Behavioral/Systems/Cognitive

Action-Specific Value Signals in Reward-Related Regions of the Human Brain

Thomas H. B. FitzGerald, Karl J. Friston, and Raymond J. Dolan

Wellcome Trust Centre for Neuroimaging, London WC1N 3BG, United Kingdom

Estimating the value of potential actions is crucial for learning and adaptive behavior. We know little about how the human brain represents action-specific value outside of motor areas. This is, in part, due to a difficulty in detecting the neural correlates of value using conventional (region of interest) functional magnetic resonance imaging (fMRI) analyses, due to a potential distributed representation of value. We address this limitation by applying a recently developed multivariate decoding method to high-resolution fMRI data in subjects performing an instrumental learning task. We found evidence for action-specific value signals in circumscribed regions, specifically ventromedial prefrontal cortex, putamen, thalamus, and insula cortex. In contrast, action-independent value signals were more widely represented across a large set of brain areas. Using multivariate Bayesian model comparison, we formally tested whether value–specific responses are spatially distributed or coherent. We found strong evidence that both action-specific and action-independent value signals are represented in a distributed fashion. Our results suggest that a surprisingly large number of classical reward-related areas contain distributed representations of action-specific values, representations that are likely to mediate between reward and adaptive behavior.

Introduction

Adaptive decision-making requires an agent to link their evaluation of current and future states to the actions that might cause them. One needs to know that one wants to drink a cup of tea and which actions are needed to realize this goal. On standard accounts, linking the two involves assigning values to particular actions, which are then compared and used to guide choice (one opens the tea caddy rather than the coffee jar). In addition, tracking action-specific values is useful for rapid and appropriate updating during learning (Gershman et al., 2009), for example, in Q-learning (Watkins and Dayan, 1992). Thus, action-specific values assume great importance for understanding decision-making in humans and other animals.

Despite the importance of such information, few studies have directly considered how it is encoded. Primate studies demonstrate the existence of action-specific value signals in the striatum (Samejima et al., 2005; Pasquereau et al., 2007; Lau and Glimcher, 2008; Hori et al., 2009; Seo et al., 2012) and parietal cortex (Platt and Glimcher, 1999; Sugrue et al., 2004) but not, to our knowledge, elsewhere. In humans, there is evidence of such representations in motor areas (Wunderlich et al., 2009), but not those most

often associated with valuation, such as the striatum. One explanation for this is that the conventional univariate fMRI analysis methods used in most studies are insensitive to signals with subject-specific profiles.

Multivariate analysis, by contrast, is sensitive to signals from distributed neuronal populations with subject-specific profiles, and does not require focal, spatially coherent activations (Norman et al., 2006). Importantly, here we applied multivariate Bayes (MVB) (Friston et al., 2008), a decoding technique that allowed us to test for representations of parametric variables. Additionally, MVB allows the comparison of different spatial-encoding models, enabling us to test whether or not value signals show local spatial coherence. For example, one can ask whether adjacent voxels encode value-related signals in a similar manner, or whether there is a distributed pattern of encoding best modeled by considering voxels in isolation (Fig. 1) (Friston et al., 2008).

Subjects performed an instrumental learning task (Fig. 1), responding with either their right or left index finger. Value signals related to each action (Q_R and Q_L, respectively) were estimated using a Q-learning algorithm. To model action-specific values, we defined a quantity AV as ($Q_R - Q_L$). This models differential responding to one action value compared with the other—indicating the presence of action-specific value signals. To test for action-independent value signals, we defined CV ($|Q_R - Q_L|$) (Boorman et al., 2009; FitzGerald et al., 2009), which models the comparison between different options along a common value scale (Padoa-Schioppa and Assad, 2006).

