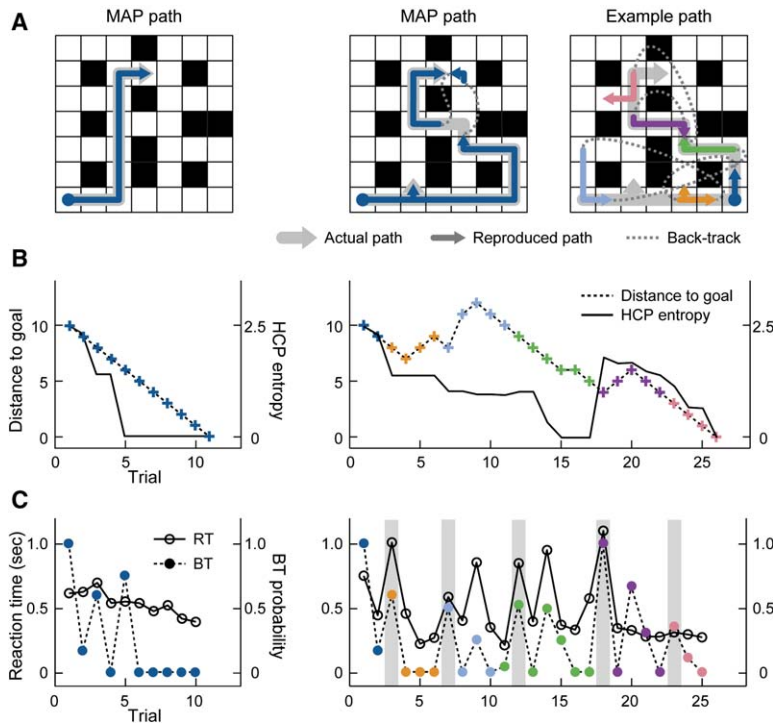
Figure 2. Behavior and Performance Profiles for a Good (Left) and a Poor (Right) Subject during a Specific Goal-Search Task, with HMM Estimation Results Based on the Profiles

(A) The actual paths selected by the subjects (gray arrows) and example paths corresponding to where the model predicts that the subjects think they are, reproduced by the HMM (blue and colored arrows). The left panel shows the actual and most probable (maximum a posteriori [MAP]) path for the good subject. The middle and right panels, corresponding to the poor subject, show the actual plus two reproduced example paths, and each panel corresponds to the MAP (middle) and non-MAP example (right). The dotted lines represent the back-track operation, in which the subjects re-estimate their position to be somewhere else in the maze. In the less probable example on the right panel (multiple colored arrows), for example, the subject is estimated to have performed numerous back-track operations. Thus, the reproduced paths sometimes jump to physically inconsistent grids (dotted lines), showing that the estimate state was switched following back-track operations that re-estimated the hidden current position. Note, for the poor subject in the middle and right panel examples, the short vertical arrows on the third from left and bottom grid indicate that this subject selected a left-turn action in the fourth trial but selected a right-turn action next, returning to the previous position in the fifth trial. The reproduced paths start on the grid with the blue closed circle, but may end on a grid other than the correct goal position (as in the right panel) because the subject would recognize the goal achievement after the path termination. The colors in the right panel correspond to those in the right panels of (B) and (C).

(B) The minimum numbers of steps to the goal (distance to goal: shown as dotted black lines with colored cross marks) and HCP entropies (black lines) are plotted against the number of trials. Plateaus in the distance to goal (e.g., the 15th and 16th trials of poor performance) exhibit missed trials, in which the subject could not press an action button within the given time (1.8 s). HCP entropy generally decreased with the resolution of position uncertainty, although it occasionally increased rapidly with an increase in the number of position candidates due to back-track operations (e.g., the 18th trial of poor performance).

(C) Reaction times (RTs: lines with open circles) and BT probabilities (dotted lines with colored circles) are plotted. The shaded trials in the right panel correspond to the trials when back-track operation was estimated to occur in the reproduced path described in (A).

(BT) probability (Figure 2C), respectively, were then used to parametrically modulate event-related stimulus functions for each trial. After convolution with a hemodynamic response function, the stimulus functions were entered into a general linear convolution model of evoked hemodynamic responses in the usual way to form statistical parametric maps.

Behavioral Results
A paired t test analyzing the average reaction times (RTs) demonstrated that the RTs in the goal-search tasks (533 ms) were significantly longer than those measured in the visuomotor control tasks (447 ms) (t = 5.20, p < 0.0001). We defined early and late phases as the former and latter halves of trials in each task block and found that the RTs in the early phase of goal-search tasks (594 ms) were significantly longer than those measured in the early visuomotor phase (467 ms) (t = 3.75, p < 0.001). There was no statistically significant difference, however, between the late goal-search (483 ms) and visuomotor phases (438 ms) (t = 1.67, p = 0.101). Within the goal-search tasks, the RTs at branch (T intersection) states (701 ms), where action selection was explicitly required, were significantly longer than those at other (straight and curve) states (556 ms) in the early goal-search phase (t = 3.10, p < 0.005); in contrast, there

was no significant difference between these states in the late goal-search phase (branch = 546 ms, other = 488 ms, t = 1.44, p = 0.156). These results provide behavioral evidence of the importance of position estimation, over and above that related to action selection alone.

