

# Economic choices can be made using only stimulus values

Klaus Wunderlich<sup>a,b,1</sup>, Antonio Rangel<sup>b,c</sup>, and John P. O'Doherty<sup>b,c,d</sup>

<sup>a</sup>Wellcome Trust Center for Neuroimaging, University College London, London WC1N 3AR, United Kingdom; <sup>b</sup>Computation and Neural Systems Program and <sup>c</sup>Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125; and <sup>d</sup>Trinity College Institute of Neuroscience and School of Psychology, Dublin 2, Ireland

Edited by Ranulfo Romo, Universidad Nacional Autonoma de Mexico, Mexico D.F., Mexico, and approved July 15, 2010 (received for review February 22, 2010)

**Decision-making often involves choices between different stimuli, each of which is associated with a different physical action. A growing consensus suggests that the brain makes such decisions by assigning a value to each available option and then comparing them to make a choice. An open question in decision neuroscience is whether the brain computes these choices by comparing the values of stimuli directly in goods space or instead by first assigning values to the associated actions and then making a choice over actions. We used a functional MRI paradigm in which human subjects made choices between different stimuli with and without knowledge of the actions required to obtain the different stimuli. We found neural correlates of the value of the chosen stimulus (a postdecision signal) in ventromedial prefrontal cortex before the actual stimulus-action pairing was revealed. These findings provide support for the hypothesis that the brain is capable of making choices in the space of goods without first transferring values into action space.**

decision-making | functional MRI | goods space | reinforcement learning | ventromedial prefrontal cortex

Imagine you are thirsty and walk up to a vending machine that is serving a variety of soft drink beverages. On the machine you see the brand marks of the offered beverages, and because you had previously sampled them, you easily assign values to each drink based on their taste. To get the desired beverage, you press the button that is distinctively associated with the preferred option. This situation exemplifies many of the decisions that humans and animals make in daily life. It is a well-established belief among economists, psychologists, and neuroscientists that the brain solves such choice problems by first computing a value for each alternative and then selecting the one that has the highest value (1–4). Neuroscientists have considered two possible alternative ways for how values might be compared in order to make a choice in these situations.

First is the actions-based model, in which choices are embedded in premotor processes of action selection: the values of stimuli are passed as action values to the motor plans required to obtain them, and the decision is then made in action space by comparing action values (5–7). Values are learned through experience and associated with each motor plan, and during choice a single motor act eventually emerges through a winner-takes-all process. Although this action based model of making decisions may seem convoluted, it is in fact the predominant view among many decision neuroscientists who have found value signals in areas of the brain known to be involved in representing and planning movements such as lateral parietal and premotor cortices (8–11). Further evidence for the view that decisions are computed by a comparison between actions comes from the finding of action value signals in several regions of the brain, including the caudate nucleus (5, 6), supplementary motor cortex (12), and action-related value signals in lateral intraparietal cortex (9, 13). Additionally, the action-based model is sometimes presented as a more general psychological model of behavior because it builds more or less directly on theories of reinforcement learning (RL) and seems to provide a flexible and adaptable unitary model for universal

problem solving. This model has its origins in Thorndike's law of effect (14), describing the behavior of the animal in terms of stimulus and motor response associations, and more recently, several influential models of decision-making can be viewed as variations of the action-based model (15, 16).

Second is the goods-based model, in which the values of the available stimuli are compared directly to make a choice without taking into account any action information (i.e., the choice takes place in goods space), and only after a stimulus is chosen are the necessary motor plans identified and executed (17, 18). This view proposes a sequential choice process in which action selection is temporally separated from the actual process of choice. There is evidence consistent with the presence of abstract goods-based value representations in orbitofrontal cortex (17, 19–22), which could be used as the input to such a choice process. Furthermore, lesion studies indicate double dissociations between the effects of lesions of orbitofrontal cortex and anterior cingulate cortex on learning of stimulus-reward and action-reward associations, respectively (23, 24).

Note, however, that although there is considerable evidence for goal values and stimulus-based learning, it is as yet unknown whether such signals are used to compute choices directly, or whether by contrast such signals need to be converted into action space before choice signals can be computed.

