Behavioral/Cognitive

Neural Correlates of the Divergence of Instrumental Probability Distributions

Mimi Liljeholm, Shuo Wang, June Zhang, and John P. O'Doherty
Division of the Humanities and Social Sciences and Computation and Neural Systems Program, California Institute of Technology, Pasadena, California 91125

Flexible action selection requires knowledge about how alternative actions impact the environment: a "cognitive map" of instrumental contingencies. Reinforcement learning theories formalize this map as a set of stochastic relationships between actions and states, such that for any given action considered in a current state, a probability distribution is specified over possible outcome states. Here, we show that activity in the human inferior parietal lobule correlates with the divergence of such outcome distributions–a measure that reflects whether discrimination between alternative actions increases the controllability of the future–and, further, that this effect is dissociable from those of other information theoretic and motivational variables, such as outcome entropy, action values, and outcome utilities. Our results suggest that, although ultimately combined with reward estimates to generate action values, outcome probability distributions associated with alternative actions may be contrasted independently of valence computations, to narrow the scope of the action selection problem.

Introduction

Theories of goal-directed behavior originated with a seminal series of early demonstrations that animals are able to learn about the structure of their environment in the absence of primary rewards (Blodgett, 1929; Tolman and Honzik, 1930). Specifically, in stark contrast to the, then dominating, view of behavior as being controlled exclusively by the incremental modulation of stimulus-response (S-R) associations based on contingent reward or punishment (Thorndike, 1933), these studies suggested that when given the opportunity to explore a maze, nonrewarded rats constructed a valence-neutral "cognitive map" of instrumental and environmental relationships, that could be flexibly integrated with subsequent reward information to generate an optimal course of action (Tolman, 1948).

Contemporary accounts of behavioral control characterize instrumental performance as being governed both by the reinforcement-based, S-R, component and by a more cognitive, goal-directed, system (Balleine and Dickinson, 1998). These separate strategies have been formalized as distinct classes of reinforcement learning (RL): An automatic "model-free" system, in which the values of actions are acquired by means of a reward prediction error (RPE), and a "model-based" class that constructs a mental map of the environment and generates decisions by flexibly combining estimates of state-transition probabilities with outcome utilities (Doya et al., 2002; Daw et al., 2005). Thus, in model-based RL, relationships between actions and future states of the world are represented explicitly and independently of associated motivational features.

Notably, given equivalent costs, actions that yield identical outcome states need not be contrasted further in terms of motivational features, reducing the demand for a computationally costly binding of outcome probabilities with utilities. Consequently, the extent to which actions differ in terms of their relationships to future states, that is, the divergence of their outcome probability distributions, can be used to prune searches of the mental map. Instrumental divergence also serves as a measure of agency–the more actions differ with respect to contingent states, the more flexible control an agent has over its environment. Because of these important characteristics, we hypothesized that a neural signature of instrumental divergence, dissociable from that of motivational variables, would be discernible during human goal-directed performance.

We scanned human participants with functional magnetic resonance imaging (fMRI) as they performed a simple task in which available actions yielded various food rewards with different probabilities (Fig. 1*A*). Our primary objective was to assess neural correlates of the difference between outcome distributions associated with alternative actions, formalized as Jensen–Shannon (JS) divergence–a measure that quantifies the distance between probability distributions. The relationship between JS divergence and other decision variables is illustrated in Figure 1*B*. First, in this example, uncertainty about which outcome will be obtained (i.e., outcome entropy) is the same across actions; for each action, one can be almost certain that a particular food reward will occur while the alternative food and the "no food" outcome will not. Likewise, provided that one enjoys oranges as much as Twix bars, the actions are equivalent in terms of their

Received March 30, 2013; revised June 14, 2013; accepted June 18, 2013.

Author contributions: M.L. and J.P.O. designed research; M.L., S.W., and J.Z. performed research; M.L. analyzed data; M.L. and J.P.O. wrote the paper.

This work was funded by a National Institutes of Health Grant (DA033077-01) to J.P.O. The authors thank Daniel McNamee for helpful discussions.