To establish the face validity of our HMM and the Bayesian inversion, we examined the RTs in detail to see whether they were consistent with the cognitive states inferred by the model. We were particularly interested to ascertain whether being in a back-track mode prolonged RTs and whether increasing conditional uncertainty (measured by HCP entropy) had a similar effect. The RT profiles (Figure 2C) demonstrated a low-frequency trend of a gradual RT decrease with the progression of trials. There is also a high-frequency component with a relatively jagged pattern; these two components are likely related to the manipulation of position candidates. Interestingly, the RT peaks (high-frequency component) correlated temporally with the BT probability peaks estimated by the HMM (Figure 2C). The peaks of BT probability occurred as estimated in 988 trials (32.5% of total of 3038 trials). Surprisingly, the majority of these trials (584 trials, 59.1%) coincided exactly with the peak timings of highly scattered RTs; this coincidence was significant ($F_0 = 2.99$, $p < 10^{-5}$), based on a goodness-of-fit test for binominal distributions.

Cortex

Basal ganglia

(1) caudate

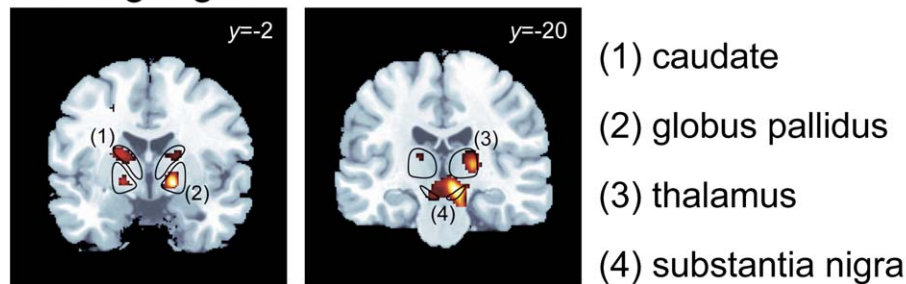(2) globus pallidus

(3) thalamus
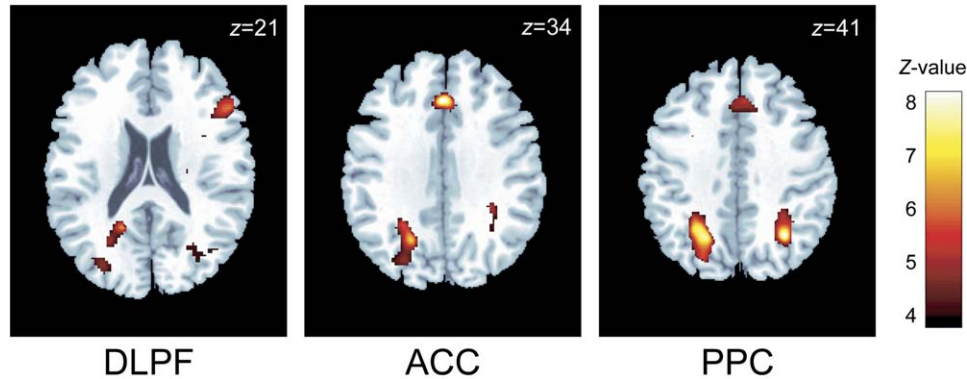
(4) substantia nigra

Figure 3. Brain Areas that Are More Active in the goal-Search Tasks than in the Visuomotor Tasks: Regions of the Cerebral Cortex, Including the DLPF, ACC, and PPC; Areas of Basal Ganglia, Including the Caudate, Globus Pallidus, and Substantia Nigra; and the Thalamus

The inclusion of adjacent trials of BT peaks, which allowed for small differences in timing attributable to subjects' probabilistic behaviors, increased the rate to 86.1% (851 trials). To support this robust synchronism, we examined for each subject the correlation coefficients between the RT sequence and the binary occurrence sequences of BT peaks and branch states in which the RT was longer. The correlations of RT with BT peaks were significant for all 13 subjects ($p < 10^{-5}$ for 12 subjects, $p < 0.05$ for one subject), while with the branch states, significant positive correlations were found only for four subjects but negative for the other four subjects, and this result indicated that the principal factor responsible for the RT peaks was likely to be the back-track operations rather than the branch states. By correlation analysis, smoothed RT (low-frequency component of RT), in which the peaks responsible for back-track operations were removed, correlated positively with the HCP entropy for all 13 subjects; the correlations were significantly positive ($p < 0.05$) for ten of them. In contrast, the distance (the minimum number of steps) to the goal (Figure 2B) demonstrated significant positive correlations with the smoothed RT for only five subjects ($p < 0.05$), and negative correlations for two subjects. We compared the correlation coefficients over 13 subjects, confirming that the average correlation of the RT with the HCP entropy ($r = 0.23 \pm 0.16$) was significantly higher than that with the distance to the goal ($r = 0.17 \pm 0.16$) ($t = 2.39$, $p < 0.05$), even though the distance was obtained directly from the subjects' profiles. The precision of our HMM-based approach was confirmed by the high accordance between the subjects' actions and those