We acquired high-resolution fMRI data from cortical and subcortical structures known to be important in processing value. Our guiding hypothesis, based on the increased sensitivity of multivariate analysis to reward-related activity, was that action-specific value signals would be found across sev-
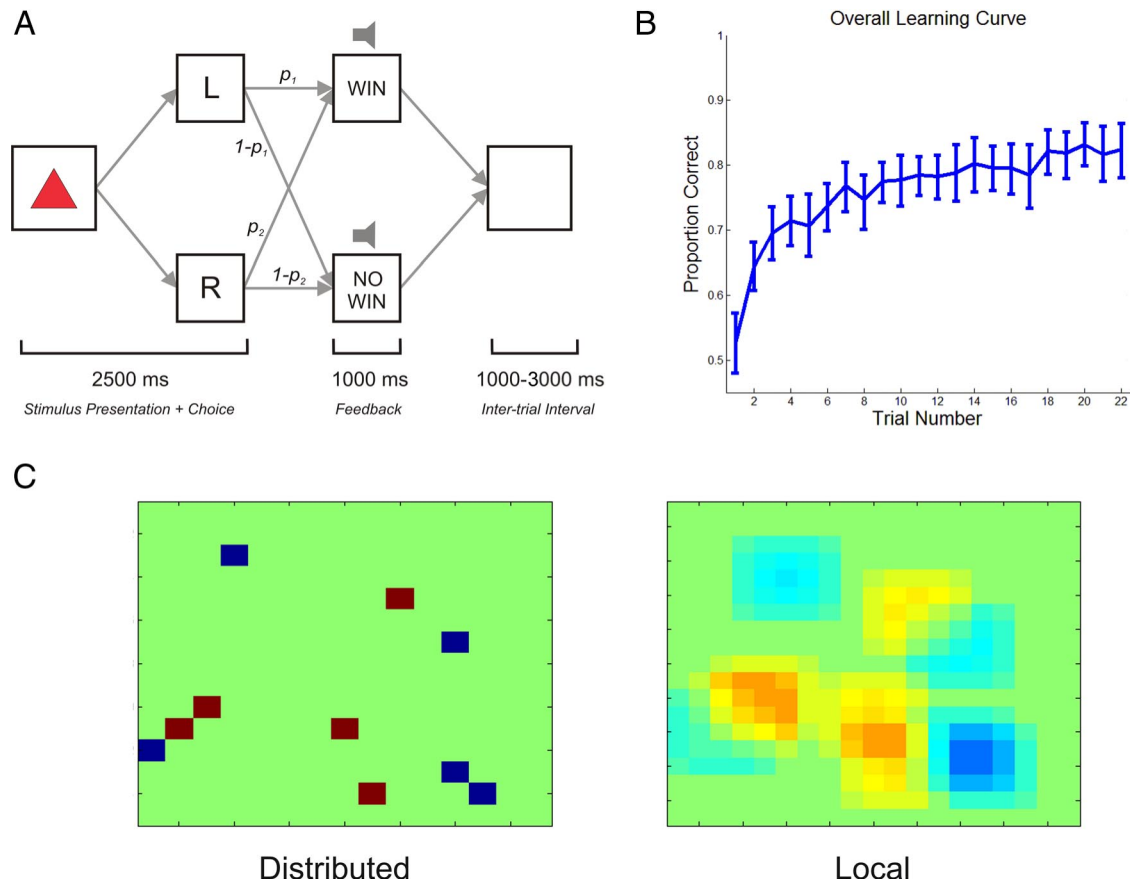
Figure 1. ***A***, Instrumental learning task. Subjects were presented with an arbitrary stimulus and had 2500 ms to make one of two responses via a button box in either hand. Each stimulus–action pairing was associated with a certain probability of reward (10 pence) versus no reward. Outcomes were signaled with two different sounds, which were presented for 1000 ms, followed by a variable intertrial interval (1000–3000 ms, uniform distribution). ***B***, Proportion of trials on which subjects chose the objectively higher-valued action, pooled over all subjects, sessions, and cues. Subjects chose the objectively better action on an increasing proportion of trials—showing that they were able to learn the task contingencies. Error bars indicate bootstrapped 95% confidence intervals. ***C***, Schematic of distributed and coherent coding schemes on a two-dimensional surface. Under a distributed encoding scheme (left), nearby voxels do not show similar responses, in contrast to a scheme with local coherence (right), where clumping of responses is observed. Red/orange, positive response to arbitrary parameter; blue, negative response; green, no response.

eral regions, in particular the striatum and ventromedial prefrontal cortex (vmPFC), and would show a distributed pattern of encoding over voxels.

Materials and Methods

Subjects. Twenty-six (10 female) right-handed subjects, age range 19–28 years, participated in the study. All subjects were free of neurological or psychiatric disease and consented to participate. The study was approved by the Joint National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS trust) and Institute of Neurology (University College London) Ethics Committee. After scanning, subjects received a sum of money according to their performance during the task (£21.80-£28.80).

Stimuli and task. Subjects performed an instrumental learning task with visual cues and auditory feedback (Fig. 1). On each trial of the experiment, a colored box was presented and the subjects were required to make either a "left" or a "right" response by pressing a button on the corresponding keypad. After 2.5 s, they heard either a higher pitched, game-show-like "win" sound, or a lower "no win" sound, each lasting for 1 s. The box disappeared at the end of the sound. There was then a variable intertrial interval of 1–3 s before the next trial began. Subjects were instructed (truthfully) that every time they heard a "win" sound, they would receive 10 pence, but they would receive nothing for a "no win" sound. These winnings were summed and given to the subjects at the end of the task.

Boxes differed according to their outcome contingencies. There were four types, with win probabilities [0.05 0.30], [0.05 0.55], [0.3 0.55], and

[0.4 0.9]. In the course of the experiment, six cues were presented with each of the four sets of contingencies, three where $P(\text{Win}|\text{Chose Right}) > P(\text{Win}|\text{Chose Left})$, and three where the converse was the rule. The experiment was separated into blocks of 44 trials. In each block, two boxes appeared in pseudorandomized order (presentation was constrained so that no box was presented more than three trials in a row). Outcome contingencies were selected such that each pair of contingency types was presented four times throughout the experiment (cues with identical or mirror image contingencies were never presented together in the same block).

There were six blocks in each of two scanning sessions (12 in total). Each box was presented in only one block, and the contingencies assigned to each box were counterbalanced across subjects. Contingency types over blocks were also fully counterbalanced across subjects. At the end of each block, subjects were asked to indicate which box they thought was better, and were then given 4.5 s to rate how confident they were in this judgment. They were then shown a running total of their winnings for 4.5 s before the next block started.

Behavioral analysis. Three subjects who reported using deterministic strategies were excluded from further analysis, leaving a total of 23 subjects. To generate the regressors used to analyze the imaging data, each subject's behavior was fit with a *Q*-learning model (Watkins and Dayan, 1992) incorporating a softmax decision rule. *Q*-learning updates the values of individual stimulus action pairs, $Q(s, a)$, according to a reward prediction error weighted by a learning rate α.

$$Q(s_{t+1}, a_{t+1}) = Q(s_t, a_t) + \alpha(R_t - Q(s_t, a_t))$$

Crucially, this generates trial-by-trial estimates (Q_R and Q_L) for the value of each of the two actions available to a subject, given the stimulus that was on screen at the time.