The aim of the present study was to directly address this question. We used functional MRI (fMRI) in human subjects while they performed a variant of a two-armed bandit task to obtain probabilistically delivered monetary rewards (Fig. 1A). In every trial, subjects made a choice between two stimuli and selected one by executing the action that was randomly paired with the chosen stimulus (either button press or saccade). A critical feature of the task was that, in half of the trials [stimulus condition (SC)], subjects were first presented with the two stimuli alone in a horizontal arrangement that did not contain any information about the actions required to obtain them. The actions were revealed only after a variable interval by randomly flipping the stimuli in vertical alignment. At this stage, subjects could choose the upper stimulus by making a saccade to a target in the right hemifield, and the lower stimulus by pressing a button with their right hand. In the other half of the trials [action condition (AC)], the first screen was not shown and instead the stimuli appeared immediately in the vertical action-pairing position. To avert the possibility that subjects made the decision in the preceding intertrial interval, they were, in every trial, presented with a choice between two of three possible stimuli (triangle, square, and circle) in pseudorandom appearance. Subjects therefore did not know until the trial onset which pair of stimuli would be

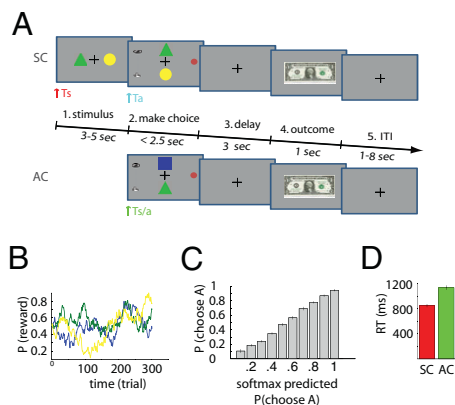
Author contributions: K.W., A.R., and J.P.O. designed research; K.W. performed research; K.W. analyzed data; and K.W., A.R., and J.P.O. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

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

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002258107/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1002258107/-DCSupplemental).



**Fig. 1.** Experimental design and behavior. (A) Subjects were presented with two stimuli (in every trial pseudorandomly selected among three possible stimuli) in horizontal arrangement (screen 1, for a variable time between 3 and 5 s). Stimuli then flipped to a vertical arrangement indicating the actions required to obtain each stimulus (making a saccade or pressing a button, screen 2). When a response had been registered, the screen was immediately cleared for a short delay and subsequently the outcome was revealed (screen 4 at 6 s after screen 2), indicating receipt of reward or no reward. There were two conditions: an SC as just described and an AC in which the first screen was not shown and subjects immediately saw the stimulus–action pairing. Red, green, and cyan arrows indicate onset times for regressors used to estimate the effects reported in Figs. 2 and 3 (which are illustrated using the same color code). (B) Example reward probability paths for the three stimuli as a function of the trial number. The probabilities of being rewarded following choice fluctuated slowly and independently for each stimulus across the experiment. (C) Actual choice probability plotted against fitted model choice probability (binned 0.1 wide), averaged across subjects (lines represent SEM). (D) Reaction time (after the action pairing is revealed in screen 2) is significantly lower in SC trials than in AC trials (paired *t* test,  $P < 10^{-11}$ , vertical lines represent SEM).

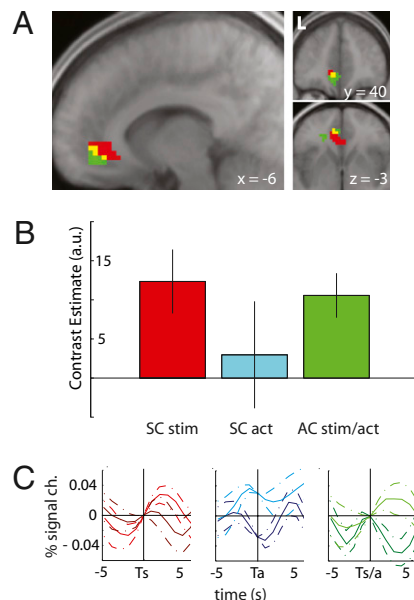
presented. The probability of being rewarded on selecting each of the three stimuli drifted randomly over time and was independent of the probability of being rewarded on the others (Fig. 1B). We estimated the value of taking each stimulus in every trial by calculating the stimulus values using a computational RL model in which the values  $V_{\text{triangle}}$ ,  $V_{\text{square}}$ , and  $V_{\text{circle}}$  were updated in proportion to a prediction error on each trial. The model also assumed that stimulus selection in every trial followed a soft-max probability rule based on the difference of the estimated values, which provided a good description of behavior (Fig. 1C, Fig. S1, and Table S1).

We reasoned that, if choices can be computed in goods space, we should observe neural correlates of the chosen value, a post-decision signal, already at the time of stimulus presentation. Note that the prediction was that these chosen value signals would be observed before any action-related information was made available to the subjects, thus making it unlikely that they would be able to process the choice in the motor system. Based on several previous fMRI studies, we expected to see a neuronal representation of the chosen value in ventromedial prefrontal cortex (vmPFC) (12, 25, 26). If our hypothesis is correct, it will provide unique evidence that the brain can compute choices solely in goods space. We also speculated that subjects would respond faster in SC trials than in AC trials because the time necessary to make a decision between the desired stimuli had already been provided before the action pairing.