predicted by our model ($88.7\% \pm 2.9\%$). The model's prediction accuracy correlated significantly with the subjects' task performance ($r = 0.49$, $p < 10^{-5}$); better performance blocks, in which the subject selected nearly optimal paths (e.g., Figure 2A), were accompanied by a higher prediction accuracy. Note that we are treating the RT as a dependent or response variable, in the same manner with which we treated the hemodynamic response. Both can be explained or caused by changes in the cognitive state.

Imaging Results

A subtraction analysis comparing the goal-search and visuomotor task blocks revealed significant activation during goal-search tasks in regions of cerebral cortex, including the right dorsolateral cortex (DLPF; BA46), anterior cingulate cortex (ACC; BA32/8), and bilateral posterior parietal cortex (PPC; BA40/7), the thalamus, and basal ganglia areas (putamen, caudate, globus pallidus, and substantia nigra) (Figure 3). The statistics can be found in Table S1.

To identify neural activity involved in the estimation of the hidden current position, we subsequently conducted a regression analysis using the two regression functions estimated by our HMM. Both the medial PFC (mPFC; BA9/6) on the medial frontal gyrus (Figure 4A) and the bilateral PPC correlated with the BT probability (Table S2). According to a region of interest (ROI) analysis, activity in mPFC was significantly higher during goal-search than during visuomotor tasks ($t = 3.59$, $p < 0.005$), as well as in trials associated with back-track mode compared to update mode within goal-search
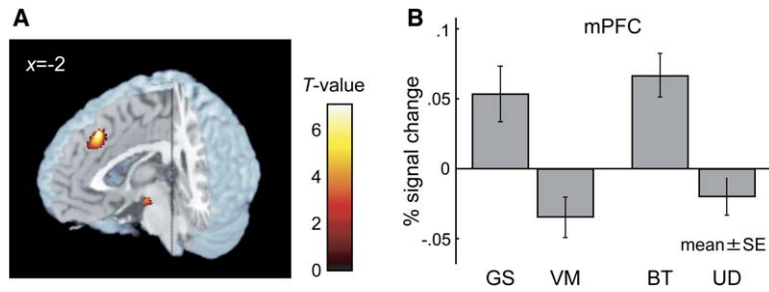
Figure 4. Brain Activation Involved in the Processes Performed during Back-Track Mode Trials

(A) Medial PFC (mPFC) activity was significantly correlated with the BT probability.

(B) Relative percent signal change of BOLD responses in mPFC (BA9/6) during goal-search (GS) and visuomotor (VM) tasks (left bars) and in back-track (BT) and update (UD) mode trials within GS tasks (right bars). The error bars indicate the standard error.

tasks (t = 4.20, p < 10^{-4}) (Figure 4B). The mean correlation coefficient between the mPFC activity and the BT probability over all goal-search blocks (r = 0.12 ± 0.09) was significantly higher than that between the mPFC activity and the RT (r = −0.03 ± 0.04) (t = 5.22, p < 0.001), indicating that our estimation method based on HMM filtering is better able to explain the imaging data than the performance data itself. For the HCP entropy, we observed correlated activity only in the anterior prefrontal cortex (APF), specifically in the bilateral medial frontal gyrus (BA10) and superior frontal gyrus (BA9) (Figure 5A and Table S2). In an ROI analysis, the percent signal change in both right and left medial frontal gyrus (BA10) positively correlated with the HCP entropy for all 13 subjects (left, r = 0.29 ± 0.22; right, r = 0.27 ± 0.19), and for ten of them it was significantly positive (p < 0.05), according to a strict significance test using the relatively small number of samples from each subject (Figure 5B and 5C). Regardless of the subjects' task performance, this positive correlation supported our

hypothesis that the APF is activated during the maintenance of partial observability (uncertainty) in accordance with the HCP entropy, which evaluates the essential computational load for the hidden position estimation. We also created a model that included the distance to the goal (which is unobservable by the subject) and the RT as statistical regression functions. These did not identify any significant correlations, excepting a small area correlated with the DG in the left APF (x, y, z = −32, 49, −1; Z = 3.87; k = 285); the correlated area was within but substantially smaller than that associated with the HCP entropy.