The softmax decision rule gives the probability of choosing action R (P_R) based on the difference in value between action R and action L ($Q_R - Q_L$), and the temperature parameter β, which determines the preference sensitivity between the two options.

$$P_R = \frac{1}{1 + e^{(Q_L - Q_R)/\beta}}$$

The learning rate α and softmax temperature β parameters were fitted individually for each subject using maximum likelihood (accuracy) estimators.

Learning was assessed by comparing the estimated probability of choosing the correct (higher value) action during their first five exposures to a stimulus with their last five exposures. To check for value-related effects on reaction times, we used a general linear model in which the absolute difference in value between the two options (CV) was regressed on reaction time. We performed this separately for each subject, with group level inference implemented using a single-sample t test on the ensuing parameter (regression slope) estimates in the standard (summary statistic) way.

fMRI data acquisition. Three-dimensional gradient-echo T2*-weighted echo-planar (EPI) images were acquired on a 3T Trio Siemens scanner with a resolution of 1.5 mm isotropic. Thirty-two slices were acquired (echo time, 32.86 ms; repetition time, 3.2 s; interleaved acquisition order), which allowed data acquisition from a partial volume of thickness 48 mm that was angled and positioned in each subject to ensure coverage of the vmPFC, ventral striatum, and dopaminergic midbrain.

Data were acquired using a gradient-echo 3D EPI sequence on a 3T Trio Siemens scanner with an isotropic resolution of 1.5 mm. At this resolution, this 3D EPI sequence was shown to yield improved fMRI sensitivity compared with standard 2D EPI sequences (Lutti et al., 2012). Data were acquired from a partial volume that was angled and positioned in each subject to ensure coverage of the vmPFC, ventral striatum, and dopaminergic midbrain. Acquisition parameters were identical to these used in Lutti et al., (2012). Thirty-two slices (partitions) were acquired with 25% oversampling along the partition direction to avoid wrap-around of excited signal outside the field-of-view into the image volumes. The resulting volume was 3.2 s. Parallel imaging was used to optimize the fMRI sensitivity of the 3D EPI sequence. Images were reconstructed using the GRAPPA algorithm (Griswold et al., 2002) available on the scanner console. The resulting echo time was 32.86 ms. The image field-of-view was 192 × 192 × 48 mm. No dropout compensation methods were used, since signal dropout is reduced at the high image resolution used here (Weiskopf et al., 2006). In each session, 485 images were collected (~25 min each, two per subject). Subjects lay in the scanner with foam head-restraint pads to minimize any movement. They responded using two fMRI-compatible button boxes, one held in each hand. To optimize coregistration, five whole-brain EPIs were collected with identical scanning parameters. Whole-brain multiparameter maps were collected at 1 mm isotropic resolution (Helms et al., 2009).

Preprocessing and statistical analysis were performed using SPM8 (Wellcome Trust Centre for Neuroimaging, London, UK, www.fil.ion.ucl.ac.uk/spm). After discarding the first five images from the task sessions to allow for T1 equilibration effects, EPI images were realigned with the first volume and unwarped using field maps generated using the Fieldmap toolbox as implemented in SPM8 (Hutton et al., 2002). This corrects for both static distortions and motion-related alterations in these distortions. The mean whole-brain EPI was then coregistered with the T1-weighted structural EPI, and the smaller-volume EPIs coregistered to the whole-brain EPI. To allow the comparison of data between subjects in a common space, the DARTEL (Ashburner, 2007) toolbox was used to normalize structural scans and coregistered EPIs to MNI space.

Regions of interest. Where appropriate, anatomical regions of interest (ROIs) were based on the Automatic Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). The vmPFC ROI was created by combining the medial orbital part of the superior frontal gyrus and the gyrus rectus AAL regions. Putamen ROIs were based on the AAL putamen region, excluding regions that overlapped with the nucleus accumbens (NA) ROI (see below). Hippocampus, amygdala, insula, and thalamus ROIs were taken directly from the appropriate AAL regions.

The NA ROIs were defined as 8 mm spheres centered at [11.11, 11.43, −1.72] and [−11.11, 11.43, −1.72], exclusively masked to remove overlap (all coordinates in MNI space) (Guitart-Masip et al., 2011). The midbrain ROI, taken to comprise the substantia nigra and ventral tegmental areas (SN/VTA) was manually defined on the mean magnetization transfer structural image for the group (Bunzeck and Düzel, 2006; Helms et al., 2009; Guitart-Masip et al., 2011). As a check analysis, we also used a substantia nigra ROI taken from the WFU Pickatlas (Maldjian et al., 2003). The same pattern of MVB results was observed in both midbrain regions, so we report those from the hand-drawn ROI. Finally, a check analysis was performed using a sphere of 6 mm radius placed in the ventricles (center, [−1, −37, 5]). No above-chance decoding was possible from this region.

Multivariate Bayes analysis. MVB is a hierarchical Bayesian decoding scheme that solves the ill-posed many-to-one (here, voxel-to-target variable) mapping by invoking empirical priors on data features, specified by a second level of the model of the mapping. Prior spatial covariance is separated into spatial patterns U, specified a priori, and pattern weights η that are estimated during Bayesian model inversion. This means that different spatial patterns of activity (for example, distributed vs locally clustered models) can be compared by changing the spatial patterns U (Fig. 1) (Morcom and Friston, 2012). Our preferred model was a distributed model, in which patterns are single voxels that show no characteristic spatial coherence; in other words, U is the identity matrix.

