

Neural computations underlying action-based decision making in the human brain

Klaus Wunderlich^{a,1}, Antonio Rangel^{a,b}, and John P. O'Doherty^{a,b,c}

^aComputation and Neural Systems Program, California Institute of Technology, Pasadena, CA; ^bDivision of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA; and ^cInstitute of Neuroscience and School of Psychology, Trinity College, Dublin, Ireland

Edited by Ranulfo Romo, Universidad Nacional Autónoma de México, Mexico, D.F., Mexico, and approved August 6, 2009 (received for review February 4, 2009)

Action-based decision making involves choices between different physical actions to obtain rewards. To make such decisions the brain needs to assign a value to each action and then compare them to make a choice. Using fMRI in human subjects, we found evidence for action-value signals in supplementary motor cortex. Separate brain regions, most prominently ventromedial prefrontal cortex, were involved in encoding the expected value of the action that was ultimately taken. These findings differentiate two main forms of value signals in the human brain: those relating to the value of each available action, likely reflecting signals that are a precursor of choice, and those corresponding to the expected value of the action that is subsequently chosen, and therefore reflecting the consequence of the decision process. Furthermore, we also found signals in the dorsomedial frontal cortex that resemble the output of a decision comparator, which implicates this region in the computation of the decision itself.

acc | action value | reinforcement learning | sma | vmPFC

Consider a goalkeeper trying to stop a soccer ball during a penalty kick. Within a brief amount of time he needs to choose between jumping to the left or right goal posts. Repeated play against the same opponents allows him to learn about their scoring tendencies, which can be used to compute the values of a left and a right jump before making a decision. It is a long-established view in economics, psychology, and computational neuroscience that the brain makes choices among actions by first computing a value for each possible action, and then selecting one of them on the basis of those values (1–3). This raises two fundamental questions in decision neuroscience: (1) where in the brain are the values of different types of actions encoded? and (2) how and where does the brain compare those values to generate a choice?

An emerging theme in decision neuroscience is that organisms need to make a number of value-related computations to make even simple choices (4). Consider the case of action-based choice exemplified by the goalkeeper's problem. First, he needs to assign a value to each action under consideration. These signals, known as action values, encode the value of each action before choice and regardless of whether it is subsequently chosen or not, which allows them to serve as inputs into the decision-making process (5–7). Second, these action values are compared to generate a choice. Third, the value of the option that is selected, known as the chosen value, is tracked to be able to do reinforcement learning. In particular, by comparing the value of the outcome generated by the decision to the chosen value, the organism can compute a prediction-error signal that can be used to update the action value of the chosen option. Note that while the action values are computed before the decision is made, the chosen value and outcome of the comparator process signals are computed afterward.

Although a rapidly growing number of studies have found neural responses that are correlated with some form of value signals, little is known about how the brain encodes action values or about how it compares them. This is central to understand how the brain makes action-based choices. For example, a number of chosen value signals have been found in the orbital and medial prefrontal cortex

(8, 9) and amygdala (10, 11). Note that these signals are quite distinct from action values, and are not precursors to choice, because they reflect the value of the actions that were selected in the decision. For similar reasons, the value signals that have been found in lateral intraparietal cortex (LIP) during saccadic action-based choice (12, 13) are also not pure action values since they are strongly modulated by whether an action is subsequently taken. This suggests that instead of serving as inputs to the comparison process, they reflect its output. Several studies found orbitofrontal cortex to encode the value of different goals (14–16). Although these signals are precursors of choice, they are not instances of action values since they are stimulus-based and independent of the action required to obtain them. To date, only three monkey electrophysiology studies have found evidence for the presence of action-value signals for hand and eye movements in the striatum during simple decision-making tasks (5–7). This study extends their findings in three directions. First, as of yet no evidence has been presented for the existence of action-value signals in the human brain. Second, using fMRI we are able to look for action-value signals in the entire brain, whereas the previous electrophysiology studies have limited their attention to the striatum. As a result, no previous study has looked for action-value signals in the cortex. This is important because, as discussed below, there are a priori reasons to believe that action value signals might be found in the motor and supplementary motor cortices. Finally, we investigate how such signals might be compared to actually compute the decision itself and where neuronal correlates of the output of this decision process are represented, an issue about which very little is known.

We studied these questions using fMRI in humans while subjects performed a variant of a two-armed bandit task to obtain probabilistically delivered monetary rewards (Fig. 1A). A critical feature of the task was that they had to select a motor response in one of two distinct response modalities: in every trial, they could choose to make either a saccade to the right of a fixation cross, or to press a button with the right hand. This design allowed us to exploit the fact that different regions of the cortex are involved in the planning of eye and hand movements (17). We hypothesized that value representations for the two actions would be separable within these cortical areas at the spatial resolution available to fMRI. The probability of being rewarded on each of the two actions drifted randomly over time and was independent of the probability of being rewarded on the other (Fig. 1B). This characteristic ensured that value estimates for eye and hand movements were uncorrelated, which gave us maximum sensitivity with which to dissociate the neural representations of the two action values.