## Results

**Response Times.** Consistent with the reaction time hypothesis, we found that subjects responded significantly faster (paired *t* test,  $P < 10^{-11}$ ) in the SC than the AC condition (Fig. 1D).

**Neural Correlates of Chosen Values.** To look for neural correlates of value signals we entered the trial-by-trial estimates of the values of the two stimuli under consideration into a regression analysis against the fMRI data. We found that, in SC trials, neural activity in vmPFC ( $x = -3, y = 27, z = -9; T = 3.54$ ) correlated with the value of the stimulus that is subsequently chosen before the stimulus–action pairing was revealed (Fig. 2A and Table S2). We also tested for a value-chosen signal in SC trials at the time of action pairing but did not find any significant correlation for this contrast at our omnibus threshold. In AC trials, the vmPFC ( $x = -6, y = 39, z = -12; T = 5.74$ ) also correlated significantly with the value-chosen signal. Although the peak of the area encoding the chosen value in AC trials was found to be located slightly more anterior and ventral than the peak in SC trials (Fig. 2A and Table S3), a further interaction contrast showed that no area correlated with chosen value more strongly in either the SC or AC condition, even at a liberal threshold of  $P < 0.005$  (uncorrected). Effect size plots (Fig. 2B) and time course plots (Fig. 2C) in the overlapping area (center at  $x = -9, y = 42, z = -3$ ) confirmed that, in SC trials, activity in vmPFC was correlated with the chosen value only at the time of stimulus presentation but not at the succeeding time of stimulus–action pairing. Time courses separated according to the value signal in SC trials already at the time of the stimulus presentation, which preceded the action–pairing screen by 3 to 5 s. All reported activations are significant after family wise error correction for multiple com-



**Fig. 2.** Neural correlates of chosen value. (A) Activity in vmPFC showed significant correlation with the value of the stimulus that was subsequently chosen before the stimulus–action pairing was revealed (SC trials, red). The value chosen signal in AC trials was represented slightly more ventrally (green). Voxels overlapping in both contrasts are shown in yellow. Activations survive correction for multiple comparisons as described in *Methods*. (B) A comparison of effect size at the overlapping region confirms that, in SC trials, the value chosen is represented only at the time of the stimulus screen (red) but not at the time of the following stimulus–action pairing (cyan). In AC trials, the value chosen is represented at the coinciding stimulus/action screen (green). Bars indicate SEM. (C) Event-related blood oxygen level-dependent responses in SC trials time locked to the stimulus presentation (Left), the stimulus–action pairing (Center), and in AC trials time locked to the coinciding stimulus–action pairing (Right). Time courses are plotted separately for trials in which the chosen values were small (dashed,  $V < 0.5$ ) and large (solid,  $V > 0.5$ ). Note that, consistent with the effect sizes shown in B, time courses split in SC trials after stimulus presentation (Ts, Left) but already approximately 4 s before the action pairing (Ta, Center). Dashed lines represent SEM.

parisons at the cluster level. Data used to calculate effect sizes were independent of the data used in the functional definition of the region of interest.

An important concern with this result is that the identified correlations with the chosen value signals might reflect the encoding of other prechoice signals that correlate with the chosen value signal. One variable of special concern is the average stimulus value, which is positively correlated with the chosen value. We addressed this concern in two ways. First, we ran a separate generalized linear model (GLM) in which activity at the time of stimulus evaluation in SC trials was modulated by the average stimulus value and by the chosen value. Importantly, to avoid any spurious correlation, the chosen value modulator was orthogonalized with respect to the average stimulus value signal. The activation in this contrast was observed at the same cluster as in the original design (peak,  $-9, 33, -3$ ;  $T = 2.95$ ;  $P < 0.004$ ), indicating that the chosen value signal accounts for variance in vmPFC activity over and above that accounted for by stimulus value. Second, we compared the signals from value chosen with both the average stimulus values and the maximum stimulus values within vmPFC to determine the relative probability with which these signals explain the measured neural signal in every subject. We used a Bayesian model comparison approach (27) to calculate posterior model probabilities and compared the likelihoods of the three value signals. The advantage of this method is that it circumvents any collinearity issues occurring from correlated regressors in the GLM. Consistent with our previous results, we found that, out of the three value signals, value chosen was the most likely underlying explanatory variable for activity in a cluster in vmPFC ( $-3, 27, -9$ ). Using the Bayesian method, we compared all three value signals simultaneously as well as pair-wise chosen value with average stimulus value and maximum stimulus value. The exceedance probability in this analysis reflects the likelihood that one model explains the neural data better than the other models in the comparison (Table S4). The exceedance probability of the value chosen model is larger than 95% in both of our pairwise comparisons. Note that the sum of the stimulus values is identical to the average stimulus value after mean correction, and a comparison of value chosen with the sum of stimulus values would yield the same outcome. To further rule out the possibility that the activation is merely stemming from an unequal representation of the two options in vmPFC before choice, we also directly compared the value of the chosen option to the value of the option not chosen in a separate analysis. If the activation in the vmPFC cluster were caused by some form of representation of both value options, we would see in this contrast some variance loading on both regressors. We found significant activation for value chosen (peak at  $-3, 27, -12$ ;  $T = 3.89$ ;  $P < 0.0005$ ) similar to our original design, but no brain areas correlating significantly with the value of the unchosen option at a liberal threshold of  $P < 0.005$  uncorrected. We also tested for the difference between value chosen and the value not chosen at the contrast level and did not find any significant correlation at  $P < 0.005$  (uncorrected).