MVB assumes that coding is sparse in pattern space. In other words, only a small proportion of patterns make a large contribution to the decoding. The set of pattern weights is iteratively partitioned into subsets based on the size of the weights, where all patterns within a subset are assumed to have the same variance. This optimization of subsets constitutes a greedy search using a standard variational scheme (Friston et al., 2007). The variational approximation to model evidence increases with each iteration until the optimal set size is reached. The final set of patterns then constitutes the decoding model with the greatest evidence for the pattern weights. Different models (spatial patterns) can then be compared directly in terms of their log evidence, without the need for cross-validation. Note that this inference pertains to different models or hypotheses about distributed representations in the brain and it is not an inference about whether the representations exist or not. In what follows, we compare distributed and spatially coherent representations of value. However, we first establish the existence of these representations using a randomization procedure that converts the log evidence for any particular model (relative to a null model) into a classical p value (see below).

Table 1. Group-level ROI average results

ROI	p value	
	AV	CV
vmPFC, right	0.445	**<0.001**
vmPFC, left	0.343	**<0.001**
Putamen, right	0.276	0.392
Putamen, left	0.219	**0.028**
NA, right	0.835	**0.003**
NA, left	0.489	0.095
Insula, right	0.429	0.237
Insula, left	0.306	0.351
Thalamus, right	0.273	0.409
Thalamus, left	0.849	0.673
Hippocampus, right	0.118	0.236
Hippocampus, left	0.306	**0.017**
Amygdala, right	0.976	0.371
Amygdala, left	0.290	0.514
SN/VTA	0.869	**0.046**

No regions showed a significant response to AV, while bilateral vmPFC, left putamen, right NA, left hippocampus, and SN/VTA showed significant responses to CV. p values are taken from two-tailed tests. Bold text indicates results which were significant at $p < 0.05$ uncorrected.

Table 2. Regional responses to AV in the group-level MVB analysis

ROI	p value					Exceedance probability (distributed > local)
	AV	AV orthogonalised to Act	AVrej	Q_R	Q_L	
vmPFC, right	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
vmPFC, left	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Putamen, right	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Putamen, left	**<0.001**	0.001	<0.001	<0.001	<0.001	>0.999
NA, right	0.609	—	—	—	—	—
NA, left	0.522	—	—	—	—	—
Insula, right	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Insula, left	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Thalamus, right	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Thalamus, left	**<0.001**	<0.001	<0.001	<0.001	<0.001	>0.999
Hippocampus, right	0.029	0.012	0.144	<0.001	0.019	>0.999
Hippocampus, left	0.182	—	—	—	—	—
Amygdala, right	0.609	—	—	—	—	—
Amygdala, left	0.574	—	—	—	—	—
SN/VTA	0.325	—	—	—	—	—

Bilateral vmPFC, putamen, insula, and thalamus showed a significant response to AV, as did the right hippocampus. All areas showing a significant response to AV survived a check analysis where the model included a binary choice regressor (Act). Action-specific rejected value signals (AVrej) were present in all regions, as was information about both the value of the right action (Q_R) and the left action (Q_L). In all regions, Bayesian model selection strongly favored distributed models, suggesting that the coding of AV lacks clear spatial clustering. Bold text indicates decoding that was significant at a threshold of $p < 0.05$, FDR-corrected. We only present FDR-corrected results for AV, since the other analyses are *post hoc* analyses designed to either check or qualify our inferences.

Table 3. Regional responses to CV in the group-level MVB analysis

ROI	p value (CV)	Exceedance probability (distributed > local)
vmPFC, right	**<0.001**	>0.999
vmPFC, left	**<0.001**	>0.999
Putamen, right	**<0.001**	>0.999
Putamen, left	**<0.001**	>0.999
NA, right	**0.007**	>0.999
NA, left	**0.036**	>0.999
Insula, right	**<0.001**	>0.999
Insula, left	**<0.001**	>0.999
Thalamus, right	**<0.001**	>0.999
Thalamus, left	**<0.001**	>0.999
Hippocampus, right	**<0.001**	>0.999
Hippocampus, left	**0.005**	>0.999
Amygdala, right	**0.009**	>0.999
Amygdala, left	**0.017**	>0.999
SN/VTA	0.643	—

Bilateral vmPFC, putamen, NA, insula, thalamus, hippocampus, and amygdala showed a significant response to CV. In all regions, Bayesian model selection strongly favored distributed models, suggesting that the coding of CV lacks clear spatial clustering. Bold text indicates decoding that was significant at a threshold of $p < 0.05$, FDR-corrected.

MVB analysis can be based upon the same design matrices used in a conventional mass-univariate SPM analysis (Friston et al., 1994). In a conventional SPM, the design matrix (X) is used to predict the time course at a particular voxel (Y). Here, activity in several voxels is used to predict a target variable in the design matrix, specified with a contrast in the usual way. To remove confounding or uninteresting effects, the resulting null space of the design matrix is used to adjust the target and predictor variables.

To provide a classical p value testing for the significance of the mapping over subjects, we used a randomization procedure to produce a null distribution over the log evidence ratio for each model. Here, the target vector is phase shuffled 20 times and the log evidence ratio (the difference between the log evidence under a prior of no mapping and the log evidence generated by the greedy search) is recalculated each time. This null distribution was used to convert the log evidence ratio into a p value for each region and for each subject.