Discussion

Neural Activity Related to Resolution of Uncertainty
Our results provide evidence that activity in different regions of the prefrontal cortex reflect critical computational components involved in decision making in uncertain environments. This fits well with the proposed role
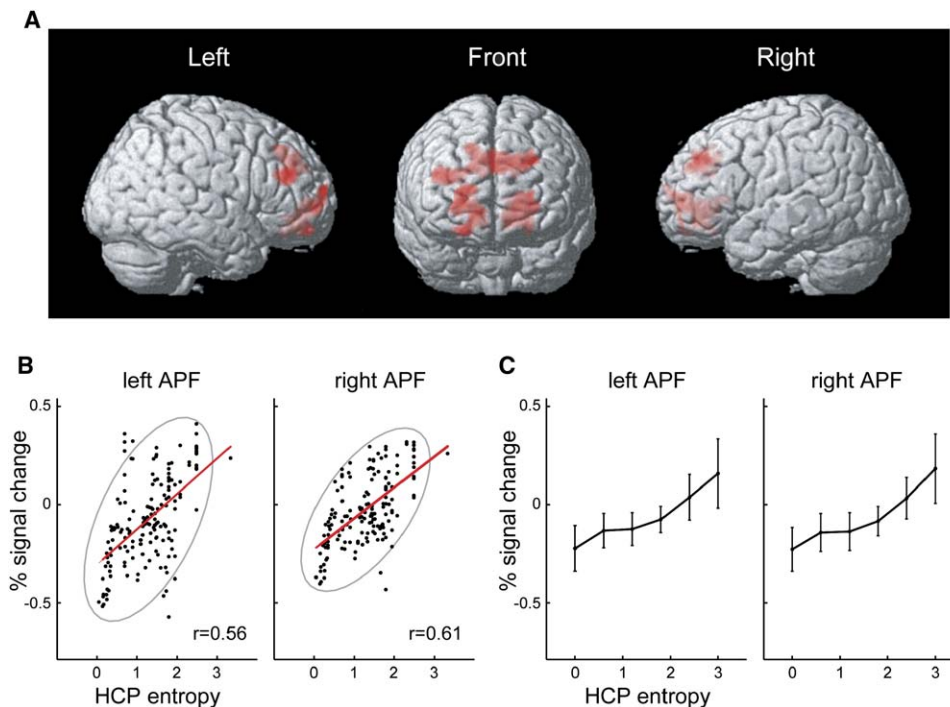

Figure 5. Brain Activation Involved in the Processes Corresponding to Maintenance and Updating of the Estimate State

(A) Bilateral APF activity was associated with the HCP entropy.

(B and C) The HCP entropy was significantly correlated with BOLD signal change for ROI in the left and right APF (BA10). The percent signal change of these areas relative to the mean activity during the GS task for a typical subject (B), and the moving average for all subjects (C). The error bars indicate the standard deviations.

of these regions in decision making, which is likely to be crucial in complex real-world environments. We also illustrate the utility of statistical model-based inference and regression in delineating key task parameters that may be represented in spatially distinct brain regions.

From a general neuroanatomical perspective, our block-based analysis of goal-search versus visuomotor tasks demonstrates brain areas that are known to play important roles in navigation and goal search. The PPC is thought to be involved in the maintenance of egocentric spatial representation (Colby and Goldberg, 1999), particularly in 3D space (Tsutsui et al., 2002), and may be related to the maintenance of 3D maze topography in our task. The DLPF is thought to be responsible for higher-order functions such as judgment and prediction (Petrides, 1996), and so would be expected to be involved in a prediction-based optimal action selection task as utilized here. The ACC is robustly associated with the detection of behavioral errors (Braver et al., 2001) and response conflict (Botvinick et al., 1999), and the activation during the goal-search task that we observed may relate to the response conflict induced by multiple action candidates for each 3D scene and/or by the lack of clarity concerning the optimal action due to the ambiguity of the current position. We also observed activation of both the thalamus and the basal ganglia nuclei during the goal-search task; these areas are known to constitute parallel loops with the cortex that are important for the control of involuntary motion (Alexander et al., 1986; Middleton and Strick, 2000) and that may be involved in motor-information processing based on the goal (reward), which is required for maze exploration.

Our model-based analysis data provide evidence that more precise computational functions may be engaged in prefrontal cortex. The mPFC is known to be coactivated with self-referential stimuli (Northoff and Bermpohl, 2004), and recent recording (Matsumoto et al., 2003) and lesion (Gehring and Knight, 2000) studies have demonstrated evidence of a link between the mPFC and monitoring of one's individual performance (self-generated actions) based on reward prediction, which engages the lateral PFC to execute performance adjustments. The processes performed during back-track mode, such as error detection and consequent action selection, corresponded closely to these hypothesized functions. However, unlike previous tasks in which the subjects select actions based purely on a given stimulus, the action in our goal-search tasks is selected based on both the observed scenes (stimuli) *and* the current estimate of the position. Thus, our results suggest that the mPFC is responsible for selecting actions based on both the currently observable and an internally evaluated cognitive state. The correlation of bilateral PPC with BT probability is more speculative, but might represent memory reload to enumerate new position candidates.