**Neural Correlates of Stimulus Values.** We then looked for representations of the individual stimulus values, because clearly such signals should be a precursor of choice in that these values need to be compared in order to work out which option is ultimately chosen. In view of the spatial limitations of fMRI, we assumed it would be unlikely to detect activity patterns encoding the value of the individual stimuli. Instead, we assumed that neurons encoding such values would be spatially intermixed within the same region, and that such intermixed neural signals would be reflected at the level of the blood oxygen level-dependent signal as an average of the values of the two stimuli under consideration on a given trial. Consistent with our hypothesis, we found such an averaged stimulus value signal in SC trials within a subregion of vmPFC ( $x = -9, y = 48, z = -3$ ;  $T = 4.41$ ; Fig. 3 and Table S5).

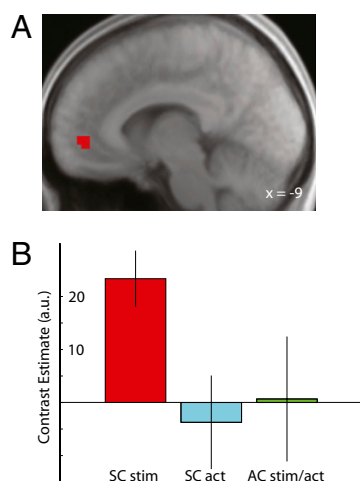
We also performed an analogous Bayesian model comparison for pre- and postdecision values at this vmPFC location, and activity in this subcluster was explained best by the average stimulus value (Table S4). Furthermore, consistent with the notion that individual stimulus values are only required at the time of decision-making, we found evidence for an averaged stimulus value signal only at the time of stimulus presentation (and not after the action pairings had been presented).

**Neural Correlates of Action.** A natural question is whether action values are still computed after a choice is made in goods space in the SC condition, and if they are computed in the AC condition. To pursue the possibility that stimulus values are still transferred to the motor system for action selection, we looked for neural representations of action value signals (5, 12, 28). We did not find any area correlating with the value of eye or hand movements in either SC or AC conditions at a liberal threshold of  $P < 0.005$  (uncorrected).

## Discussion

Our results show that there is activity in vmPFC that correlates with the value of the stimulus that is chosen in a trial, a post-decision signal, before the action pairing was revealed. Because there exists no other way to represent the value of the option that will eventually be selected in a trial without prior value comparison that converged to this particular option, these findings provide support for the hypothesis that the brain is capable of computing choices completely within an abstract representation of stimuli.

One possible alternative explanation of our findings is that, in the SC, subjects make decisions by assigning temporary action pairings to the stimuli before the action pairings are revealed, and then compare the temporary actions. These temporary action pairings could then be substituted for the real action assignments at the time of action presentation. Although we cannot completely rule out these explanations on the basis of our data alone, there are several reasons why this type of explanation is unlikely to account for our results. First, when subjects are in a situation in which it is necessary to make decisions over actions (i.e., when there is no unique stimulus information to discriminate different options), regions of the brain known to be involved



**Fig. 3.** Neural correlates of stimulus value. (A) Activity in vmPFC showed significant correlation with the average value of the two stimuli that were presented in SC trials at the time of screen 1 (red). Activations survive correction for multiple comparisons as described in *Methods*. (B) Comparison of effect sizes at the peak region between conditions. Neural activity in this region correlated with stimulus values in SC trials but not in AC trials. Bars indicate SEM.

in motor planning and initiation such as supplementary motor cortex, lateral intraparietal cortex, and anterior cingulate cortex have been found to contain action-related value signals in earlier imaging studies (12, 29). However, these regions did not show significant correlations with action value signals in the present paradigm even at a liberal uncorrected threshold ( $P < 0.005$ ), suggesting that neural systems involved in action representations were not directly engaged during the decision process in the present study. Second, on a more conceptual level, although encoding of conditional action pairings might be feasible in the present simplified experimental paradigm, such a mechanism is unlikely to scale well in many real-world decision problems with large numbers of sequential conditional action pairings, because decisions in such contexts would require encoding of long strings of conditional action pairings that could rapidly become computationally intractable. By contrast, the parsimonious alternative proposed here, whereby in such contexts a decision is made between the stimuli, would not suffer from the same scaling problem.