Group level inference was performed using Fisher's method for combining independent p values (Fisher, 1925). Here, for k-independent p values, a statistic χ^2 is calculated as:

$$\chi^2 = -2\sum_{i=1}^{k}\ln(p_i)$$

Under the null hypothesis, this has a χ^2 distribution with $2k$ degrees of freedom. This statistic is then used to generate a group-level p value. Multiple-comparison correction was performed using the false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995).

To test whether the value signals were spatially coherent, we performed additional analyses using three encoding models with differing degrees of local spatial structure, parameterized respectively as 1 mm^3, 2 mm^3 and 4 mm^3 FWHM Gaussians (Friston et al., 2008; Morcom and Friston, 2012). We then performed Bayesian model selection (BMS) on an ROI-by-ROI basis using log evidence ratios and a random effects procedure that accounts for group heterogeneity (Stephan et al., 2009).

Design matrix specification. Based on our behavioral analysis, we created a number of subject-specific regressors to model the imaging data. To establish the presence of an action-specific value signal, it is not enough simply to find activity correlated with the value of one or another action (Q_R or Q_L) considered independently, as this also picks out regions encoding both values equivalently. The key variable is thus the difference between the two action values (AV := $Q_R - Q_L$), since this reflects only action-specific values. In addition, we calculated the comparative value of the actions (FitzGerald et al., 2009), defined as the absolute difference between their values (CV := $|Q_R - Q_L|$), and the reward prediction error (PE) associated with each outcome. The existence of such a signal implies that the brain compares the values of the options available to it, which in turn requires the existence of some form of common currency within which this comparison can be made. CV is action independent because it will take the same value regardless of which action is valued more highly, provided that the difference between action values is the same. The presence of such a signal thus does not necessitate that a region contain action-specific value information, hence providing a useful companion (and comparison) to our analyses of action-specific value.

Design matrices were created as for a standard SPM analysis. A general linear model was specified with events at cue onset times (Cue), modulated by parametric regressors encoding CV and AV, as well as events at outcome time (Out) modulated by PE. To test whether the responses correlated with AV could be explained by simple action signaling, we created another model that included an additional binary regressor, encoding the action taken on each trial (Act). This ensures that any significant decoding of AV could be attributed to action-specific value, having accounted for action *per se*.

To test whether the signals we observed truly reflected the values of both actions (Boorman et al., 2009; FitzGerald et al., 2009), or whether they simply reflected the chosen value modulated by action (Plassmann et al., 2007; Wunderlich et al., 2009), we created a model with separate action-specific regressors for the chosen and rejected options (AVchs and AVrej, respectively). Significant decoding of AVrej suggests that the value of both options is signaled in an action-specific way.
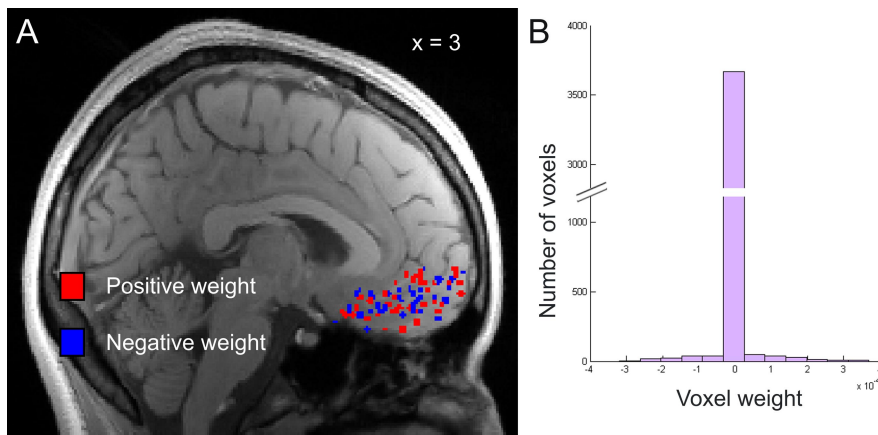
Figure 2. Voxel weights (η) from the MVB analysis for AV in the right vmPFC ROI for Subject 5. ***A***, Voxels with positive (red) and negative (blue) voxel weights overlaid on Subject 5's T1-weighted structural scan ($x = 3$ mm). Voxels with a positive weight show a positive response to AV and voxels with a negative weight show a negative response to AV according to our multivariate analysis. Image thresholded at $\eta > 0.00005$ for positive weights and $\eta < 0.00005$ for negative weights. Positive and negative weights are interspersed without any obvious pattern, suggesting a lack of spatial coherence. ***B***, Histogram of voxel weights. Only a small proportion of voxels have large weights, showing that only a small proportion are important for decoding (this is the hallmark of sparse distribution).

To ascertain whether regions from which AV could be decoded contained information about Q_R, Q_L, or both, we created a model where instead of CV and AV, Cue was modulated by Q_R and Q_L. Because of the computationally intensive nature of MVB, we did not run these extra analyses for regions that did not show a significant response to AV.

The resulting stimulus functions were then convolved with a hemodynamic response function. Regression was performed using standard maximum likelihood in SPM. Low-frequency fluctuations were removed using a high-pass filter (cutoff, 128 s) and remaining temporal autocorrelations were modeled with a two parameter auto-regression model.