Author contributions: K.W., A.R., and J.P.O. designed research; K.W. performed research; K.W. contributed new reagents/analytic tools; K.W. analyzed data; and K.W., A.R., and J.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: kwunder@caltech.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0901077106/DCSupplemental.

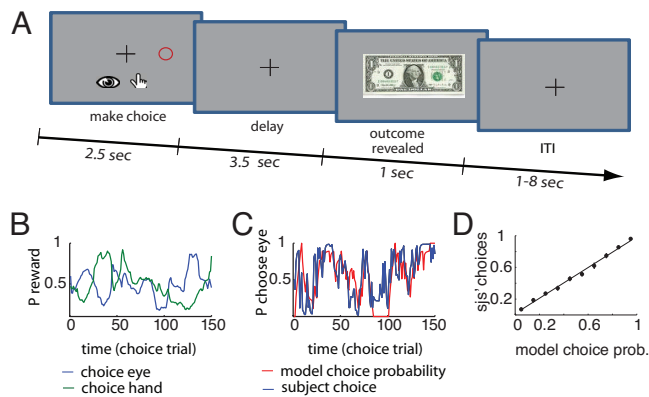


Fig. 1. Experimental Design and Behavior. (A) Subjects were presented with a choice cue after which they had to respond within 2.5 s by performing a saccade to the red target circle or a right handed button press. Once a response was registered the screen was immediately cleared for a short delay and subsequently the outcome was revealed (6 s after trial onset) indicating either receipt of reward or no reward. Inter-trial-intervals varied between 1 and 8 s. (B) Example reward probabilities for saccades and button presses as a function of the trial number. The probability of being rewarded following choice of either the hand or eye movement was varied across the experiment independently for each movement. (C) Fitted model choice probability (red) and actual choice behavior (blue) shown for a single subject. (D) Actual choice behavior versus model predicted choice probability. Data are pooled across subjects, the regression slope is shown as a line, vertical bars, SEM.

To look for neural correlates of action values we had to estimate the value of taking each action in every trial. We calculated the action values using a computational reinforcement-learning (RL) model in which the value of each action, V_{eye} and V_{hand} , was updated in proportion to a prediction error on each trial (see Table S1 for a summary of how the different types of value signals relate to the components of the experiment). The model also assumed that action selection in every trial followed a soft-max probability rule based on the difference of the estimated action values (8). To test for the presence of action value signals in the brain we took the model predicted trial-by-trial estimates of the two action values and entered these into a regression analysis against the fMRI data. In addition to a whole brain screening for the presence of action-value signals, we specifically looked for them in areas known to be involved in the planning of motor actions, including supplementary motor cortex (18–21) and lateral parietal cortex (22, 23). Given that both of these areas have previously been shown to contain value-related signals for movements in nonhuman primates, and that they are closely interconnected with the area of motor cortex involved in carrying out motor actions (24–26), we considered these areas prime candidates for containing action-value representations that could then be used to guide action-based choices. It is important to emphasize, however, that the tasks used in previous studies did not make it possible to determine if the value signals identified were chosen values or action values.

We also looked for areas that are involved in comparing the action values to make a choice. Two areas of a priori interest were the anterior cingulate cortex (ACC) and the dorsal striatum. ACC has been previously implicated in action-based choice, both in the context of a human imaging study reporting activity in this area during a task involving choices between different actions compared to a situation involving responses guided by instruction (27), and in a monkey lesion study where ACC lesions produced an impairment in action-outcome based choice but not in mediating changes in responses following errors (28). Dorsal striatum has been implicated in both goal-directed and habitual instrumental responding for reward in rodents (29, 30). Moreover, human fMRI studies reveal increased activity in both of these regions when subjects

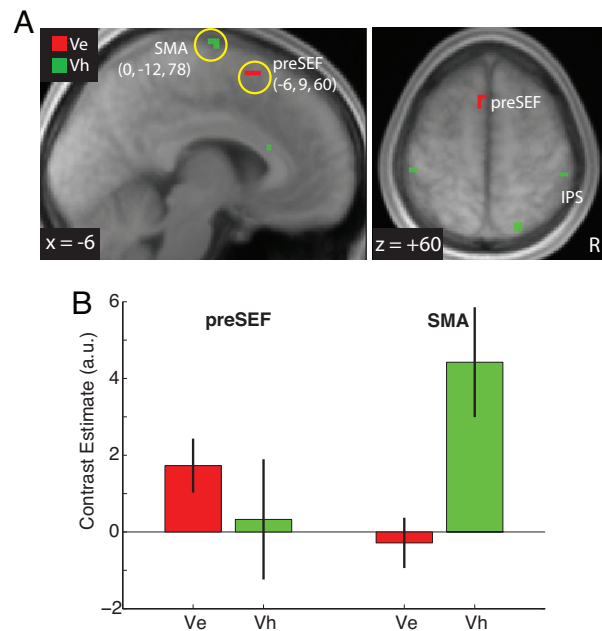


Fig. 2. Action values. (A) Region of supplementary motor area showing correlations with action values for hand movement (Vh/green) and a region of pre-SEF showing correlations with action-values for eye movements (Ve/red). T-maps are shown from a whole brain analysis thresholded at $P < 0.001$ uncorrected (see Fig. S1 for a version with color bars relating to t stats). (B) Average effect sizes of Ve (red) and Vh (green) extracted from SEF and SMA. The effects shown here were calculated from trials independent of those used to functionally identify the ROI. Note that only Ve but not Vh modulate the signal in preSEF, and that activity in SMA shows the opposite pattern. Vertical lines, SEM.

make choices to obtain reward compared to an otherwise analogous situation in which the rewards are obtained without the need to make a choice (31–34).

