Brain Hemispheres Selectively Track the Expected Value of Contralateral Options

Stefano Palminteri,^{1,3} Thomas Boraud,⁴ Gilles Lafargue,⁵ Bruno Dubois,^{1,3} and Mathias Pessiglione^{1,2,3}

¹Institut du Cerveau et de la Moëlle épinière (CR-ICM), INSERM UMR 975 and ²Centre de Neuroimagerie de Recherche (CENIR), Hôpital Pitié-Salpêtrière, F-75013 Paris, France, ³Université Pierre et Marie Curie (Paris 6), F-75005 Paris, France, ⁴Basal Gang, CNRS UMR 5227, Université Victor Segalen (Bordeaux 2), F-33076 Bordeaux, France, and ⁵CNRS UMR 8160, Université Lille Nord-de-France (Lille 3), F-59653 Villeneuve d'Ascq, France

A main focus in economics is on binary choice situations, in which human agents have to choose between two alternative options. The classical view is that decision making consists of valuating each option, comparing the two expected values, and selecting the higher one. Some neural correlates of option values have been described in animals, but little is known about how they are represented in the human brain: are they integrated into a single center or distributed over different areas? To address this issue, we examined whether the expected values of two options, which were cued by visual symbols and chosen with either the left or right hand, could be distinguished using functional magnetic resonance imaging. The two options were linked to monetary rewards through probabilistic contingencies that subjects had to learn so as to maximize payoff. Learning curves were fitted with a standard computational model that updates, on a trial-by-trial basis, the value of the chosen option in proportion to a reward prediction error. Results show that during learning, left and right option values were specifically expressed in the contralateral ventral prefrontal cortex, regardless of the upcoming choice. We therefore suggest that expected values are represented in a distributed manner that respects the topography of the brain systems elicited by the available options.

Introduction

Some patients complain that their left (L) hand does not obey them anymore. For example, the left hand may close doors the right (R) has opened, or snatch money the right has offered to a store cashier (Baynes et al., 1997). This is called the anarchic hand syndrome and is often seen after damage to the corpus callosum resulting in disconnection of the two brain hemispheres (Sperry, 1961; Gazzaniga, 2005). It suggests that the nonspeaking right hemisphere may pursue goals different from those of the speaking left hemisphere. Hence, different expected rewards may be represented in the left and right hemispheres and conveyed to motor systems controlling the contralateral side of the body. To test this idea, we designed an experimental paradigm in which we manipulated the rewards associated with left and right hand responses. Then, we used functional magnetic resonance imaging (fMRI) to examine whether the reward values of the two options were represented in the contralateral hemispheres.

In addition to the issue of interhemispheric competition, identifying representation of option values is crucial for understanding how the human brain resolves binary choice situations

that are at the core of economic theories. Indeed, when two alternatives are available, the rational economic agent is supposed to value the two options, compare the expected values, and select the higher one (Rangel et al., 2008). Some neural correlates of option values have been described in the monkey striatum and orbitofrontal cortex (Samejima et al., 2005; Padoa-Schioppa and Assad, 2006; Lau and Glimcher, 2008). Option value representation has also been identified in the human brain when there is just one alternative that the subject can accept or refuse (Knutson et al., 2007; Plassmann et al., 2007; Hare et al., 2008). In situations involving two independent alternatives, previous studies have only located representation of the chosen option value, as well as state values, calculated as a linear combination of the two option values (O'Doherty et al., 2004; Valentin et al., 2007; Gläscher et al., 2009). Here we intend to spatially dissociate simultaneous representations of the two option values by locating them in two different brain hemispheres.

Our experimental design used a probabilistic instrumental learning task adapted from previous study (Pessiglione et al., 2006), in which subjects have to select one of two visual cues so as to maximize payoff in the long run (Fig. 1A). To avoid confounds between reward- and movement-related activations, the different responses in instrumental learning tasks are traditionally given with the same hand, such that their neuronal representation cannot be distinguished with fMRI (Haruno et al., 2004; O'Doherty et al., 2004; Tanaka et al., 2004; Daw et al., 2006; Kim et al., 2006). Here, we purposely asked subjects to respond with the right hand to select cues appearing on the right side of the screen, and with the left hand for left-appearing cues. Thus, the cue, hand, and button on each side were naturally integrated into competing

Received March 30, 2009; revised July 3, 2009; accepted Aug. 12, 2009.

This study was funded by the Fyssen Foundation. S.P. received a PhD fellowship from the Neuropôle de Recherche Francilien. We are grateful to Chris D. Frith for helpful suggestions on an earlier version of the manuscript. We also thank the Centre de Neuroimagerie de Recherche staff (Eric Bardin, Eric Bertasi, Kevin Nigaud, and Romain Valabregue) for skillful assistance in MRI data acquisition and analysis. Soledad Jorge, Martin Guthrie, and Shadia Kawa checked the English.