We did not find action value signals (5, 12, 28) in the SC or AC conditions even at liberal thresholds. A possible explanation for the lack of action value signals could be that, in contrast to previous paradigms, in the current study subjects did not have to keep track of action values over trials and doing so would have not improved performance. Although a negative finding should not be interpreted as definitive evidence against the presence of action values, it supports the idea that, in this particular task, the choices were being implemented in goods space in both the SC and AC conditions.

Interestingly, we found a representation of the stimulus values only in the SC but not in the AC condition. Although we cannot rule out the possibility that subjects indeed compute choices differently in the two conditions, it is conceivable that the stimulus value signals are very transiently represented only while the options are under consideration. Whereas in SC trials, the stimuli screen is shown for 3 to 5 s during which subjects could consider the options, they are allowed an immediate response in AC trials, allowing the value comparison process to reach convergence very quickly. With respect to recent hypotheses that ventromedial frontal cortex might be comparing the values of possible choices, discarding the representation of the unchosen option and maintaining the representation of the chosen one (30, 31), it would be interesting to know whether the stimulus values in the ventromedial frontal cortex temporally precede the presence of a signal related to the value of the chosen stimulus in the adjacent ventromedial frontal cortical region. Although the temporal resolution of fMRI makes it very challenging to answer this question, other techniques such as EEG or single-unit neurophysiology could be deployed in the future to resolve this issue.

It is important to emphasize that our data do not imply that all decisions are made in goods space, but rather that the brain is capable of computing a decision purely in goods space when action pairings are not available. There is ample experimental evidence that behavioral decisions can be and are made over actions in many contexts (12, 29, 32–35). Given the results presented here and taken alongside evidence for the presence of action value signals in caudate (5, 6) and supplementary motor system (12), it is natural to conjecture that both mechanisms may coexist during certain types of choices. In fact, this is proposed in some associative models of goal-directed behavior, which posit that choice can occur over outcomes that then subsequently retrieve actions, or over actions that subsequently retrieve outcomes (36, 37).

## Methods

**Subjects.** A total of 24 healthy subjects [age 18–31 y; all right-handed as assessed by self-report with an adapted version of the Edinburgh handedness inventory (38)] with no history of neurological or psychiatric illness

participated in the study. The study was approved by the Institutional Review Panel of the California Institute of Technology.

**Task.** The task is a variant of a two-armed bandit problem in which subjects make pair-wise choices between subsets of two stimuli that were pseudorandomly selected among three stimuli used in the experiment: a green triangle, a blue square, and a yellow circle.

There were two conditions presented pseudorandomly intermixed during the experiment. In the first one (SC), subjects were initially presented with the stimuli in horizontal arrangement without the information of what action they had to perform to choose a stimulus. After a variable time (3, 4, or 5 s, uniform distribution), the stimuli flipped to vertical position that indicated the action associated with each stimulus. The assignment of stimuli to actions was made randomly in every trial. At this stage, subjects could press a button with their right index finger to choose the bottom stimulus or perform a saccade from a central fixation cross to a target located at  $10^\circ$  of visual angle in the right hemifield to choose the top stimulus.

In the second condition (AC), the trials were identical except that the first screen was not shown and subjects were immediately presented with the stimuli in vertical arrangement at the beginning of the trial.

The probability ( $Q_i, t$ ) of stimulus  $i$  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  $\lambda$  was 0.8, the decay center  $\theta$  was 0.50, and the diffusion noise  $\nu$  was zero-mean Gaussian with an SD  $\sigma$  of 0.2. Two different probability trajectories were generated by using this method and were assigned across subjects randomly. Fig. 1B depicts one of the probability paths used in the experiment. An important feature of this design is that the probability of being rewarded on one of the three stimuli is independent of the probability of being rewarded on the others. This feature is useful because it implies that the RL-based estimates of the stimulus values are uncorrelated with each other, which increases our ability to dissociate the neural correlates.

The task consisted of four sessions of 75 trials each, separated by a short break. Subjects had to select an action within 2.5 s after onset of the stimulus-action pairing screen; otherwise the trial was counted as an invalid missed trial. Subjects very rarely failed to make a response within this time window: none of the subjects had more than two such events during the entire experiment, and most subjects did not miss any trials at all. After the response was registered, the screen changed to a fixation cross until 6 s after trial onset. At this time the outcome was displayed for 1 s by showing an image of a dollar bill in rewarded trials or a scrambled dollar bill in nonrewarded trials. Trials were separated by a fixation cross that lasted between 1 and 8 s (uniform distribution).