The most simple type of comparison process would be to compute a difference between the two action values. We tested for such a difference, but as we had no a priori hypothesis about the directionality of the computation, we tested for both the difference between the value of the action chosen and the value of action not chosen ($V_{chosen} - V_{unchosen}$), and one involving the opposite difference ($V_{unchosen} - V_{chosen}$). As we found evidence for such an action-value comparison signal in the brain, we then proposed a simple computational model to provide a conceptual explanation as to how such a signal could reflect the output of a computationally plausible decision mechanism.

Results

RL Model Fits to Behavioral Choice Data. A comparison of the choice probabilities predicted by the RL model and the soft-max procedure to subjects' actual behavior suggests that the model matches subjects behavior well. Fig. 1C compares both variables for a typical subject. Fig. 1D compares the predicted choice probability (binned) against the actual choice probabilities for the group. A similar linear regression analysis at the individual level generated an average R^2 across subjects of 0.83 and regression coefficients that were significant at $P < 0.001$ in each subject.

Action Values. We found neural activity correlating with the action values for making a hand movement in left supplementary motor area (SMA; Fig. 2A and Table S2). A region of interest (ROI) analysis showed that activity in this area satisfied the properties of a hand action value: it was sensitive to the value of hand movements, and it showed no response selectivity to the value of eye movements (Fig. 2B). Activity in lateral parietal cortex, ACC, and right dorsal

putamen also correlated with hand action values. In contrast, activity in a region of left supplementary motor cortex anterior to the SMA (presupplementary eye fields, preSEF, Fig. 2*A* and Table S2) correlated with action values for eye movements. A similar ROI analysis showed that the area satisfied the properties of an eye action value: it was sensitive to the value of eye movements, but showed no sensitivity to the value of hand movements (Fig. 2*B*). We tested this by performing a two-way ANOVA with the factors of area (SEF vs. SMA) and action value (eye vs. hand). There was no significant main effect for either area or action value but the interaction was significant at $P = 0.03$ ($F = 5.6$, $df = 1$). Another important feature of an action value signal is that, since it is a precursor of choice, it should not depend on which action is actually chosen. We tested for this property by computing the following two voxel-wise interaction contrasts ($V_e|eye - V_e|hand$) $\neq 0$ and ($V_h|hand - V_h|eye$) $\neq 0$. We found no significant interaction between action value and chosen action in either SMA or preSEF at $P < 0.05$ uncorrected. A post-hoc plot of the average percent signal change within each cluster plotted as a function of high and low action values are shown in Fig. S2.

One potential explanation for these correlations is that activity in the SEF and SMA reflect motor preparation. To help exclude the possibility, we carried out two additional analyses. First, we estimated a model that used reaction times (RT) as a proxy index of the degree of motor preparation on a given trial and found hand and eye RTs did not show the same pattern of differential correlations in SMA and SEF as exhibited by our action-value regressors. Second, we estimated a version of our main general linear model in which the RTs were included as a covariate of no interest alongside our action-value signals, and found the action-value results in SMA and SEF still survived at $P < 0.005$ uncorrected. Both results suggest that simple motor preparation is unlikely to account for the correlations with action values identified above.

Another alternative potential explanation for the correlations between activity in SMA/pre-SEF and action values is that signal fluctuations in these areas depend on the degree to which subjects currently choose those motor actions. For example when the value of a hand movement is high, the subject may tend to choose hand actions more often, and therefore activity in SMA may be increased as a result of enhanced overall motor excitability. We tested for this possible confound by regressing BOLD signals against the degree to which subjects favored one or other action in the recent past. We found activation most prominently within the occipital lobe, primary motor cortex, cerebellum, and dorsal medial frontal gyrus. However, we did not find any significant correlation within our previously identified action value areas, ruling out this possible explanation for the action-value signals in SMA, SEF, and elsewhere (see Table S2 and Fig. S3).

Chosen Values. We then looked for correlates of the value of the action that is chosen on a particular trial, irrespective of its modality. Consistent with previous findings (8, 9), we found chosen value modulated activity in a number of brain areas, most prominently the ventromedial prefrontal cortex (vmPFC) extending onto the medial orbital surface (Fig. 3*A* and Table S2). The parietal cortex, including bilateral IPS and right LIP were also activated by this contrast.