Correspondence should be addressed to Mathias Pessiglione, Institut du Cerveau et de la Moëlle épinière, INSERM UMR 975, Hôpital Pitié-Salpêtrière, 47 Boulevard de l'Hôpital, 75013 Paris, France. E-mail: mathias.pessiglione@gmail.com.

DOI:10.1523/JNEUROSCI.1500-09.2009

Copyright © 2009 Society for Neuroscience 0270-6474/09/2913465-08\$15.00/0

Figure 1. Behavioral task and results. **A**, Trial structure. Subjects selected between left and right hand responses, corresponding to the two symbols displayed on screen, and subsequently observed the outcome (+0.5€ or nothing). In the illustrated example, the subject pressed the left button, corresponding to the left symbol, which was associated with a 75% probability of winning 0.5€ (vs nothing). **B**, Learning curves. Data points represent, trial by trial, the percentage of left versus right hand responses for cue pairs with equal (circles) or unequal (diamonds) reward probabilities. Dotted lines represent the learning curves generated from the computational model. **C**, Choice rate and option value. Histograms show the observed percentage of choice and modeled *Q* values for each option in the different pairs (gray: 25/25%, black: 75/75%, red: 75/25%, blue: 25/75%), at the end of learning sessions.

options that we hereafter consider at an abstract level encompassing both sensory and motor aspects.

Materials and Methods

Subjects. The study was approved by the Pitié-Salpêtrière Hospital ethics committee. Participants were recruited via e-mail and screened for exclusion criteria: left-handedness, age under 18, regular usage of drugs or medication, history of psychiatric or neurological illness, and contraindications to MRI scanning (pregnancy, claustrophobia, metallic implants). All subjects gave informed consent before inclusion in the study. They believed that they would be playing for real money, but to avoid discrimination, payoff was rounded up to a fixed amount of 100€ for every participant. Twenty subjects (11 females) aged between 19 and 31 (mean: 24.0 ± 2.8) years were included.

Behavioral tasks and analyses. Subjects were first asked to read the task instructions (see appendix in the supplemental material, available at www.jneurosci.org), which could be reformulated orally if necessary. The task was a probabilistic instrumental conditioning task with two motor responses (left or right) and two monetary outcomes (0.5€ or nothing). It was programmed on a PC using Cogent 2000 software (Wellcome Trust for Neuroimaging, London, UK). Subjects were trained on a practice version outside the scanner before performing the three test sessions within the scanner. Each session lasted 11 min, contained 96 trials, and used four different pairs of visual cues, which were letters taken from the Agathodaimon font. Each of the three test sessions was an independent task, containing new cues to be learned. Each cue was associated with a stationary reward probability (25 or 75%). The four cue pairs were randomly constituted and assigned to the four possible combinations of probabilities (25/25, 25/75, 75/25, 75/75%).

At the beginning of every trial, one pair was randomly selected and the two cues were simultaneously presented on the left and right side of a

central fixation cross. Subjects were required to select between the two cues by pressing one of the corresponding two buttons, with their left thumb to select the leftmost cue or their right thumb to select the rightmost cue. The two cues that comprised a pair always appeared together and at the same location, such that each cue could be directly mapped onto the corresponding motor response (left or right button press). To select a symbol, subjects had to continue pressing the corresponding button until the end of the response time interval (3000 ms). The selected symbol was then indicated by a red pointer appearing on screen for 500 ms. If the subject was not actively pressing one of the two buttons at the end of the response interval, the pointer remained under the fixation cross and the outcome was nil. At the end of every trial, the outcome was illustrated as a 50 cent coin picture, either boxed in a red square and labeled with “+0.5€” (for hits) or crossed out and labeled with “0€” (for misses), which remained on screen for 3000 ms. Random time intervals (jitters), drawn from a uniform distribution between 0 and 2 s, were inserted between trials to ensure better sampling of the hemodynamic response and to avoid the sleepiness that can arise from monotonous pace.

Subjects were encouraged to accumulate as much money as possible and were informed that some cues would result in a win more often than others. However, they were given no explicit information regarding reward probabilities, which they had to learn through trial and error. Of course, as far as payoff is concerned, learning mattered only for pairs with unequal probabilities (25/75 and 75/25%) and not for those with equal probabilities (25/25 and 75/75%). To assess learning, we examined the percentage of left or right hand responses in conditions with unequal versus equal probabilities. To control for manual lateralization bias we also compared overall response percentage and time between hands. Statistical significance was assessed using two-tailed paired *t* tests.