Correspondence should be addressed to Mimi Liljeholm, Division of the Humanities and Social Sciences and Computation and Neural Systems Program, California Institute of Technology, Pasadena, CA 91125. E-mail: mlil@hss.caltech.edu.

DOI:10.1523/JNEUROSCI.1353-13.2013

expected value. And yet, the two actions clearly differ with respect to contingent states; this difference is captured by JS divergence.

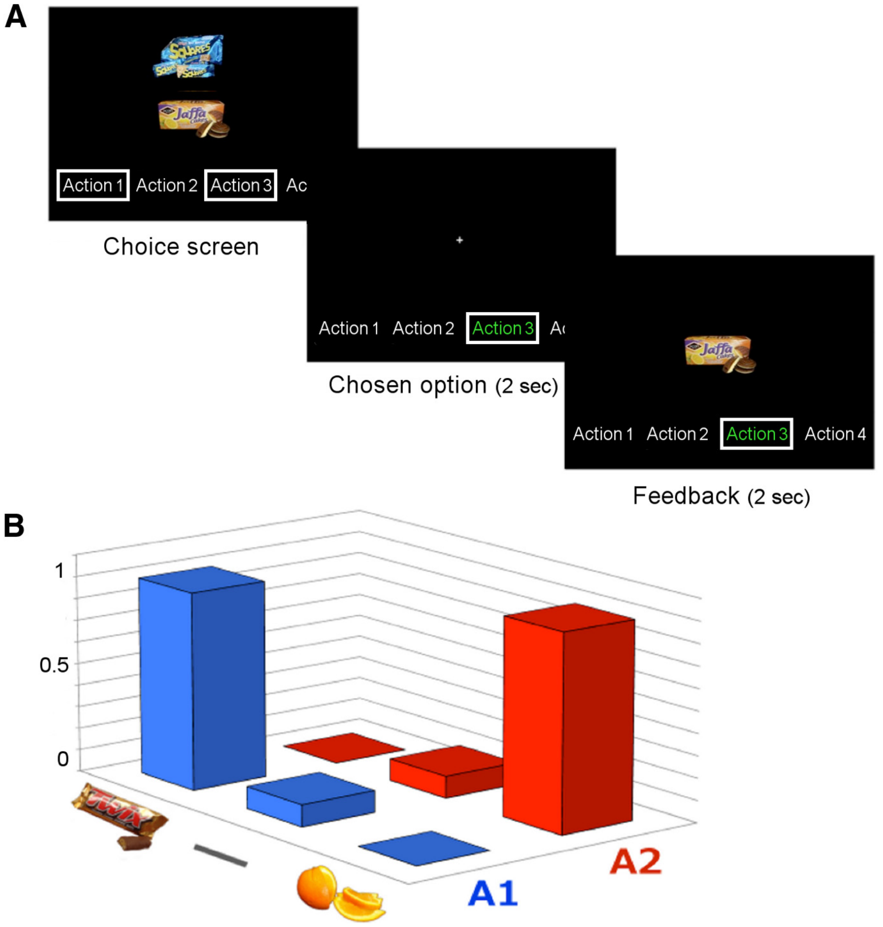

Figure 1. Illustration of the task and the concept of outcome divergence. ***A***, Illustration of trial in the choice phase: at the onset of a trial, two of four alternative actions are highlighted in white, indicating their availability, together with depictions of potential trial outcomes. After participants choose, the chosen action is highlighted in green for 2 s, followed by either a picture of the obtained food outcome (on rewarded trials) or a white line in the center of the screen (nonrewarded trials). Trials are separated by a jittered 6 s intertrial interval. ***B***, An illustration of the relationship between actions, outcomes, and associated probabilities. The graph shows two available actions, A1 and A2 (coded in blue and red, respectively), where the bars represent the probability distribution of each action across a set of three potential outcomes: a Twix bar, an orange, and a "no-food" outcome state, indicated by a gray bar. JS divergence is a measure of the distance between the two distributions.