We also tested for areas correlating with the chosen value only on occasions when the action chosen was a hand movement, and for areas correlating with chosen value only on trials in which the eye movement was chosen. Intriguingly, we found evidence for a topographical arrangement of action specific chosen value signals in vmPFC along the anterior-posterior axis, whereby a mid-vmPFC region correlated with hand values only when hand movements were selected, and a region of more posterior vmPFC correlated with the value of eye movements only on trials when eye movements were selected. These two action specific representations were both located caudal to the value chosen signal reported above (Fig. 3*B*).

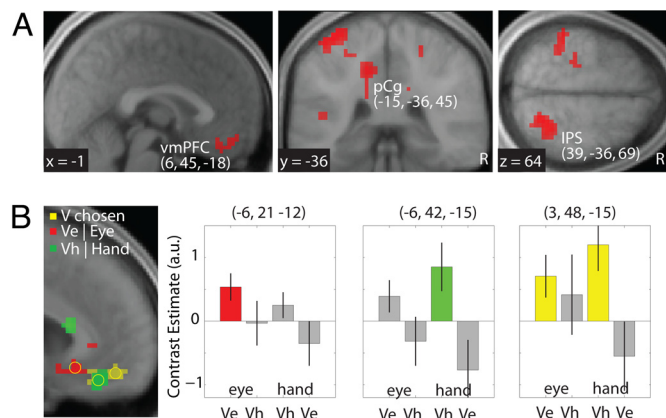


Fig. 3. Chosen values. (A) Brain regions showing significant correlations with the value of the action chosen. Areas shown include vmPFC, intra-parietal sulcus and posterior cingulate cortex. Threshold is set at $P < 0.001$. (B) Distinct forms of the value chosen signal are present within vmPFC. The area depicted in yellow indicates voxels that correlate with the value of the chosen action irrespective of whether the action taken is a hand or an eye movement. The area depicted in green correlates only with the value chosen on trials when the hand movement is chosen but not when the eye movement is chosen. Finally the area depicted in red indicates voxels correlating with value chosen only on trials when the eye movement is selected but not the hand movement. The results suggest an anterior to posterior trend in the selectivity of voxels to these different types of value chosen signals. Bar plots show effect sizes averaged across subjects for the action specific value chosen signals in the three areas (left: red area, middle: green area, right: yellow area). Bars shown in chromatic color are significantly different from zero (t test, $P < 0.05$). Similar to bar plots in Fig. 2*B*, effects were calculated from a data sample independent of the one used to functionally identify the ROI. Vertical lines, SEM.

Action-Value Comparison: Decision Computation. The most straightforward decision process to compare the action values is to compute the value difference and choose the one with the higher value. We looked for areas in which BOLD activity was correlated with the value difference between the two action values. As any difference in values can be computed by subtracting the lower from the higher value and also by subtracting the higher value from the lower value, and we had no a priori hypothesis for the directionality of the computation, we tested for correlates of both. We did not find any areas where activity was correlated with $V_{chosen} - V_{unchosen}$ at our omnibus statistical threshold of $P < 0.001$ uncorrected. However, we found a strong correlation with $V_{unchosen} - V_{chosen}$ in anterior cingulate cortex, extending dorsally into Brodmann area 9 (dmFC; Fig. 4*A* and Table S2).

To provide a conceptual explanation as to how the brain might implement the value difference computation, we constructed a computational model called the competition difference model (CDM). (Fig. S4). This model is a simple neural network that carries out value comparisons by stochastic mutual inhibition between two populations of neurons: one encoding the value of a hand movement, and one encoding the value of an eye movement. The model takes into account the stochasticity that leads to non-optimal choices in a proportion of the trials, consistent with actual behavior choices. It produces an output that closely resembles but is not identical to, the value comparison regressor used above. To validate the model behaviorally, we compared the performance of the model on subjects' actual choice behavior and found that the model predicted subjects' actual choices as well as the soft-max procedure used with reinforcement learning (Table S3). We then used the output of this model as a parametric regressor in our fMRI analysis instead of the value difference. This model was found to correlate robustly with activity in the same anterior cingulate cortex region identified as correlating with the value difference (Fig. 4*B*). Thus, the model proposed here provides a possible description of the