During the anatomical scan, and following the three test sessions of the instrumental conditioning task, subjects performed a supplementary

task designed to assess explicit knowledge of cue–outcome contingencies. Every cue seen during the conditioning task was displayed above an analog scale ranging from 0 to 100. Subjects were told to indicate, by moving the cursor along the scale, the probability of winning 0.5€ upon choosing the displayed cue. They had 10 s to move the cursor, using left and right buttons, toward the estimated appropriate position on the scale. To examine whether subjects were able to differentiate among the probabilities, we compared their estimations for poorly (25%) versus highly (75%) rewarded cues, again using two-tailed paired *t* tests.

Computational model. We fit the learning curves to a standard Q-learning algorithm (Sutton and Barto, 1998), which has been previously shown to offer a good account of instrumental choice in both human and nonhuman primates (O'Doherty et al., 2004; Samejima et al., 2005; Pessiglione et al., 2006; Frank et al., 2007b). For each pair of cues, the model estimates the expected values of left and right options, Q_L and Q_R , on the basis of individual sequences of choices and outcomes. These Q values essentially represent the expected reward obtained by taking a particular option in a given context. Q values were set at 0.25€ before learning, corresponding to the a priori expectation (50% chance of winning 0.5€). After every trial t of >0 , the value of the chosen option (e.g., L) was updated according to the following rule: $Q_L(t+1) = Q_L(t) + \alpha \cdot \delta(t)$. Following this rule, adapted from the Rescorla–Wagner model (Miller et al., 1995), option values are increased if the outcome is better than expected and decreased in the opposite case. Note that in our design, subjects could learn only about the chosen option, since they could not infer the fictive outcome of choosing the other option. In the equation, $\delta(t)$ was the prediction error, calculated as $\delta(t) = R(t) - Q_L(t)$, and $R(t)$ was the reward obtained as an outcome of choosing L at trial t . In other words, the prediction error $\delta(t)$ is the difference between the expected reward $Q_L(t)$ and the actual reward $R(t)$. The reward magnitude R was +0.5 for winning 0.5€ and 0 for winning nothing. Given the Q values, the associated probability (or likelihood) of selecting each option was estimated by implementing the softmax rule for choosing L, which is as follows: $P_L(t) = \exp(Q_L(t)/\beta) / (\exp(Q_L(t)/\beta) + \exp(Q_R(t)/\beta))$. This is a standard stochastic decision rule that calculates the probability of selecting one of a set of options according to their associated values. The learning rate, α , is a scaling parameter that adjusts the amplitude of value changes from one trial to the next. The temperature, β , is another scaling parameter that adjusts the randomness of decision making. These two free parameters, α and β , were adjusted to maximize, across all sessions and subjects, the log likelihood of the actual choices under the model. A single set of parameters was used to fit the choices for all pairs of cues, such that the estimation was based on the entire dataset. Optimal parameters were $\alpha = 0.10$ and $\beta = 0.17$. Log likelihoods did not differ between left and right hand responses (-235.9 ± 10.5 vs -241.2 ± 10.8), indicating that the model fit equally well on both sides. Statistical regressors for analysis of brain images were then generated using these optimal parameters.

Image acquisition and analysis. T2*-weighted echo planar images (EPIs) were acquired with BOLD contrast on a 3.0 tesla magnetic resonance scanner (Siemens Trio). A tilted-plane acquisition sequence was used to optimize functional sensitivity in the orbitofrontal cortex (Deichmann et al., 2003; Weiskopf et al., 2006). To cover the whole brain with a TR of 1.98 s, we used the following parameters: 30 slices; 2 mm slice thickness; 2 mm interslice gap. T1-weighted structural images were also acquired, coregistered with the mean EPI, segmented, and normalized to a standard T1 template and averaged across all subjects to allow group-level anatomical localization. EPI images were analyzed in an event-related manner, within a general linear model, using the statistical parametric mapping software SPM5 (Wellcome Department of Imaging Neuroscience, London, UK). The first 5 volumes of each session were discarded to allow for T1 equilibration effects. Preprocessing consisted of spatial realignment, normalization using the same transformation as structural images, and spatial smoothing using a Gaussian kernel with a full width at a half-maximum of 8 mm.