Before the experiment, subjects received full instructions about the task and the two conditions, they were informed that the probabilities of being rewarded on each stimulus changed as a continuous function over time (but were not given details about the underlying stochastic process), and they were instructed to try to maximize their earnings, which were paid to them at the end of the experiment. Subjects accumulated \$0.25 in each rewarded trial. 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 \$20.

The task was presented to the subjects via back projection on a translucent screen, viewable through a head coil-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  $8^\circ$  of visual angle to the right from fixation. Subjects were instructed to maintain central fixation during the entire experiment when not deliberately making a saccade.

**RL Model.** RL is concerned with learning values in different states of the world in a model-free environment in which subjects do not have complete knowledge about the underlying reward generating process. Thus, it is ideally suited to model how subjects learned the value of choosing the different stimuli over time.

In this study, we used RL learning, whereby stimulus values are updated using a simple Rescorla-Wagner rule. If a stimulus is not selected in a trial its value is not updated. In contrast, if stimuli  $S_1$  and  $S_2$  are shown and  $S_1$  is selected on trial  $t$ , its value is updated via a prediction error,  $\delta$ , as follows:

$$V_{s1}(t+1) = V_{s1}(t) + \eta\delta(t), \quad [1]$$

where  $\eta$  is a learning rate between 0 and 1. The prediction error is given by  $\delta(t) = r(t) - V_{s1}(t)$ . The delivered probabilistic rewards were used in updating value predictions of S1 in the RL model. The value of the stimulus that was not chosen (S2) or not presented in a trial (S3), i.e., those stimuli on which subjects did not receive feedback, were allowed to decay toward the mean (13) and updated as follows:

$$V_{s2/3}(t+1) = V_{s2/3}(t) - \tau(V_{s2/3}(t) - 0.5), \quad [2]$$

where the parameter  $\tau$  is the decay rate, accounting for possible forgetting of past experiences (leaky integration). To generate choices, we first used a soft-max procedure in which, in every trial, the probability (P) of choosing stimulus s1 is given by the following:

$$P_{s1,t} = \sigma(\beta(V_{s1}(t) - V_{s2}(t)) - \alpha) \quad [3]$$

where  $\sigma(z) = 1/(1 + e^{-z})$  is the Luce choice rule or logistic sigmoid,  $\alpha = 0$  denotes the indecision point (at which both actions are selected with equal probability), and  $\beta$  determines the degree of stochasticity involved in making decisions.

The model decision probabilities  $P_{s1}$  and  $P_{s2}$  were fitted against the discrete behavioral data  $B_{s1}$  and  $B_{s2}$  to estimate the free parameters ( $\eta$ ,  $\beta$ , and  $\tau$ ). This was done using maximum likelihood estimation subject by subject. The associated likelihood function is given by

$$\log L = \frac{\sum B_{s1} \log P_{s1}}{N_{s1}} + \frac{\sum B_{s2} \log P_{s2}}{N_{s2}}, \quad [4]$$

where  $N_{s1}$  and  $N_{s2}$  denote, respectively, the number of trials in which S1 and S2 were chosen, and  $B_{s1}$  ( $B_{s2}$ ) equals 1 if S1 (S2) was chosen in that trial, and 0 otherwise. We fitted this function similarly for the other two stimulus combinations (S1/S3 and S2/S3) and found the optimal parameters by minimizing the sum of the three negative log-likelihoods.

We compared the choice probabilities predicted by the RL model using the soft-max procedure to subjects' behavior by binning P into 10 bins (bin size of 0.1) and calculating for each bin the fraction of trials in which subjects chose one stimulus. For this test we pooled data from the three sigmoid choice probability functions. To test the fit between the model and the behavioral data, we performed a linear regression, subject by subject, of the fraction of choices on the binned choice probability versus the predicted bin. Overall, the regression results suggest that the model captures actual action value estimation and choice behavior well: on average, the model could explain 92% of the variance in the actual choice data (based on R<sup>2</sup>; Fig. 1C, Fig. S1, and Table S1). We also tested for any structure in the residuals of the regression by looking at the autocorrelations, as any serial correlation in the residuals would mean that there is room for improvement in the model. We have no evidence for such a violation of independence in our data. For every individual subject, the autocorrelations at all possible lags were inside the 95% CI for a stochastic process.