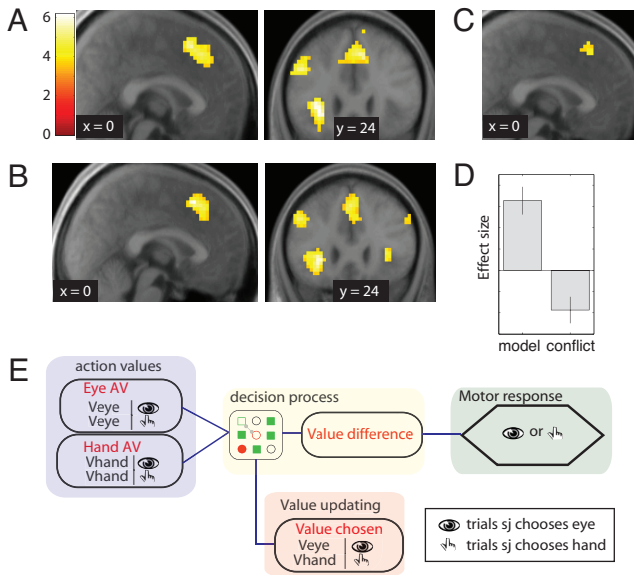


Fig. 4. Value comparison. (A) Region of dmFC and adjacent ACC showing significant correlations with the $V_{\text{unchosen}} - V_{\text{chosen}}$ value difference contrast. Additional areas correlating with this comparison signal are bilateral anterior insula and left dlFC. (B) Output of our stochastic decision model for the value comparison showing correlations with activity in the same brain regions. (C) The model explains activity in dmFC even on a subset of trials where subjects clearly choose the “correct” and not “erroneous” choice (where $V_{\text{chosen}} - V_{\text{unchosen}} > 0.2$). This suggests that the result in (B) cannot be fully explained by error monitoring. (D) Average beta values in the random effects analysis of the model described in the text showing that neural activity in dmFC/ACC is explained better by the output of our decision model than by a decision difficulty based index of decision conflict ($P < 10^{-7}$). The vertical lines, the SEM. (E) Illustration of the different stages involved in action based decision-making: action-based decision-making requires the computation of distinct value representations for both choice alternatives (purple box). These action values are compared against each other in a decision comparator (yellow box) to decide on a particular action. Such a comparator could yield a signal that approximately resembles the difference in the action values of the two actions. The output from this comparator could then be passed through a nonlinear function to inhibit a response of the unchosen action in primary motor areas (green box). The value of the chosen action is used to update future action values on the basis of experience, and to generate prediction errors (red box).

output of a decision comparator, and captures activity related to such a comparison process in ACC.

An important question is the relationship between our suggested role for dmFC/ACC in the decision comparison process and prior findings implicating this area in error monitoring (35). An error-monitoring signal would be strongest on trials where subjects chose the lower valued action and in which there is a large difference between the values of the two available actions (as on those trials it should be most clear to the subjects that they have erroneously chosen the “wrong” action). However, we still find significant correlations between the dmFC signal and our decision model when we restrict the analysis to trials in which the value of the chosen action largely exceeded the value of the unchosen action (Fig. 4C). Another possibility is that subjects were deliberately choosing the lower value action in some trials to explore and anticipated in those deliberate ‘error’ trials a negative outcome. We therefore also looked for regions that had stronger activity during the choice period on trials that were subsequently not rewarded compared to subsequently rewarded trials. Activity in the frontal poles showed such a pattern but not activity in dmFC/rostral ACC. Together this suggests that the decision signal is unlikely to be accounted solely as a side effect of error monitoring.

Another pertinent issue is the extent to which the activity in

dmFC/ACC is related to conflict monitoring, another cognitive function that has been attributed to this area (36). To compare these explanations we constructed a measure of decision conflict by testing for areas showing a maximal response when the action values are the same, and a minimal response when they are maximally different. We found that activity in rostral dmFC is significantly better explained by the decision signal than by this simple decision conflict signal (Fig. 4D), and this is true even if the measure of decision conflict includes subject specific biases to either eye or hand movements. To further address this point, we also tested for correlations of reaction time with decision difficulty, but did not observe any such influences ($r = 0.01$ across all subjects).

Discussion

Action based decision-making involves different kinds of value signals that play specific roles in the various stages of the decision process (Fig. 4E). Action values are by definition precursors of choice that are used to guide the decision process. Here we provide evidence that action-values for different physical actions are present in the supplementary motor area. These value signals are not modulated by choice, that is, they are present for a given action on trials when that action is chosen and on trials when that action is not chosen. We found neural correlates for action values in supplementary motor cortex, an area traditionally associated with motor planning (37). This finding supports the hypothesis that during decisions involving motor actions, action-value signals are encoded in brain regions directly connected with, and involved in, the generation of motor output (38). This finding is broadly consistent with a number of previous studies that have investigated the role of the motor system in decision making. For example, two studies (25, 26) have found monkey medial premotor cortex involved in the entire discrimination process between haptic stimuli, and the findings of another study (24) suggest that formation of the decision and formation of the behavioral response share a common level of neural organization.

In contrast to the supplementary motor cortex, activity in both the vmPFC and intraparietal sulcus pertained predominantly to the value of the action that is chosen. Such signals are a consequence of the decision process, emerging after the subject has decided which action will ultimately be taken. We suggest that the functional contribution of such signals to the decision process is likely not in guiding choice directly, but rather in learning the action values. Reinforcement learning theory stipulates that updating of action values occurs via a prediction error, which computes the difference between actual and expected outcomes (2, 39, 40). A major function of the chosen value signals in these areas could be to facilitate the generation of a prediction error signal that can then be used to update future action values. It is notable that the two signals required to compute a prediction error, namely the actual outcome and the expected outcome (of the chosen action) are both represented in ventromedial prefrontal cortex (8, 9, 41, 42). Therefore this region is ideally placed to facilitate computation of the prediction error signal that could then be transferred on to dopaminergic neurons in the midbrain for subsequent broadcast (43, 44). Another intriguing feature of our results is that we observed a number of different types of chosen value signals within vmPFC. While one region of vmPFC was responsive to the value of the chosen action irrespective of whether that action was a hand or an eye, distinct regions more posterior within vmPFC appear to be sensitive to action specific chosen values. These findings provide evidence that values of different types of movement might be represented separately within vmPFC, adding further support to the suggestion that this region plays a role in encoding the value of chosen actions as well as possibly contributing to encoding stimulus-values (45). The apparent topographical arrangement of action modality specific value signals within vmPFC may relate to distinct cortico-striatal loops concerned with processing hand and eye movements (46).