We used two statistical linear regression models to generate statistical parametric maps (SPMs) as follows (also see illustration in the supplemental material, available at www.jneurosci.org). Trials were sorted into two types according to the response given (left or right), which were distributed over separate regressors. Each trial was modeled as having

two time points, corresponding to cues and outcome display onsets, again separated in two different regressors. In the first model, stick functions were modulated by state values (calculated as the weighted sum of the two option values) at the time of cue onset and by prediction errors at the time of outcome. This first model hence contained six regressors of interest: three onset vectors (cue onsets followed by left and right responses plus outcome onsets), each modulated by one parameter (state value for cue onsets, prediction errors for outcomes). In the second model, we inserted two parametric modulations corresponding to left and right option values, Q_L and Q_R , to replace the regressors that contained state values. We also split the outcome vectors with respect to response side. There were therefore 10 regressors of interest in this second model—that is, 5 per response side: cue onset/left option value/right option value/outcome onset/prediction error. For both models, all regressors of interest were convolved with a canonical hemodynamic response function. To correct for motion artifacts, subject-specific realignment parameters were modeled as covariates of no interest.

Linear contrasts of regression coefficients were computed at the individual subject level and then taken to a group-level random effect analysis (one-sample *t* test). Activity related to hand movement was identified by contrasting cue onsets between the two response sides. Activity reflecting the modeled variables (state value, option value, and prediction error) was isolated in contrasts including the relevant parametric modulation of onsets for the two response sides. An inclusive mask, made up of all voxels that significantly correlated with state value, was applied in the group-level analysis of option values. All activations disclosed on glass brains and reported in the text survived a threshold of $p < 0.001$ (uncorrected) and contained a minimum of 30 contiguous voxels. To show the extent of these activations, we also took slices with less conservative threshold ($p < 0.01$). Finally, to check the specificity of the reported activations, we independently defined regions of interest (ROIs). Ventral prefrontal cortex (VPFC) ROIs were identified as clusters reflecting the state value at cue onset ($p < 0.001$, uncorrected). Primary motor cortex ROIs were defined as the precentral gyri of the AAL (automated anatomical labeling) atlas provided by the MRIcro software (www.mricro.com). Lateralization effects were assessed in terms of both size (number of voxels) and magnitude (regression coefficients) of activations obtained with the different regressors of interest. Size effects were assessed using χ^2 tests that compared the distribution of activated voxels over left and right ROIs to a symmetrical distribution (half on the left and half on the right). Magnitude effects were assessed using paired *t* tests that compared regression coefficients between left and right ROIs.

Results

Behavior

Behavioral data showed significant effects ($t_{(19)} = 5.5$, $p < 0.001$, two-tailed paired *t* test) of learning when cue pairs had unequal probabilities (25/75 and 75/25%), relative to those with equal probabilities (25/25 and 75/75%): average percentages of left/right hand responses were 50.0/50.0, 21.5/78.5, 77.3/22.7, and 46.5/53.5% for the 25/25, 25/75, 75/25, and 75/75% pairs, respectively. Percentage of correct responses did not differ between the two unequal pairs, indicating that subjects learned to move the most rewarded hand equally well (77.3 ± 3.2 and $78.5 \pm 3.3\%$ for left and right, respectively), regardless of the side (Fig. 1*B*). Moreover, neither response rate (48.9 ± 2.1 vs $51.1 \pm 2.1\%$) nor response time (1607 ± 60 vs 1563 ± 59 ms, respectively) differed between left and right hand responses, adding evidence that behavioral performance was in all respects symmetrical between the two sides. The model provided an equally good fit on both sides, and the Q values approximately converged to the reward probability associated with each option (Fig. 1*C*). Note that introduction of symmetrical pairs allowed disentangling option value (much lower for the 25/25 than for the 75/75% pair) from choice propensity ($\sim 50\%$ in both cases).

Learning appeared to be rather implicit, since during the explicit probability estimation task that followed conditioning sessions, sub-

jects failed to distinguish between highly (75%) versus poorly (25%) rewarded cues (average estimations were 48.2 ± 1.9 and $49.2 \pm 0.8\%$, respectively). If subjects were not explicitly tracking reward probabilities, different hypotheses can be formulated regarding the underlying mechanisms that incorporate reward information to improve next choices. A first possibility is that rewards reinforce relevant stimulus–response circuits via modulation of synaptic efficacy, such that representation of reward probabilities (option values) would be useless. Another possibility is that option values are indeed calculated and represented in some brain regions, such that they can influence decisions without necessary access to the subject's conscious mind.