In this study, we observed activation in medial cortical structures by both goal-search > visuomotor subtraction analysis and regression analysis using the BT probability, which were primarily distributed over the ACC (BA8/32) and the mPFC (BA9), respectively. These results indicate that the ACC was constantly activated during goal-search tasks, while mPFC activity increased in relation to back-track operations. Recently, medial cortical structures have been shown to become activated when action changes are induced by either negative stimuli (error) or positive ones (reward) (Knutson et al., 2001; Shidara and Richmond, 2002). A recent review suggests functional segregation in this area, distinguishing three regions: the orbitomedial PFC (BA10/11/12), dorsomedial PFC (BA8/9, mPFC in our study), and the anterior cingulate (BA24/32), and postulated that these regions correspond to representation, evaluation, and monitoring of self-referential stimuli, respectively (Northoff and Bermpohl, 2004). During each trial of the goal-search task, the subjects were required to check the 3D scenes that resulted from their actions; subjects were to re-estimate their positions only during back-track trials. These processes can be regarded as the monitoring and evaluation of self-referential stimuli; our results indicating that ACC and mPFC activity, respectively, are consistent with the recent view presented above.

Recent imaging studies have demonstrated APF activation during complex cognitive tasks, in particular those involving multiple competitive rules (Braver and Bongiolatti, 2002; Koechlin et al., 2000; Strange et al., 2001). The tasks used in these studies can be regarded as optimal decision problems in partially observable environments: given that the correct rule was not explicitly provided by the environment, the subjects were required to maintain multiple candidates, such as two subtasks in a branching task (Koechlin et al., 2000) or six rules in a categorization task (Strange et al., 2001), and then to examine their possibilities based on feedback. These processes correspond to the maintenance and updating of a hidden state. In these tasks, however, the number of hidden rules was small enough to allow the subjects to memorize them completely. In addition, as feedback clearly indicated whether the used rule had been correct, updating the candidates at each trial did not require the estimation of probabilities for each candidate, but only deletion or retention of the used rule based on feedback. As these studies performed a subtraction analysis between the conditions or events, the results did not fully illuminate the processes functioning during the tasks. Our regression analysis has clarified that APF activities vary according to the HCP entropy, which is the objectively estimated measure of the internal conflict among candidates of the hidden state, adding a dynamic aspect to previous results. Our result is also consistent with recent assertions that the APF is involved in the processing of internally generated information (Christoff et al., 2003) and is pivotal in the PFC network by integrating information represented in lateral PFC to select an appropriate behavioral rule (Koechlin et al., 2003).

Computational Considerations

From a computational viewpoint, an optimal action selection problem without hidden states is formulated as a Markov decision process (MDP), and reinforcement learning, which is an approach to solving MDPs, may be realized within brain networks primarily including the PFC and the basal ganglia (Daw et al., 2005; O'Doherty et al., 2004; and Seymour et al., 2004). If the environment involves unobservable information such as hidden states, such a problem is called a partially observable MDP (POMDP) (Astrom, 1965). In POMDPs, the hidden-state estimate is called a "belief" and is often

represented as a probability distribution over the state space (Kaelbling et al., 1998), and we have previously suggested that these beliefs are calculated using Bayesian inference (Ishii et al., 2002). In the current study, we assumed subjects' process information as an HMM and calculated the belief by incremental Bayes, in which the posterior probability of their estimate state is updated based on the previous one as a priori information, after each scene observation. By introducing Bayesian filtering based on an HMM, we were able to dissociate the high- and low-frequency components of noisy RTs successfully and to explain the subjects' behavioral profiles, providing strong support for the hypothesis that subjects perform the partially observable maze task by executing incremental inference (filtering) to estimate hidden states. Using the entropy about the current estimate as an explanatory variable in our model presupposes that the brain is representing, explicitly, conditional uncertainty about the current estimate. In fact, one could interpret our model in terms of a neuronal implementation of the Bayesian update scheme, under the assumption that each subject was an ideal Bayesian observer. This approach has been used in a useful way in the context of perceptual learning and uncertainty or unpredictability (Strange et al., 2005).

Although several recent studies using probabilistic models have been successful in reproducing internal information processing (Smith et al., 2004; Tanaka et al., 2004), the regression functions have often been calculated based on the maximum likelihood or maximum a posteriori inference for hidden variables, and statisticians have recognized that such a point estimate may introduce instability in the analysis, especially when there are a relatively large number of hidden variables, as in our task. Even though our HCP entropy is a subjective value, because it was obtained by mimicking the subjects' process, this value depends on subjects' internal state. In such a case, a Bayesian approach employing the marginalization technique (MacKay, 2003) eliminates the instability arising from the point estimate and hence provides for a reliable regression analysis. Although this statistical technique is novel, it is potentially a powerful technique for use in imaging studies, and we believe that such approaches will become increasingly important for revealing the principles underlying higher brain functions.

As we acquire new information from experience and observation, our beliefs may become either more convincing or be discarded in favor of another. The maintenance and refutation of beliefs represent critical processes involved in sequential decision making, realizable with the simple incremental processes discussed above. Our imaging results suggest that the former, belief maintenance, and the latter, belief back-track, are performed principally by the APF and the mPFC, respectively, implying that a large-scale circuit in the PFC is responsible for optimal decision making during uncertainty.