Our results also suggest that the dmFC/ACC plays a role in the decision process. Interestingly, this area has been previously implicated in action-based choice: in the context of a human neuroimaging study reporting activity in this area during a task involving choices between different actions compared to a situation involving responses guided by instruction (27) and single neuron recordings have shown that cells in this area were activated only by particular action-reward combinations (47). Another study suggests that this region plays a part in processing the reward information for motor selection (48). Consistent with our findings, Seo and Lee (22) found neural signals resembling the difference between action values in this region. In addition, ACC lesions have been shown to produce an impairment in action outcome-based choice, but not in mediating changes in responses following errors (28, 49). Our results provide evidence that these deficits might be the results of impairments in the mechanisms in ACC/dmFC responsible for comparing action values. Heekeren et al. (50, 51) [see also (52)] have used fMRI to look for regions that might be involved in computing perceptual decisions. They found evidence that activity in left dorso-lateral PFC encodes a value signal that is proportional to the absolute value difference between the two signals, while our value difference related signal is represented in dorso-medial PFC. The specific form of the comparison signal we found in ACC was well captured by a simple network model that we called the CDM. This model relies on a mutual inhibitory competition between distinct populations of neurons representing eye and hand movements to generate a decision. Although it bears a conceptual relationship to many other models used to generate decisions such as the drift diffusion model (DDM) (Fig. S5) (53–55), the predictions of the CDM and the DDM model are in fact very different (see *SI Text* for more details). Indeed, while the CDM model provides a good account for the comparison signal we observed in ACC, the DDM model fails to capture such an output. Because our study was not designed to address the presence of DDM-related signals we cannot rule out the contribution of such computations to the decision process. However, it is worth noting that while there is now considerable evidence concerning the applicability of the DDM model to the neural mechanisms underlying decision making in the perceptual domain (56, 57), to our knowledge very little evidence exists regarding the applicability of such a model to value-based decision making. Thus, it is possible that these two different types of decisions rely on distinct computational processes.

Interestingly, the signal reflecting the output of the action-value comparator represented the difference between the action not chosen and the action chosen, instead of the more intuitive difference given by the action chosen minus the action not chosen. A speculative interpretation for this finding is that the outcome of the comparator process is used to inhibit the opposite action, instead of exciting the motor plan that it represents. Interestingly, our activation pattern looks very similar to one found in a study of volitional motor inhibition (58). Such a mechanism based on preinnervation and inhibition could provide a better execution speed after the values become available compared to a mechanism where the motor response is planned only after the decision has been made. Although we cannot distinguish between excitatory and inhibitory processes based on the measured BOLD (59), our hypothesis resonates with previous findings that preinnervation and inhibition play an important role in motor execution and volitional action initiation (60, 61).

Since activity in ACC/dmFC has been associated with error monitoring and conflict detection in previous studies we carried out several controls to help exclude the possibility that the activity we observed in this area can be explained by these alternative computations. We emphasize that our results don't rule out a contribution of ACC to either conflict or error monitoring, but rather suggest that these explanations are unlikely to account fully for the results we observe here. Instead we provide a mechanistic account for how action comparison signals in ACC/dmFC could form an

integrated part of the decision process. Note that because of limitations in the spatial and temporal resolution of our fMRI signal it is not possible to determine whether the signal we observe reflects solely the output of a decision comparator or whether the dmFC/ACC is involved in the comparison process itself. Therefore the possibility exists that the actual computation of the decision is carried out elsewhere and the output then transferred to dmFC/ACC.

Choices between different physical actions, such as those studied here, represent a large subset of the decisions made by humans and other animals. The present study has identified neural mechanisms involved in these types of choices and provides insight into the general neural mechanism that might be involved in action-based decision making. An important question for future studies is whether similar mechanisms are at play when goal-directed decisions are made between more abstract choices not tied to specific physical actions.

Methods

Subjects. Twenty-three healthy subjects [10 female; 18–29 years old; right-handed, assessed by self-report with an adapted version of the Edinburgh handedness inventory (62)] with no history of neurological or psychiatric illness participated in the study. The study was approved by the Institutional Review Board of the California Institute of Technology.