Neuroimaging

To test whether our design was able to replicate classical findings of the neuroimaging literature, we first examined the representation of state value and hand movement at the time of cue onset and of prediction error at the time of outcome (Fig. 2). State value was defined as the sum of the two options weighted by the likelihood that they will be chosen ($P_L \cdot Q_L + P_R \cdot Q_R$), which was calculated using a softmax function (see Materials and Methods). It can be thought of as the reward to be expected given the cues displayed on screen, regardless of the upcoming action. Regression coefficients were estimated separately for trials that resulted in left and right hand responses and then integrated in the contrast done to isolate brain activity correlating with state value independent of the chosen option. SPMs revealed large bilateral clusters: principally in the ventral prefrontal cortex (ventrolateral, orbitofrontal, and anterior cingulate) but also in the middle cingulate cortex, angular and middle temporal gyri, precuneus, and cerebellum (Fig. 2A, top). Contrasts between left and right responses yielded contralateral clusters in the motor cortex (primary, premotor, and supplementary motor areas), putamen, and thalamus, as well as ipsilateral clusters in the cerebellum (Fig. 2B). Neuronal activity correlating with reward prediction errors was found bilaterally in the basal ganglia (ventral striatum, putamen, pallidum, and thalamus), limbic temporal regions (amygdala and hippocampus), primary motor cortex, ventral and dorsal prefrontal cortex, middle and posterior cingulate cortex, angular and parietal inferior gyri, precuneus, and cerebellum (Fig. 2A, bottom).

We now turn to the representation of option values, which, to our knowledge,

Figure 2. Representation of state value, hand movements, and prediction errors. Coronal slices correspond to the blue lines on axial glass brains. Areas colored in gray-to-black gradient on glass brains and in red-to-white gradient on slices showed significant effects ($p < 0.001$, uncorrected). [$x y z$] coordinates of maxima refer to the Montreal Neurological Institute (MNI) space. Color bars indicate t values. **A**, Statistical parametric maps of activity correlating with state value (top) and prediction error (bottom). State values were calculated as a weighted sum of the two option values generated from the computational model and modeled at the time of cue onset. Prediction errors were modeled at the time of outcome onset, whatever the chosen option. **B**, Statistical parametric maps of activity related to left and right hand movement. Hemodynamic responses were modeled at the time of cue onset and contrasted between left and right movements.

Figure 3. Parsing representations of the two options. **A**, Statistical parametric maps of activity correlating with option values. Sagittal slices correspond to the blue lines on glass brains. Areas colored in gray-to-black gradient on glass brains and in red-to-white gradient on sagittal slices showed significant effects ($p < 0.001$ and $p < 0.01$ uncorrected, respectively). ACC, Anterior cingulate cortex; OFC, orbitofrontal cortex; VLPFC, ventrolateral prefrontal cortex. **B**, Statistical parametric maps of activity correlating with prediction errors, calculated as the difference between the actual outcome and the value of the chosen option. Coronal slices correspond to the blue lines on glass brains. Areas colored in gray-to-black gradient on glass brains and in red-to-white gradient on sagittal slices showed significant effects ($p < 0.001$ uncorrected). PE_L , Prediction error following left response; PE_R , prediction error following right response. $[x y z]$ coordinates of maxima refer to the MNI space. Color bars indicate t values.

have never been distinguished in the human brain. We used the same analysis as above, replacing the single regressor that previously incorporated state values by two regressors incorporating left and right option values (Q_L and Q_R). We found that option values were specifically represented in the contralateral ventral prefrontal cortex (Fig. 3A). To formally test this lateralization of option value representation, we compared the left/right distribution of activated voxels within the VPFC to a symmetrical distribution (half on the left and half on the right). The comparison was highly significant ($p < 0.001$, χ^2 test) for both Q_L (left/right VPFC: 70/242 voxels), and Q_R (left/right VPFC: 90/0 voxels) ($p < 0.001$). The comparisons remained highly significant ($p < 0.001$, χ^2 test) when counting the voxels activated at a lower threshold ($p < 0.01$, uncorrected), for both Q_L (left/right VPFC: 464/750 voxels) and Q_R (left/right VPFC: 958/486 voxels). Importantly, the same significance levels ($p < 0.001$, χ^2 test) were attained in all cases when restricting the analysis to trials when the option was chosen (left/right VPFC: 153/221 voxels with Q_L for left responses and 146/30 voxels with Q_R for right responses) or not chosen (left/right VPFC: 0/34 voxels with Q_L for right responses and 172/66 voxels with Q_R for left responses). Hence, the apparent representation of state value in the VPFC was actually driven by each hemisphere encoding the value of the contralateral option, whether it was chosen or not. Note that at the lower threshold ($p < 0.01$, uncorrected), VPFC activation extended to the ventral striatum, in relation to option as well as state values, with a similar lateralization effect between Q_L and Q_R representations. This result must however be taken with caution, given that $p < 0.01$ (uncorrected) may be too permissive a threshold.