Conventional ROI averaging. For comparison with our MVB analysis, we also applied conventional ROI averaging to our data, using the models described above. Parameter estimates were averaged over voxels within each ROI, and a two-tailed *t* test was applied to the group-level data. Because the purpose of the univariate analyses was an illustrative comparison with our MVB results, we did not apply a multiple-comparisons correction to maximize sensitivity at the expense of specificity. Generally, we would not recommend this (straw man) analysis of regional averages over the use of ROIs to provide a small search volume for (mass-univariate) effects within the ROI.

Results

Behavior
Averaging over all stimuli, each and every subject selected the higher value action more often in the last five trials compared with the first five ($p < 0.0001$, Wilcoxon rank sum test; Fig. 1). Reaction times showed a strong negative effect of CV ($p < 0.0001$, Wilcoxon signed rank test), consistent with our previous findings (FitzGerald et al., 2009). This means subjects responded more quickly when there was a greater disparity (clarity) in the value of the options available to them.

Conventional ROI averaging
None of the ROIs showed a significant correlation with AV (Table 1). This remained true even at a liberal threshold of $p < 0.1$. Activity in bilateral vmPFC, SN/VTA, left putamen, left hippocampus, and right NA showed a significant correlation with CV ($p < 0.05$, two-tailed *t* test) (Table 1).

MVB analysis
We observed a significant decoding of AV bilaterally in vmPFC, putamen, insula, thalamus and right hippocampus (Table 2).

Secondary analyses showed that, in all of these regions, this mapping could not be explained simply by action (Act) (Table 2).

We next tested whether these signals reflected true action-specific value signaling, which could be used to influence choice, or whether they simply reflected action-specific chosen value signals that evolve postdecision (Wunderlich et al., 2009). To do this, we tested for the presence of action-specific signals relating to the value of the rejected option (AVrej). Such signals were present in bilateral vmPFC, putamen, insula, and thalamus (Table 2), suggesting that activity in these regions is driven at least in part by true AV signals. Note that AVrej could not be decoded from right hippocampus at a rate significantly above chance. Because of this, and because we only found evidence for AV unilaterally rather than bilaterally (in this and only this region), we regard evidence for AV in the right hippocampus as weaker than in other areas.

All ROIs from which AVs were significantly decoded also permitted decoding of Q_L and Q_R, considered individually. This suggests that essential information about the value of both actions was available to these regions in both hemispheres.

A significant decoding of CV was possible from bilateral vmPFC, putamen, NA, insula, thalamus, hippocampus, and amygdala ROIs (Table 3). Interestingly, although we observed a significant CV response in our SN/VTA ROI using conventional ROI averaging, decoding was not above chance in this region using MVB. This suggests that our distributed decoding model was not appropriate for detecting spatially coherent correlates of CV, or that the ROI average result was a false positive (given the multiple comparisons we performed).

Spatial distributions of activity
In addition to testing for the presence of a signal, MVB allows one to examine its spatial distribution (Morcom and Friston, 2012). This is based on comparing evidence for models containing different priors on the spatial distribution of activity using BMS (Friston et al., 2008). We examined whether value signals show spatial coherence (where nearby voxels have similar responses) or not (voxels are best considered in isolation from their neighbors) (Fig. 1). For both AV and CV, distributed encoding models were strongly favored over spatially coherent models (Tables 2, 3; Fig. 2). This suggests that, at least at the resolution of our 1.5 mm isotropic voxels, representations of both value signals are spatially distributed (which would render them difficult to detect with ROI averages).

Discussion
Action-specific value signals play a key role in both learning and choice behavior. Previous studies in humans and other primates suggest the existence of such signals in a small number of structures. These studies are limited, on the one hand, by the restricted brain coverage of single-unit recording studies and, on the other hand, by an inherent difficulty in detecting many kinds of signal using mass-univariate and ROI-averaging techniques. Exploiting recent developments in multivariate analysis, we provide evidence of action-specific value signals in the vmPFC, putamen, insula, and thalamus that are spatially distributed.

In accordance with single-unit studies, which report both action-specific value signals (Samejima et al., 2005; Pasquereau et al., 2007; Lau and Glimcher, 2008; Hori et al., 2009; Seo et al., 2012) and action-specific prediction errors (Stalnaker et al., 2012) in the striatum, we found evidence for action-specific value signals in the putamen. This is consistent with a key role for this region in linking reward to action (Redgrave et al., 2010), as well as action planning in general (Monchi et al., 2006). It is also consistent with recent evidence that this region encodes habit as opposed to goal-directed value during choice behavior (Wunderlich et al., 2012). A previous study reported evidence of effector-specific outcome signaling in the ventral striatum, but not action-specific value signals *per se* (Gershman et al., 2009). Interestingly, we did not observe any evidence for AV in the NA, perhaps suggesting that this region is more involved in generalized, or stimulus-locked, signaling of value. A similar observation pertains to the SN/VTA, which showed no evidence of AV. We are mindful here that it is unwise to draw conclusions from the negative results, but we note that these observations agree with direct recording data from dopaminergic midbrain neurons, which likewise failed to find evidence for action-specific PE signals (which would be needed to update action-specific values) (Nakahara et al., 2004; Morris et al., 2006; Roesch et al., 2007).