Experimental Design and Task. The task is a variant of a two-armed bandit problem in which subjects chose between two actions: a button press with the right index finger, and a saccade from a central fixation cross to a target located at 10° of visual angle in the right hemifield. In every trial each action yielded either a prize of 10 cents, or nothing. We did not reveal the exact reward per trial to subjects before the experiment but instead instructed them only that they will get a small amount of money for each rewarded trial. At the end of the experiment, subjects were paid their accumulated earnings in addition to a flat amount of \$25.

The probability ($Q_{i,t}$) of action i being rewarded in trial t evolved over time as a decaying Gaussian random walk process, with $Q_{i,t+1} = \max\{0, \min[1, \lambda Q_{i,t} + (1 - \lambda)\theta + \nu]\}$; where the decay parameter λ was 0.79836, the decay center θ was 0.50, and the diffusion noise ν was zero-mean Gaussian with standard deviation $\sigma_d = 0.208$. Five different probability trajectories were generated using this method and were assigned across subjects randomly. The task consisted of two sessions of 150 trials separated by a short break. There were three trial types. In free-choice trials (150 trials) the subject had to choose one of the two actions and both were rewarded according to their current reward schedule. Free-choice trials were pseudorandomly interspersed with forced-choice trials (50 eye trials and 50 hand trials) and null-choice trials (50 trials). Subjects were instructed that in forced trials only the displayed action would be rewarded with its current probability, while the other action would lead to a zero prize with certainty. Subjects did not get a prize in null-trials, but were still required to make a choice.

The task was presented via back projection on a translucent screen, viewable through a headcoil mounted mirror. Subjects chose the hand action by pressing a button on a button box with their right index finger. Eye positions were monitored at 120 Hz with a long-range infrared eye-tracking device (ASL Model L6 with control unit ASL 6000, Applied Science Laboratories). An eye action during the choice period was registered when the median horizontal eye coordinate during the past 200 ms exceeded 8° of visual angle to the right from fixation. Subjects were instructed to maintain fixation during the entire experiment when not deliberately making a saccade.

Reinforcement Learning (RL) Model. A RL model was used to estimate the value that the brain assigned to the two actions on the basis of trial-by-trial experience. In this study we used a version of RL called Q-learning, where action values are updated using a simple Rescorla-Wagner rule (see *SI Text* for details).

Computational Model of the Choice Process (Decision Model). We were also interested in identifying brain regions involved in comparing the action values to make decisions. The most basic value comparison process that one could consider involves calculating the difference between the action values to identify and select the largest one. A problem with such a model is that it does not account for the choice stochasticity that is observed in the data, and thus it cannot explain behavior in those trials where subjects chose the action with the lower action value. We therefore constructed an extremely simple neural network type model that characterizes the properties of aggregate activity that identify putative decision making regions (Fig. S6). We then use these trial-by-trial predictions as

parametric regressors in our fMRI analysis to identify areas where the value comparison computation might be carried out (see *SI Text* for details).

fMRI Data Acquisition and Analysis. Data were acquired with a 3T scanner (Trio, Siemens) using an eight-channel phased array head coil (see *SI Text* for details).

We estimated two different general linear models with AR (1) for each individual subject (see the *SI Text* for details). In each case we computed contrasts of interest at the individual level using linear combinations of the regressors and, to enable inference at the group level, we calculated second-level group contrasts using a one-sample *t* test.

Whole brain inference was carried out at $P < 0.001$ uncorrected. We also computed small volume correction (SVC) for multiple comparisons at the $P < 0.05$ level in areas or a priori interest (*SI Text*).