We also sought to dissociate representations of prediction errors corresponding to chosen options. Although a trend could be noticed over the striatum on the maps (Fig. 3B), there was no clear evidence for the left and right prediction errors being differentially expressed in the two hemispheres. Lateralization effects were searched for by testing the left/right distribution of activated voxels against a symmetrical distribution, as was done for option values. Activated voxels were counted at a threshold of $p < 0.001$ (uncorrected), within an anatomically defined mask of the striatum. There was no significant effect, neither for right prediction errors (left/right striatum: 387/419

Figure 4. Dissociation between value- and response-induced lateralization effects. **A**, VPFC ROIs refer to the clusters reflecting state value at the time of cue onset. M1 ROIs were taken from a standard anatomical atlas (AAL atlas). **B**, Histograms represent Z-scored laterality, calculated as the difference in regression coefficients (betas) between right and left ROIs. Bars are inter-subject SEs of the mean. Regressors of interest: Q_L and Q_R refer to parametric modulation of cue onsets by left and right option values (top); L and R refer to cue onsets for left and right responses (middle); Q_C and Q_{NC} refer to parametric modulation of cue onsets by chosen and nonchosen option values (bottom). * $p < 0.05$, *** $p < 0.001$, paired t test.

voxels, $p > 0.1$, χ^2 test) nor for left prediction errors (left/right striatum: 667/655 voxels, $p > 0.5$, χ^2 test). A similar analysis was done for voxels activated ($p < 0.001$, uncorrected) by left and right responses, which were counted within anatomically defined M1 masks. There was a highly significant effect ($p < 0.001$; χ^2 test) for both left responses (left/right M1: 582/3230 voxels) and right responses (left/right M1: 2338/878 voxels). Thus, contralateral activation was obtained at the time of cue onset, in relation to option values for VPFC, and to hand movement for motor cortex (M1).

To further verify the specificity of lateralization effects, we extracted the regression coefficients from the VPFC (all voxels reflecting state values) and M1 (all voxels in the precentral gyrus) ROIs (Fig. 4A). Note that definition of the two ROIs was independent from the parameter they were found to reflect (option value for VPFC and hand movement for M1). Laterality was calculated as the difference between regression coefficients obtained from right versus left ROIs (Fig. 4B, illustration). The laterality

was negative for Q_R (more activation in the left VPFC) and positive for Q_L (more activation in the right VPFC), with a significant difference between the two ($t_{(19)} = 1.9$, $p < 0.05$, paired t test). Importantly, the laterality of activations was not significantly different in VPFC regions when comparing between left and right hand responses (L and R) or chosen and nonchosen option values (Q_C and Q_{NC}). Contralateral representation of option values in the VPFC was hence not driven by the upcoming choice or movement. In contrast, the laterality of M1 activations was significantly different between L and R hand responses ($t_{(19)} = 13.9$, $p < 0.001$, paired t test) but not between Q_L and Q_R or between Q_C and Q_{NC} . Thus, VPFC expressed contralateral option values independent of the chosen option, and M1 the contralateral movement made for choosing the option, independent of its value.

Discussion

We first replicated common findings on brain representation of expected rewards, movement execution, and prediction errors. Indeed, expected rewards have been reported to involve the ventral prefrontal cortex as well as the ventral striatum (Tremblay and Schultz, 1999; O'Doherty et al., 2002; Knutson et al., 2005; Padoa-Schioppa and Assad, 2006), whereas movement execution has been reported to implicate the motor cortex as well as the posterior putamen (Middleton and Strick, 2000; Haber, 2003; Lehericy et al., 2004; Mayka et al., 2006). Thus, our analysis was able to dissociate limbic circuits known to encode reward-related information from motor circuits related to movement execution. The terms used for this ventro/dorsal dissociation are voluntarily general, as our design cannot disentangle between close reward-related computational variables. It confounds for instance a weighted state value (the reward to be expected given the two cues) versus a cue prediction error (the latter minus the average reward prediction). Also consistent with previous reports (Pagnoni et al., 2002; O'Doherty et al., 2004; Pessiglione et al., 2006; Lohrenz et al., 2007), regions expressing prediction errors at the time of outcome largely overlapped with the above regions related to both expected reward and hand movement at the time of cue onset. These results support the notion that reward prediction errors are used as a teaching signal to improve both reward prediction and movement selection in the corresponding limbic and motor brain circuits.

The replicated findings were important steps in discovering the neural underpinnings of reinforcement learning, but did not provide full account of how the human brain makes an economic choice. Indeed, the state value cannot tell the subject what button to press, contrary to option values, which provide a criterion for making a decision. When we replaced in our linear regression model the state value by the option values, we found that significant activation was confined to the contralateral ventral prefrontal cortex. This cannot be due to the upcoming movement itself, since ventral prefrontal activity was not different for left and right hand responses. Conversely, motor cortex activity distinguished left and right hand responses, but did not incorporate option values. Moreover, the option value representation in the ventral prefrontal cortex was unaffected by choices: it remained contralateral regardless of which hand was eventually moved. Such results indicate that each brain hemisphere tracks the value of the option under its control, regardless of whether it is eventually chosen or not. Note however that our design cannot tell whether the lateralization is due to the response or to the stimulus, since the two factors were confounded to increase chances of dissociating option value representations. Further experiments, in which cue lo-

cations on the screen would be crossed with response sides, would be needed to clarify this question.