Activity in the vmPFC was strongly modulated by AV. Although there has been speculation about this previously (O'Doherty, 2011) and some neuroimaging findings are consistent with the idea that the vmPFC encodes such values (Gläscher et al., 2009; Palminteri et al., 2009; Wunderlich et al., 2009), to our knowledge, this is the first clear demonstration that such signals do exist in vmPFC. Palminteri et al. (2009) showed that activity in the vmPFC tracked the value of contralateral options in an instrumental learning task, but since subjects were always presented with two stimuli to choose between, it is difficult to separate the effects of spatial attention or lateralized stimulus-value processing from action-specific valuation. A similar caveat applies to the activity consistent with action-specific value signaling in the lateral intraparietal sulcus reported in Gershman et al. (2009). The lack of such spatial- or stimulus-bound effects may explain why we did not observe a significant correlation between AV and activity in the vmPFC in our ROI analyses, contrary to what might be predicted based on Palminteri et al. (2009).

Our MVB findings are consistent with a number of single-unit studies that suggest that the activity in the vmPFC (or regions of the orbitofrontal cortex) contain information about spatial goals (Feierstein et al., 2006) and a range of parameters important for economic choice (Padoa-Schioppa and Assad, 2006; Kennerley and Wallis, 2009; Kennerley et al., 2009). Although our results suggest that action-specific value information is present in the vmPFC, we do not see this as invalidating previous claims that the region is involved in pure stimulus- (or goods-) based choices (Wunderlich et al., 2010). Instead, our results provide evidence for heterogeneity of vmPFC function, consistent with its putative central role in both valuation and choice.

Bilateral thalamus and insula both showed activity that correlated with AV, in keeping with previous findings linking these structures both to action (Anderson et al., 1994; Fink et al., 1997) and reward (Gottfried et al., 2003; Balleine, 2005; Vickery et al., 2011). In contrast, we did not find evidence for action-specific value signals in the amygdala and left hippocampus, and only relatively weak evidence for AV in the right hippocampus. This was despite the fact that we found strong evidence of action-independent value (CV) signals in these regions. This may suggest that these structures are more involved in stimulus- or goods-based evaluation, rather than roles directly involving action (Balleine, 2005).

CV signals were more widespread than AV, being found in bilaterally in vmPFC, putamen, NA, insula, thalamus, hippocampus, and amygdala. This fits with previous multivariate fMRI studies that have found activity across large swathes of the brain for anticipated value, acquired through stimulus-outcome learning (Kahnt et al., 2010, 2011), and for outcome signaling (Vickery et al., 2011). Our results extend these findings to the domains of instrumental learning and economic choice. They also speak to the fact that value information is likely to impact upon a large variety of cognitive processes, and should thus be disseminated widely.

Comparing the results we obtained using ROI averaging with those from the MVB analysis suggests that, for certain questions, MVB can be more sensitive to AV, which argues that multivariate techniques are likely to be important for studies examining the properties of action-specific value signals. MVB was also more sensitive to CV, except in the SN/VTA ROI. This may be because encoding in the SN/VTA is actually spatially uniform and coherent. However, without further study, it is difficult to be sure what the precise explanation for this difference in inference.

Our Bayesian model comparison suggests that, rather than being spatially coherent and clustered (at least at the scale of our fMRI voxels), value-signaling is spatially distributed. This lack of spatial coherence fits with the observation that multiple decision- or value-related parameters are often encoded by neurons recorded from the same sites (Padoa-Schioppa and Assad, 2006; Lau and Glimcher, 2008; Kennerley and Wallis, 2009; Kennerley et al., 2009). It can also explain why multivariate analysis methods are more sensitive to such activity (Vickery et al., 2011), since mass-univariate analysis typically involves spatial smoothing, and thus assumes local coherence—at least over a few voxels.

One limitation of multivariate analyses is their lack of spatial resolution, depending—by definition—on responses in ROIs. This can easily be addressed by using smaller ROIs, as in searchlight procedures (Kriegeskorte and Bandettini, 2007; Kahnt et al., 2010). Because we focused our high-resolution data acquisition on the vmPFC, midbrain, and other deep structures associated with reward, we were unable to test for the presence of AV signals in other brain areas associated with action and reward, such as the anterior cingulate cortex and the caudate nucleus.

In conclusion, our analyses provide evidence for representations of action-specific value in the vmPFC, putamen, thalamus, and insula. Understanding how valuation processes influence action is critical for explaining the neurobiology of choice, and we suggest that action-specific value signals are likely to provide a critical link. Our findings thus represent a step toward a broader goal of understanding how organisms in general, and humans in particular, use and encode value information to guide their behavior.

References

Anderson TJ, Jenkins IH, Brooks DJ, Hawken MB, Frackowiak RS, Kennard C (1994) Cortical control of saccades and fixation in man: a PET study. Brain 117:1073–1084. CrossRef Medline

Ashburner J (2007) A fast diffeomorphic image registration algorithm. Neuroimage 38:95–113. CrossRef Medline

Balleine BW (2005) Neural bases of food-seeking: affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav 86:717–730. CrossRef Medline

Benjamini Y, Hochberg J (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300.

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. CrossRef Medline

Bunzeck N, Düzel E (2006) Absolute coding of stimulus novelty in the human substantia nigra/VTA. Neuron 51:369–379. CrossRef Medline

Feierstein CE, Quirk MC, Uchida N, Sosulski DL, Mainen ZF (2006) Representation of spatial goals in rat orbitofrontal cortex. Neuron 51:495–507. CrossRef Medline

Fink GR, Frackowiak RS, Pietrzyk U, Passingham RE (1997) Multiple nonprimary motor areas in the human cortex. J Neurophysiol 77:2164–2174. Medline

Fisher S (1925) Statistical methods for research workers. Edinburgh: Oliver and Boyd.