**fMRI Data Acquisition.** Data were acquired with a 3-T scanner (Trio; Siemens) using an eight-channel phased-array head coil. Functional images were taken with a gradient-echo T2\*-weighted echo-planar sequence (repetition time, 2.65 s; flip angle, 90°; echo time, 30 ms; 64 × 64 matrix). Whole brain coverage was achieved by taking 45 slices (3 mm thickness, no gap, in-plane resolution 3 × 3 mm), tilted in an oblique orientation at 30° to the anterior commissure/posterior commissure line to minimize signal dropout in the orbitofrontal cortex. Subjects' heads were restrained with foam pads to limit head movement during acquisition. Functional imaging data were acquired in four separate 370-volume runs, each lasting approximately 16 min. A high-resolution T1-weighted anatomical scan of the whole brain (magnetization-prepared rapid acquisition with gradient echo sequence, 1 × 1 × 1 mm resolution) was also acquired for each subject.

**fMRI Data Analysis.** Image analysis was performed using SPM5 (Institute of Neurology, Wellcome Department of Imaging Neuroscience, London, United Kingdom). Images were first slice time corrected to TR/2, realigned to the first volume to correct for subject motion, spatially normalized to a standard T2\* template with a voxel size of 3 mm, and spatially smoothed with a Gaussian kernel of 8 mm full-width at half-maximum. Intensity normalization and high-pass temporal filtering (using a filter width of 128 s) were also applied to the data.

First, we estimated a GLM with AR(1) for each individual subject. The following events were modeled in each trial: (i) The time of the stimulus presentation in SC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$ ; (ii) the time of the stimulus-action pairing in SC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$ ; (iii) the time of the stimulus-action pairing in AC trials, parametrically modulated by the trial-by-trial stimulus values  $V_{s1}$  and  $V_{s2}$ ; and (iv) the time of the presentation of the outcome, modulated by the prediction error  $\delta$  and a binary function encoding whether a reward was given. Trials in which subjects chose the eye action and trials in which subjects chose the hand action were modeled in separate regressors. All regressors were convolved with the canonical hemodynamic response function. In addition, the six scan-to-scan motion parameters produced during realignment and session constants were included as additional regressors of no interest.

Second, we computed contrasts of interest at the individual level using linear combinations of the regressors described earlier. Finally, to enable inference at the group level, we calculated second-level group contrasts using one-sample *t* tests.

In addition to the GLM described earlier, we estimated four additional GLMs, which were in most parts identical to the first GLM except for the following difference in parametric modulation: at the time of stimulus presentation in SC condition, the first parametric modulator was (i) the sum of both stimulus values, (ii) the maximum of the stimulus values, (iii) the difference of the stimulus values, and (iv) the value of the option not chosen. The second parametric modulator was the value of the stimulus that was later chosen in the trial. This second value chosen modulator was orthogonalized toward the stimulus value parameter. By doing this, we assured that any shared variance between value chosen and stimulus value was assigned to the stimulus value regressor. Therefore, any remaining variance captured by the value chosen regressor does not contain any confound from the stimulus value signal.

We used a Bayesian model comparison procedure (27) to determine which value signal (value chosen or stimulus value) better explained the neural activity in vmPFC. The details of this method are described in *SI Methods*.

Statistics were corrected for multiple comparisons by using a family wise error correction. For this we set a height threshold (at  $P < 0.005$ ) in combination with an extent threshold based on the number of contiguous voxels located in a cluster (39). We used AlphaSim in the Analysis of Functional NeuroImages open source to run a Monte Carlo simulation to determine the probability of contiguous cluster formations under the null hypothesis at the height threshold specified (40). AlphaSim generates an estimate of overall cluster size significance level by iteration of the process of random image generation, Gaussian filtering to simulate voxel smoothness, thresholding, image masking, and tabulation of cluster size frequencies. In our simulation we generated a series of 10,000 random images, each having 56,401 spatially uncorrelated voxels (the number of voxels in masked EPI images) by filling the masked brain volume with independent normal random numbers. The effect of voxel correlation was simulated by convolving the random image with a Gaussian function of the size of our smoothing kernel (8 mm full-width at half-maximum). The image was then scaled to provide the individual voxel probability threshold  $p_{thr}$  of 0.005 by determining the value  $z_{thr}$  such that approximately  $p_{thr} \cdot N$  voxels have intensity greater than  $z_{thr}$ . The thresholding was then accomplished by setting those voxels with intensity greater  $z_{thr}$  to 1 (activated voxels), voxels with intensity less than  $z_{thr}$  to 0. Finally, AlphaSim determined which activated voxels belonged to which clusters. When all clusters had been found, the size of each cluster in voxels was recorded in a frequency table. This simulation estimated that, in our volume of the entire brain, a cluster size of more than 51 contiguous activated voxels (with each individual voxel surviving a  $P < 0.005$  threshold; height threshold) would occur by chance with a probability of less than 0.05 (cluster extent threshold).