Locating option values in the human brain provides a neural substrate on which decision-making could rely in principle, but it does not explain how the best option is selected. A direct implementation of the rational agent envisaged in economics would postulate that a decision-making system constitutes the neuronal counterpart of conscious deliberation. In this view, the alleged system would compare option values, select the highest and send orders to dedicated motor effectors. We can only speculate here about the neuronal implementation of such a decision-making system. Because option values were distributed over the two hemispheres, it would necessarily include several distant brain regions to cover the entire process. A large-scale brain network might even be involved, such as the prefrontoparietal system that has been associated with conscious cognitive control (Baars, 2005; Maia and Cleeremans, 2005; Badre, 2008). However, the anarchic hand syndrome observed in split-brain patients would rather suggest that interhemispheric fibers, and not an external controller, are responsible for preventing each hemisphere from following its own goal. A direct competition mechanism, for instance via mutual inhibition, might therefore account for how the best option is selected.

A more parsimonious view would thus be that options are selected through direct competition between effectors, with the winner taking control of body movement. The latter view has been formalized in various computational models inspired by robotics, in which action selection can be achieved without the need for central supervision (Redgrave et al., 1999; Bar-Gad and Bergman, 2001; Joel et al., 2002; Leblois et al., 2006; Cisek, 2007; Frank et al., 2007a; Houk et al., 2007). In the neuronal implementation of these models, the different options are represented in motor frontostriatal circuits, which are considered an actor system. To explain how action selection may be driven by rewards, some models postulate that a critic system, implemented in the limbic frontostriatal circuits, communicates reward predictions and prediction errors to the actor system. Our results are compatible with these models, and would imply that available options are separately represented in limbic (critic) as well as motor (actor) circuits. Thus, at the time of decision making, neuronal representation of the different options may be gradual in limbic circuits, in proportion to reward prediction, and all-or-none in motor circuits, depending on the action selected by the competition mechanism.

Such models leave aside the question of why humans experience themselves as unified agents, able to purposely select the actions in which they engage, based on conscious estimation of costs and benefits. One way to articulate the two above views would be assuming that subconscious decisions are achieved through direct competition, whereas conscious deliberation requires a controlling system, which could possibly involve large-scale prefrontoparietal networks. Whether decisions were consciously deliberated or not in our case remains unclear, but that subjects failed to distinguish the option values in their postlearning explicit estimations may give some clue. It would suggest that the influence of encoded option values on behavioral choices was largely subconscious, in keeping with a previous demonstration of subliminal instrumental learning (Pessiglione et al., 2008). Caution must be taken however, as debriefing tasks are confounded by memory decay. Subjects could have been transiently aware of the option values during the course of learning, and then forgotten them before debriefing. Further experiments would thus be needed to clarify the relation-

ships between the neural substrates that encode option values and those that underpin conscious deliberation.

The demonstration that neuronal representation of reward prediction (state value) can be parsed into available option values has been made here in the case of intermanual choice. It might be possible to generalize the idea that option value representation is spatially distributed, based on which brain region is in charge of processing sensory or motor aspects of the different alternatives. For instance, frontostriatal circuits might separately represent the expected value of moving the hand, the foot, one finger, and so forth. In other words, a reward somatotopy could exist in the human brain, similar to those already established for sensory and motor systems. This concept might also apply to higher levels, which would involve representing expected values of complex actions, or even beyond sensorimotor processes, to cognitive tasks such as reading, reasoning, or remembering. In sum, the expected value of engaging different neuronal populations, dedicated to specific tasks, might be topographically represented in frontostriatal circuits. In primate prefrontal cortex or striatum, certain neuronal activities have been reported to express both expected reward and various task dimensions, such as target location (Watanabe, 1996; Leon and Shadlen, 1999; Kobayashi et al., 2006) or movement direction (Lauwereyns et al., 2002; Matsumoto et al., 2003; Samejima et al., 2005; Pasquereau et al., 2007). However, the topographical organization of neuronal populations encoding expected values, and their connection with the corresponding effectors, remain to be established.