FitzGerald TH, Seymour B, Dolan RJ (2009) The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci 29:8388–8395. CrossRef Medline

Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ (1994) Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 2:189–210. CrossRef

Friston K, Mattout J, Trujillo-Barreto N, Ashburner J, Penny W (2007) Variational free energy and the Laplace approximation. Neuroimage 34:220–234. CrossRef Medline

Friston K, Chu C, Mourão-Miranda J, Hulme O, Rees G, Penny W, Ashburner J (2008) Bayesian decoding of brain images. Neuroimage 39:181–205. CrossRef Medline

Gershman SJ, Pesaran B, Daw ND (2009) Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J Neurosci 29:13524–13531. CrossRef Medline

Gläscher J, Hampton AN, O'Doherty JP (2009) Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex 19:483–495. Medline

Gottfried JA, O'Doherty J, Dolan RJ (2003) Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 301:1104–1107. CrossRef Medline

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47:1202–1210. CrossRef Medline

Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJ, Dayan P, Dolan RJ, Duzel E (2011) Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J Neurosci 31:7867–7875. CrossRef Medline

Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N (2009) Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. Neuroimage 47:194–198. CrossRef Medline

Hori Y, Minamimoto T, Kimura M (2009) Neuronal encoding of reward value and direction of actions in the primate putamen. J Neurophysiol 102:3530–3543. CrossRef Medline

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. Neuroimage 16:217–240. CrossRef Medline

Kahnt T, Heinzle J, Park SQ, Haynes JD (2010) The neural code of reward anticipation in human orbitofrontal cortex. Proc Natl Acad Sci U S A 107:6010–6015. CrossRef Medline

Kahnt T, Heinzle J, Park SQ, Haynes JD (2011) Decoding the formation of reward predictions across learning. J Neurosci 31:14624–14630. CrossRef Medline

Kennerley SW, Wallis JD (2009) Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables. Eur J Neurosci 29:2061–2073. CrossRef Medline

Kennerley SW, Dahmubed AF, Lara AH, Wallis JD (2009) Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci 21:1162–1178. CrossRef Medline

Kriegeskorte N, Bandettini P (2007) Analyzing for information, not activation, to exploit high-resolution fMRI. Neuroimage 38:649–662. CrossRef Medline

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463. CrossRef Medline

Lutti A, Thomas DL, Hutton C, Weiskopf N (2012) High-resolution functional MRI at 3 T: 3D/2D echo-planar imaging with optimized physiological noise correction. Magn Reson Med. Advance online publication. Retrieved Oct. 8, 2012. doi:10.1002/mrm.24398. CrossRef Medline

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated

method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239. CrossRef Medline

Monchi O, Petrides M, Strafella AP, Worsley KJ, Doyon J (2006) Functional role of the basal ganglia in the planning and execution of actions. Ann Neurol 59:257–264. CrossRef Medline

Morcom AM, Friston KJ (2012) Decoding episodic memory in ageing: a Bayesian analysis of activity patterns predicting memory. Neuroimage 59:1772–1782. CrossRef Medline

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063. CrossRef Medline

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. Neuron 41:269–280. CrossRef Medline

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424–430. CrossRef Medline

O'Doherty JP (2011) Contributions of the ventromedial prefrontal cortex to goal-directed action selection. Ann N Y Acad Sci 1239:118–129. CrossRef Medline

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223–226. CrossRef Medline

Palminteri S, Boraud T, Lafargue G, Dubois B, Pessiglione M (2009) Brain hemispheres selectively track the expected value of contralateral options. J Neurosci 29:13465–13472. CrossRef Medline

Pasquereau B, Nadjar A, Arkadir D, Bezard E, Goillandeau M, Bioulac B, Gross CE, Boraud T (2007) Shaping of motor responses by incentive values through the basal ganglia. J Neurosci 27:1176–1183. CrossRef Medline

Plassmann H, O'Doherty J, Rangel A (2007) Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J Neurosci 27:9984–9988. CrossRef Medline

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238. CrossRef Medline

Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, Agid Y, DeLong MR, Obeso JA (2010) Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nat Rev Neurosci 11:760–772. CrossRef Medline

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. CrossRef Medline

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340. CrossRef Medline

Seo M, Lee E, Averbeck BB (2012) Action selection and action value in frontal-striatal circuits. Neuron 74:947–960. CrossRef Medline

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2012) Reward prediction error signaling in posterior dorsomedial striatum is action specific. J Neurosci 32:10296–10305. CrossRef Medline

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017. CrossRef Medline

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289.

Vickery TJ, Chun MM, Lee D (2011) Ubiquity and specificity of reinforcement signals throughout the human brain. Neuron 72:166–177. CrossRef Medline

Watkins CJCH, Dayan P (1992) Q-learning. Mach Learn 8:279–292. CrossRef

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. Neuroimage 33:493–504. CrossRef Medline

Wunderlich K, Rangel A, O'Doherty JP (2009) Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci U S A 106:17199–17204. CrossRef Medline

Wunderlich K, Rangel A, O'Doherty JP (2010) Economic choices can be

made using only stimulus values. Proc Natl Acad Sci U S A 107:15005–15010. CrossRef Medline

Wunderlich K, Dayan P, Dolan RJ (2012) Mapping value-based planning and extensively trained choice in the human brain. Nat Neurosci 15:786–791. CrossRef Medline