Experimental Procedures

Experimental Details

Thirteen normal subjects (11 males and two females, aged 23–28) participated after giving written informed consent, which was approved by the ethical committee of the Advanced Telecommunications Research Institute International (ATR), Japan. Subjects performed a maze exploration task using a block design with goal-search and visuomotor control task components (Figure 1A) using an identical maze. The maze had a 7 × 7 square grid of walls and paths with no dead-ends or crossroads. The topography was intentionally designed to exhibit relatively high symmetry, so that the resolution of any position was the best way to navigate efficiently (as the subjects actually did). This organization largely precluded any efficient use of more heuristic strategies, like a phased estimation, in which a subject might first specify the broad spatial area and then estimate their position within it. On the day before scanning, subjects were given verbal and written explanations of the aim and procedures of the behavioral tasks and practiced a training task outside the MRI scanner to learn the topography of the maze. We confirmed that the subjects had memorized the whole map by asking them to recall possible two-dimensional (2D) positions from exemplar 3D scenes.

In the goal-search task (Figure 1B), after the goal position on the learned 2D map was presented on a screen (4 s), a 3D scene at an unknown start position and body orientation was displayed, which constituted the hidden current position. Subjects were then required to press one of three action buttons: forward movement, which made the subjects move to the forward position (grid); and left or right turn, which made them turn in the corresponding direction while staying in the same position, within 1.8 s. The intertrial interval was fixed at 2 s; the next trial began with presentation of the 3D scene at the next position, determined by the previously selected action. If the action was not performed within the allotted time or progress was blocked by a wall, the position did not change; the same 3D scene was displayed. Immediately after goal achievement, a yoked visuomotor task (Figure 1C) was performed as a control task in which subjects were required to reproduce all sensorimotor events in the preceding goal-search task, guided by visual instructions. Every subject underwent two successive 10 min sessions separated by a 5 min interval, and the order of the two sessions was counterbalanced across subjects. Each session comprised a fixed number of 300 trials and could be terminated even in the middle of either type of task; only the data for completed blocks were used for analyses. While the combinations of start and goal positions were different for each block and session, the minimum numbers of steps and branches (T intersections) were approximately the same over all blocks (12 ± 2 steps and 4 ± 1 branches). The subjects' performances in the first and second sessions during goal-search tasks did not differ significantly; the ratio of the actual number to the minimum number of trials from the start to goal for each block (first session: 2.87 ± 0.95, second session: 2.51 ± 0.62; t = 1.10, p = 0.29) and the mean reaction time (RT) (first: 535 ± 134 ms, second: 527 ± 146 ms; t = 0.44, p = 0.67) showed no learning effect on the goal-search tasks. Our subsequent analyses did not distinguish the data from each of the two sessions.

Bayesian Estimation for Subjects' Processing

The aim of our hidden Markov model was to reproduce all possible sequences of the subject's cognitive state, estimate state, and operant state from the sequence of both the presented 3D scene and the subject's action. As both the memory capacity and information processing ability of humans are limited, it is difficult to perform optimally in complex tasks, such as our maze navigation, based on all information given from the environment. Thus, subjects' behaviors are frequently imperfect (Figure 2). To explain such behaviors, it is necessary to introduce a kind of imperfectness into the model of the subjects' behavior. Based on retrospective inspections from the subjects, we assumed an HMM-based model of subject information processing, hypothesizing both an imperfect sampling process (concurrent maintenance and update of multiple position candidates) and a re-estimation process to compensate for this error (details can be found in the Supplemental Data). In brief (Figure 1D), in each trial of the goal-search task, subjects would maintain current position candidates that are consistent with the current and previously observed 3D scenes, estimate their position as one of these possibilities, and select an action based on that estimation. Using the selected action, the next position and the corresponding scene can be predicted; after the next 3D scene becomes accessible,

subjects can assess whether their prediction is correct by matching the predicted scene with the actual scene. If the prediction is correct, the predicted position estimate becomes more convincing, and the position candidates are maintained such that they are consistent with the new scene (update operation). If there is a discrepancy between the scenes, however, the position candidates are re-enumerated, based on the history of scenes and actions from *n*-step past trials, and a new estimate-state is selected from these possibilities (back-track operation). We compared our model to three alternative models with perfect-or-imperfect temporal-or-spatial memory. Our model surpassed alternative models in both the goodness-of-model fit and action prediction accuracy (see Supplemental Data).