The effect size/time course plots in Figs. 2B and C and 3B were computed by averaging the GLM's  $\beta$  values/time course data across subjects. To ensure the independence of the data that we analyzed within a region of interest (effect sizes and Bayesian model comparison) from the data used to select regions of interest, we performed the following leave-one-out analysis. First, we looped through all subjects and computed group averages for all but one subject. We then extracted the data within a region centered at the leave-one-out group peak voxel of the subject that was excluded in this group.

The structural T1 images were coregistered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the *t*-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas (41).

**ACKNOWLEDGMENTS.** We thank T. Behrens for insightful comments. This research was supported by grants from the Gordon and Betty Moore Foundation

(to J.P.O. and A.R.), a Gordon and Betty Moore Foundation Scholarship (to K.W.), and a Searle Scholarship (to J.P.O.), and by the Caltech Brain Imaging Center.

1. von Neumann J, Morgenstern O (1944) *Theory of Games and Economic Behavior* (Princeton Univ Press, Princeton, NJ).
2. Sutton RS, Barto AG (1998) *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA).
3. Dayan P, Abbott LF (2001) *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 (2005) Representation of action-specific reward values in the striatum. *Science* 310:1337–1340.
6. Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. *Neuron* 58:451–463.
7. Glimcher P, Camerer C, Poldrack R, Fehr E (2008) *Neuroeconomics: Decision Making and the Brain* (Academic Press, New York).
8. Roesch MR, Olson CR (2003) Impact of expected reward on neuronal activity in prefrontal cortex, frontal and supplementary eye fields and premotor cortex. *J Neurophysiol* 90:1766–1789.
9. Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. *Nature* 400:233–238.
10. Kawagoe R, Takikawa Y, Hikosaka O (1998) Expectation of reward modulates cognitive signals in the basal ganglia. *Nat Neurosci* 1:411–416.
11. Ikeda T, Hikosaka O (2003) Reward-dependent gain and bias of visual responses in primate superior colliculus. *Neuron* 39:693–700.
12. Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. *Proc Natl Acad Sci USA* 106:17199–17204.
13. Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–1787.
14. Thorndike EL (1911) *Animal Intelligence: An Experimental Study of the Associate Processes in Animals* (Macmillan, New York).
15. Glimcher PW, Dorris MC, Bayer HM (2005) Physiological utility theory and the neuroeconomics of choice. *Games Econ Behav* 52:213–256.
16. Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *J Neurosci* 16:1486–1510.
17. Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. *Nature* 441:223–226.
18. Platt M, Padoa-Schioppa C (2008) Neural representations of value. *Neuroeconomics: Decision Making and the Brain*, ed Glimcher P (Academic Press, New York), pp 441–461.
19. Padoa-Schioppa C (2007) Orbitofrontal cortex and the computation of economic value. *Ann N Y Acad Sci* 1121:232–253.
20. Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *J Neurosci* 27:9984–9988.
21. 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.
22. Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53:147–156.
23. Rudebeck PH, Bannerman DM, Rushworth MF (2008) The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. *Cogn Affect Behav Neurosci* 8:485–497.
24. Rudebeck PH, et al. (2008) Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci* 28:13775–13785.
25. 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.
26. Knutson B, Taylor J, Kaufman M, Peterson R, Glover G (2005) Distributed neural representation of expected value. *J Neurosci* 25:4806–4812.
27. Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. *Neuroimage* 46:1004–1017.
28. Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. *J Neurosci* 27:14502–14514.
29. Glascher 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.
30. FitzGerald TH, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. *J Neurosci* 29:8388–8395.
31. Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62:733–743.
32. Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–419.
33. 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.
34. Ostlund SB, Winterbauer NE, Balleine BW (2009) Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *J Neurosci* 29:8280–8287.
35. Balleine BW, Liljeholm M, Ostlund SB (2009) The integrative function of the basal ganglia in instrumental conditioning. *Behav Brain Res* 199:43–52.
36. Balleine BW, Ostlund SB (2007) Still at the choice-point: Action selection and initiation in instrumental conditioning. *Ann N Y Acad Sci* 1104:147–171.
37. de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci* 29:11330–11338.
38. Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
39. Friston KJ, Worsley KJ, Frackowiak RSJ, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. *Hum Brain Mapping* 1: 214–220.
40. Ward BD (2000) Simultaneous Inference for FMRI Data, Analysis of Functional Neuro-Images/National Institute of Mental Health. Available at <http://afni.nimh.nih.gov/pub/dist/doc/manual/AlphaSim.pdf>. Accessed September 1, 2009.
41. Duvernoy HM (1999) *The Human Brain. Surface, Blood Supply and Three-Dimensional Section Anatomy* (Springer, New York) 2nd ed.