1. von Neumann J, Morgenstern O (1944) in *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ).
2. Sutton RS, Barto AG (1998) in *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA).
3. Dayan P, Abbott LF (2001) in *Theoretical Neuroscience* (MIT Press, Cambridge, MA).
4. Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci* 9:545–556.
5. Samejima K, Ueda Y, Doya K, Kimura M (2007) Action value in the striatum and reinforcement-learning model of cortico-basal ganglia network. *Neurosci Res* 58:522.
6. Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
7. Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27:14502–14514.
8. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. *Nature* 441:876–879.
9. Hampton AN, Bossaerts P, O'Doherty JP (2006) The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci* 26:8360–8367.
10. Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science* 301:1104–1107.
11. Hommer DW, et al. (2003) Amygdalar recruitment during anticipation of monetary rewards: An event-related fMRI study. *Ann NY Acad Sci* 985:476–478.
12. Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–1787.
13. Sugrue LP, Corrado GS, Newsome WT (2005) Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat Rev Neurosci* 6:363–375.
14. Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223–226.
15. Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci* 27:9984–9988.
16. 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.
17. Fujii N, Mushiaki H, Tanji J (2002) Distribution of eye- and arm-movement-related neuronal activity in the SEF and in the SMA and Pre-SMA of monkeys. *J Neurophysiol* 87:2158–2166.
18. Campos M, Breznen B, Bernheim K, Andersen RA (2005) Supplementary motor area encodes reward expectancy in eye-movement tasks. *J Neurophysiol* 94:1325–1335.
19. Lau HC, Rogers RD, Ramnani N, Passingham RE (2004) Willed action and attention to the selection of action. *Neuroimage* 21:1407–1415.
20. Thaler D, Chen YC, Nixon PD, Stern CE, Passingham RE (1995) The functions of the medial premotor cortex. I. Simple learned movements. *Exp Brain Res* 102:445–460.
21. Seo H, Barraclough DJ, Lee D (2007) Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb. Cortex* 17:1110–1117.
22. Seo H, Lee D (2007) Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J Neurosci* 27:8366–8377.
23. Cui H, Andersen RA (2007) Posterior parietal cortex encodes autonomously selected motor plans. *Neuron* 56:552–559.
24. Gold JJ, Shadlen MN (2003) The influence of behavioral context on the representation of a perceptual decision in developing oculomotor commands. *J Neurosci* 23:632–651.
25. Romo R, Hernandez A, Zainos A (2004) Neuronal correlates of a perceptual decision in ventral premotor cortex. *Neuron* 41:165–173.
26. Hernandez A, Zainos A, Romo R (2002) Temporal evolution of a decision-making process in medial premotor cortex. *Neuron* 33:959–972.
27. Walton ME, Devlin JT, Rushworth MF (2004) Interactions between decision making and performance monitoring within prefrontal cortex. *Nat Neurosci* 7:1259–1265.
28. Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF (2006) Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9:940–947.
29. Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci* 19:181–189.
30. Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. *European J Neurosci* 22:513–523.
31. O'Doherty J, et al. (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science* 304:452–454.
32. Haruno M, et al. (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.
33. O'Doherty J, Critchley H, Deichmann R, Dolan RJ (2003) Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci* 23:7931–7939.
34. Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. *Neuron* 41:281–292.
35. Carter CS, et al. (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* 280:747–749.
36. Kerns JG, et al. (2004) Anterior cingulate conflict monitoring and adjustments in control. *Science* 303:1023–1026.
37. Hoshi E, Tanji J (2004) Differential roles of neuronal activity in the supplementary and presupplementary motor areas: From information retrieval to motor planning and execution. *J Neurophysiol* 92:3482–3499.
38. Rathelot JA, Strick PL (2006) Muscle representation in the macaque motor cortex: An anatomical perspective. *Proc Natl Acad Sci USA* 103:8257–8262.
39. Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. *Science* 275:1593–1599.
40. Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16:1936–1947.
41. Knutson B, Fong GW, Adams CM, Varner JL, Hommer D (2001) Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport* 12:3683–3687.
42. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C (2001) Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci* 4:95–102.
43. Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47:129–141.
44. Roesch MR, Calu DJ, Schoenbaum G (2007) Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10:1615–1624.
45. Glascher J, Hampton AN, O'Doherty JP (2008) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cereb Cortex* 19:483–495.
46. Gerardin E, et al. (2003) Foot, hand, face and eye representation in the human striatum. *Cerebral Cortex* 13:162–169.
47. Matsumoto K, Suzuki W, Tanaka K (2003) Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301:229–232.
48. Shima K, Tanji J (1998) Role for cingulate motor area cells in voluntary movement selection based on reward. *Science* 282:1335–1338.
49. Hadland KA, Rushworth MFS, Gaffan D, Passingham RE (2003) The anterior cingulate and reward-guided selection of actions. *J Neurophysiol* 89:1161–1164.
50. Heekeren HR, Marrett S, Ruff DA, Bandettini PA, Ungerleider LG (2006) Involvement of human left dorsolateral prefrontal cortex in perceptual decision making is independent of response modality. *Proc Natl Acad Sci USA* 103:10023–10028.
51. Heekeren HR, Marrett S, Ungerleider LG (2008) The neural systems that mediate human perceptual decision making. *Nat Rev Neurosci* 9:467–479.
52. Rorie AE, Newsome WT (2005) A general mechanism for decision-making in the human brain? *Trends Cogn Sci* 9:41–43.
53. Usher M, McClelland JL (2001) The time course of perceptual choice: The leaky, competing accumulator model. *Psychol Rev* 108:550–592.
54. Smith PL, Ratcliff R (2004) Psychology and neurobiology of simple decisions. *Trends Neurosci* 27:161–168.
55. Busemeyer JR, Johnson JG (2004) Computational models of decision making. In *Handbook of Judgement and Decision Making*, eds Koehler D, Narvey N (Blackwell Publishing Co, New York), pp 133–154.
56. Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86:1916–1936.
57. Shadlen MN, Britten KH, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16:1486–1510.
58. Brass M, Haggard P (2007) To do or not to do: The neural signature of self-control. *J Neurosci* 27:9141–9145.
59. Logothetis NK (2007) The ins and outs of fMRI signals. *Nat Neurosci* 10:1230–1232.
60. Libet B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav Brain Sci* 8:529–566.
61. Sumner P, et al. (2007) Human medial frontal cortex mediates unconscious inhibition of voluntary action. *Neuron* 54:697–711.
62. Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
63. Duvernoy HM (1999) in *The Human Brain. Surface, Blood Supply and Three-Dimensional Section Anatomy* (Springer, New York), 2nd Ed.