We then compute the probability distribution over reproduced cognitive states by our HMM using a Bayesian belief update procedure. According to the Bayes theorem, the posterior probability of a hidden variable *X* is given by

$$P(X \mid Y) = \frac{P(Y \mid X)P(X)}{P(Y)},$$

where $P(X)$ is the prior probability before observing the data *Y*, and $P(Y|X)$ is the likelihood of perceiving *Y* when the hidden variable is *X*. $P(Y)$ is the likelihood of the data, often called evidence (MacKay, 2003), and can be used for evaluating the data reproducibility by the model. When the hidden variable and observation (data) are constituted by time series, $X = \{x_1, x_2, ...\}$ and $Y = \{y_1, y_2, ...\}$, respectively, and the state transition process and observation process are given by $P(x_t|x_{t-1})$ and $P(y_t|x_t)$, respectively, the posterior probability is calculated by a simple recursive form:

$$P(x_{1:t+1} \mid y_{1:t}) = \frac{P(x_{t+1} \mid x_t)P(y_t \mid x_t)P(x_{1:t} \mid y_{1:t-1})}{P(y_t \mid y_{1:t-1})},$$

where the suffix 1:*t* denotes a data vector from time step 1 to *t*, i.e., $y_{1:t} \equiv \{y_1, y_2, ..., y_t\}$. This recursive inference is the simplest case of the general framework called incremental Bayes or Bayesian filtering. Note that the Bayes theorem provides a statistically sound basis for the simple diagram described in the Introduction. For our maze task, we intend to reproduce the sequences of the subjects' cognitive states, which are hidden variables of the HMM, from the sequence of subjects' actions and observations. Here, the likelihood $P(Y|X)$ was calculated based on the processes in the assumed subject's model (see Supplemental Data). The two HMM parameters were determined by the maximum-likelihood estimation; the number of past scenes used for the back-track operation (*n*) and the optimal action selection probability (α) were obtained as $\alpha = 0.8$ and $n = 1$ so as to maximize the product of evidence $P(Y)$ over all subjects. Even if we calculate them for each subject, they did not vary much over the 13 subjects ($\alpha = 0.81 \pm 0.08$, $n = 1.3 \pm 0.6$). In addition, the product of evidence was not significantly dependent on these parameter values, showing that our HMM has stability over these parameters. The action reproduction accuracy of our model was calculated as the proportion of successful action prediction over trials, averaged with respect to the posterior probabilities.

To identify brain areas functionally involved in essential information processes during hidden position estimation, we calculated two regression functions. One was the hidden current position (HCP) entropy, which evaluates the integrative computational load of assessing the variable estimate states. This load, related to the resolution of uncertainty, can be calculated as the amount of information (Shannon entropy) for the possible estimate states, but was dependent on the sequence of cognitive states in the subject's brain. To remove this dependence, the HCP entropy was calculated by averaging the uncertainty over possible sequences of cognitive states with respect to their posterior probabilities. This technique, called "marginalization," can remove the instability stemming from the hidden variable estimation. When the subjects enumerated position candidates either initially or in back-track trials, additional processing was required. To evaluate this additional computational load inherent to the re-estimation, the BT probability was also calculated as a "marginalized" back-track occurrence, objectively from the subjects' behaviors. Here, we defined BT peak trials as the goal-search trials in which our model estimated a BT probability exceeding 25%. RT peaks were then extracted in descending order of

the difference from the mean RT of the adjacent trials to provide the same number of BT peak trials within the same goal-search block. A smoothed RT was obtained by replacing the RT at each BT peak trial with the average RT of adjacent trials.

Imaging Data and Analysis

Functional images were obtained with T2-weighted EPIs using BOLD contrast (TE, 55 ms; FA, 90°) on a 1.5 tesla scanner (Magnetic Eclipse; Shimadzu Marconi, Kyoto, Japan). Volumes, acquired in synchronization with stimulus presentation (TR, 2 s), contained 20 slices each of a 5 mm thickness (matrix size, 64 × 64; FOV, 192 × 192 mm). The first six (12 s) EPIs in each session were not evaluated as a part of the scanning data to avoid T1 equilibrium effects. Each scanning run began with a T1-weighted anatomical image acquisition (voxel size, 1 mm^3).

Imaging data were analyzed using SPM99 (Wellcome Department of Cognitive Neurology, London, UK). All functional images from each subject were realigned to the first image as a reference, coregistered to the individual anatomical image, normalized into an MNI template, and spatially smoothed with a Gaussian kernel filter (FWHM, 10 mm). For goal-search and visuomotor tasks, sustained activity was modeled as an epoch using a boxcar function covering the whole experiment. Both subtraction analysis comparing the two tasks and multivariate regression analysis utilizing the HCP entropy and BT probability were conducted on the whole brain. These regression functions were orthogonalized by the Gram-Schmidt method to remove correlations: the HCP entropy was orthogonalized to the goal-search boxcar function first. All epochs and regression functions were convolved with a canonical hemodynamic response function. A group random effects statistic was calculated from the combination of individual contrast maps. We applied statistical thresholds at the voxel level of p < 0.001 (uncorrected) and at the cluster level of p < 0.05 (corrected). To see how the activation within a particular brain area correlated with our regression functions, we employed a region-of-interest analysis defining regions using activations from previous SPM analyses and an anatomically defined area. The percent signal change was the relative change from the mean MR signal seen throughout the experiment (left bars in Figure 4B) or during the goal-search task alone (right bars in Figures 4B, 5B, and 5C), excluding scans performed during the instruction phase.