References

- Baars BJ (2005) Global workspace theory of consciousness: toward a cognitive neuroscience of human experience. *Prog Brain Res* 150:45–53.
- Badre D (2008) Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12:193–200.
- Bar-Gad I, Bergman H (2001) Stepping out of the box: information processing in the neural networks of the basal ganglia. *Curr Opin Neurobiol* 11:689–695.
- Baynes K, Tramo MJ, Reeves AG, Gazzaniga MS (1997) Isolation of a right hemisphere cognitive system in a patient with anarchic (alien) hand sign. *Neuropsychologia* 35:1159–1173.
- Cisek P (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos Trans R Soc Lond B Biol Sci* 362:1585–1599.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
- Deichmann R, Gottfried JA, Hutton C, Turner R (2003) Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19:430–441.
- Frank MJ, Scheres A, Sherman SJ (2007a) Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. *Philos Trans R Soc Lond B Biol Sci* 362:1641–1654.
- Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007b) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci U S A* 104:16311–16316.
- Gazzaniga MS (2005) Forty-five years of split-brain research and still going strong. *Nat Rev Neurosci* 6:653–659.
- Gläscher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex* 19:483–495.
- Haber SN (2003) The primate basal ganglia: parallel and integrative networks. *J Chem Neuroanat* 26:317–330.
- Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci* 28:5623–5630.
- Haruno M, Kuroda T, Doya K, Toyama K, Kimura M, Samejima K, Imamizu H, Kawato M (2004) A neural correlate of reward-based behavioral learning in caudate nucleus: a functional magnetic resonance imaging study of a stochastic decision task. *J Neurosci* 24:1660–1665.
- Houk JC, Bastianen C, Fansler D, Fishbach A, Fraser D, Reber PJ, Roy SA, Simo LS (2007) Action selection and refinement in subcortical loops

- through basal ganglia and cerebellum. *Philos Trans R Soc Lond B Biol Sci* 362:1573–1583.
- Joel D, Niv Y, Ruppert E (2002) Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw* 15:535–547.
- Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4:e233.
- Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. *J Neurosci* 25:4806–4812.
- Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53:147–156.
- Kobayashi S, Nomoto K, Watanabe M, Hikosaka O, Schultz W, Sakagami M (2006) Influences of rewarding and aversive outcomes on activity in macaque lateral prefrontal cortex. *Neuron* 51:861–870.
- Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
- Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–417.
- Leblois A, Boraud T, Meissner W, Bergman H, Hansel D (2006) Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. *J Neurosci* 26:3567–3583.
- Lehéricy S, Ducros M, Van de Moortele PF, Francois C, Thivard L, Poupon C, Swindale N, Ugurbil K, Kim DS (2004) Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Ann Neurol* 55:522–529.
- Leon MI, Shadlen MN (1999) Effect of expected reward magnitude on the response of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–425.
- Lohrenz T, McCabe K, Camerer CF, Montague PR (2007) Neural signature of fictive learning signals in a sequential investment task. *Proc Natl Acad Sci U S A* 104:9493–9498.
- Maia TV, Cleeremans A (2005) Consciousness: converging insights from connectionist modeling and neuroscience. *Trends Cogn Sci* 9:397–404.
- Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301:229–232.
- Mayka MA, Corcos DM, Leurgans SE, Vaillancourt DE (2006) Three-dimensional locations and boundaries of motor and premotor cortices as defined by functional brain imaging: a meta-analysis. *Neuroimage* 31:1453–1474.
- Middleton FA, Strick PL (2000) Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res Brain Res Rev* 31:236–250.
- Miller RR, Barnet RC, Grahame NJ (1995) Assessment of the Rescorla-Wagner model. *Psychol Bull* 117:363–386.
- O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. *Neuron* 33:815–826.
- O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
- Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223–226.
- Pagnoni G, Zink CF, Montague PR, Berns GS (2002) Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci* 5:97–98.
- Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T (2007) Shaping of motor responses by incentive values through the basal ganglia. *J Neurosci* 27:1176–1183.
- Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* 442:1042–1045.
- Pessiglione M, Petrovic P, Daunizeau J, Palmiter S, Dolan RJ, Frith CD (2008) Subliminal instrumental conditioning demonstrated in the human brain. *Neuron* 59:561–567.
- Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci* 27:9984–9988.
- Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* 9:545–556.
- Redgrave P, Prescott TJ, Gurney K (1999) The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience* 89:1009–1023.
- Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
- Sperry RW (1961) Cerebral organization and behavior: the split brain behaves in many respects like two separate brains, providing new research possibilities. *Science* 133:1749–1757.
- Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.
- Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S (2004) Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7:887–893.
- Tremblay L, Schultz W (1999) Relative reward preference in primate orbitofrontal cortex. *Nature* 398:704–708.
- Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci* 27:4019–4026.
- Watanabe M (1996) Reward expectancy in primate prefrontal neurons. *Nature* 382:629–632.
- Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. *Neuroimage* 33:493–504.