Supplemental Data

The Supplemental Data for this article can be found online at http://www.neuron.org/cgi/content/full/50/5/781/DC1/.

References

Alexander, G.E., Delong, M.R., and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. *9*, 357–381.

Astrom, K. (1965). Optimal control of Markov processes with incomplete state information. J. Math. Anal. Appl. *10*, 174–205.

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

Braver, T.S., and Bongiolatti, S.R. (2002). The role of frontopolar cortex in subgoal processing during working memory. Neuroimage *15*, 523–536.

Braver, T.S., Barch, D.M., Gray, J.R., Molfese, D.L., and Snyder, A.Z. (2001). Anterior cingulate and response conflict: Effects of frequency, inhibition, and errors. Cereb. Cortex *11*, 825–836.

Cassandra, A.R., Kaelbling, L.P., and Littman, M.L. (1994). Acting optimally in partially observable stochastic domains. In Proceedings of the Twelfth National Conference on Artificial Intelligence, Volume 2 (Seattle, WA: AAAI Press/MIT Press), pp. 1023–1028.

Christoff, K., Ream, J.M., Geddes, L.P., and Gabrieli, J.D. (2003). Evaluating self-generated information: anterior prefrontal contributions to human cognition. Behav. Neurosci. *117*, 1161–1168.

Colby, C.L., and Goldberg, M.E. (1999). Space and attention in parietal cortex. Annu. Rev. Neurosci. *22*, 319–349.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci. *8*, 1704–1711.

Gehring, W.J., and Knight, R.T. (2000). Prefrontal-cingulate interactions in action monitoring. Nat. Neurosci. *3*, 516–520.

Ishii, S., Yoshida, W., and Yoshimoto, J. (2002). Control of exploitation-exploration meta-parameter in reinforcement learning. Neural Netw. *15*, 665–687.

Kaelbling, L.P., Littman, M.L., and Cassandra, A.R. (1998). Planning and acting in partially observable stochastic domains. Artif. Intell. *101*, 99–134.

Kersten, D., and Yuille, A. (2003). Bayesian models of object perception. Curr. Opin. Neurobiol. *13*, 150–158.

Koechlin, E., Ody, C., and Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. Science *302*, 1181–1185.

Koechlin, E., Corrado, G., Pietrini, P., and Grafman, J. (2000). Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. Proc. Natl. Acad. Sci. USA *97*, 7651–7656.

Kording, K.P., and Wolpert, D. (2004). Bayesian integration in sensorimotor learning. Nature *427*, 244–247.

Knutson, B., Fong, G.W., Adams, C.M., Varner, J.L., and Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. Neuroreport *12*, 3683–3687.

Macdonald, I.L., and Zucchini, W. (1997). Hidden Markov and Other Models for Discrete-Valued Time Series (London: Chapman and Hall/CRC Press).

MacKay, D. (2003). Information Theory, Inference and Learning Algorithms (New York: Cambridge University Press).

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

Middleton, F.A., and Strick, P.L. (2000). Basal ganglia output and cognition: Evidence from anatomical, behavioral, and clinical studies. Brain Cogn. *42*, 183–200.

Northoff, G., and Bermpohl, F. (2004). Cortical midline structures and the self. Trends Cogn. Sci. *8*, 102–107.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science *304*, 452–454.

Petrides, M. (1996). Specialized systems for the processing of mnemonic information within the primate frontal cortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. *351*, 1455–1461.

Seymour, B., O'Doherty, J., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664–667.

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

Smith, A.C., Frank, L.M., Wirth, S., Yanike, M., Hu, D., Kubota, Y., Graybiel, A.M., Suzuki, W.A., and Brown, E.N. (2004). Dynamic analysis of learning in behavioral experiments. J. Neurosci. *24*, 447–461.

Stankiewicz, B.J., Legge, G.E., Mansfield, J.S., and Schlicht, E.J. (2006). Lost in virtual space: Studies in human and ideal spatial navigation. J. Exp. Psychol. Hum. Percept. Perform., in press.

Strange, B.A., Henson, R.N.A., Friston, K.J., and Dolan, R.J. (2001). Anterior prefrontal cortex mediates rule learning in humans. Cereb. Cortex *11*, 1040–1046.

Strange, B.A., Duggins, A., Penny, W., Dolan, R.J., and Friston, K.J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable? Neural Netw. *18*, 225–230.

Sugrue, L.P., Corrado, G.S., and Newsome, W.T. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. Nat. Rev. Neurosci. *6*, 363–375.

Tanaka, S.C., Doya, K., Okada, G., Ueda, K., Okamoto, Y., and Yamawaki, S. (2004). Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. Nat. Neurosci. *7*, 887–893.

Tsutsui, K., Sakata, H., Naganuma, T., and Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. Science *298*, 409–412.

Wood, J.N., and Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. Nat. Rev. Neurosci. *4*, 139–147.