Materials and Methods

Participants. Twenty-two healthy normal volunteers (mean age = 23 0.1 ± 4.3; range = 19–38, 10 females) participated in the study. The volunteers were pre-assessed to exclude those with a history of neurological or psychiatric illness. The eating attitudes test (EAT-26) (Garner et al., 1982) was administered and indicated no eating disorders in any of the subjects (mean score, 3.6 ± 2.8; range, 0–13; all scores were under the 20 point cutoff). Before being scheduled for the experiment, the subjects were prescreened to ensure that they enjoyed sweet and salty treats, that they had no allergies or intolerances, and that they were not overweight, on a diet, or planning to go on a diet. Subjects were asked to fast for at least 4 h before their scheduled arrival time at the laboratory, but were permitted to drink water. All subjects gave informed consent and the study was approved by the Institutional Review Board of the California Institute of Technology.

Task and procedure. A simple instrumental task was used, in which four action alternatives (i.e., button presses) yielded various food rewards with different probabilities. Specifically, on any given trial, two available actions could differ with respect to the probability with which they produced their respective rewards, with respect to the subjective utility of those rewards and/or with respect to the integrated action value. Thus, by manipulating both probabilities and utilities, we were able to largely decorrelate the different components of the decision problem. To ensure sufficient variance in experienced probabilities and utilities, each subject participated in three consecutive sessions (during a single appearance by the subject in the lab), with the same four actions but with a novel set of food outcomes and outcome probabilities being used in each session. Throughout the task, available actions were indicated by corresponding rectangles on the computer screen, together with images of the food outcomes potentially produced by those actions (Fig. 1*A*). At the start of the experiment, participants were informed that they would have to remain in the laboratory for 30 min after completing the task, during which they would be allowed to consume any earned treats.

The probabilities with which actions produced their outcomes were generated so as to minimize correlations between our three decision variables (i.e., between outcome probabilities, outcome values, and action values), and thus varied slightly across subjects. Nonetheless, across sessions and action alternatives, probabilities of 0, 0.2, 0.5, 0.7, and 1.0 were used for each subject. In addition, each subject had two probabilities chosen from the set [0.3 0.6 0.8 and 0.9]. The probabilities drawn from this set differed depending on the decorrelation constraints imposed by subjective outcome utilities. A minimum of two and maximum of four distinct probabilities were used in each session.

Table 1. List of the 36 food treats used in BDM auction

Kit Kat (small pack)	MentosFruit	KeeblerChipsDelux
Hershey milk	Twix	Lays Classic potato chips
Apple (red)	Reese's	Lindor Truffles (milk)
Banana	Doritos (ranch flavor)	Orange
Famous Amos (small pack)	Peanut M&Ms	Sunmaid raisins
Ferrero Rocher chocolates	Fig Newton	Rice Krispy Treat (small)
Flaming Cheetos	Chips Ahoy (small pack)	Oreos
Fritos	Skittles	Ruffles potato chips
Ghirardelli milk chocolate	Lindt Swiss bittersweet dark	Streusel Cakes
Godiva dark chocolate bar	Milano cookies	Toblerone
HoHo	Pringle	Pepperidge Farms Cookies
Keebler Fudge Stripes cookies	Mrs. Fields chocolate chip	White grapes

The subjective values of 36 potential food rewards (listed in Table 1), represented by photographic images, were assessed using evaluative ratings of their pleasantness (on a scale from 0 to 9), as well as a Becker–DeGroot–Marschak (BDM) auction procedure that has been shown to elicit an individual's willingness to pay for a consumer good (Becker et al., 1964). The pleasantness ratings were used to set the utility of food stimuli in subsequent phases of the experiment. The BDM auction was used to obtain convergent evidence for these ratings, providing a measure of inter-rater reliability. Specifically, in the BDM auction, participants were endowed with $5 with which to bid on the various food items. They were instructed that, on each trial, they would have to indicate an amount of money from $0 to $5 that they were willing to pay for the food item displayed on that trial and that, at the end of the experiment, the computer would randomly select one trial from all presented in that phase, as well as randomly draw an amount between \$0 and \$5. Participants were further told that if the bid that they had indicated on the randomly drawn trial was less than the amount generated by the computer, they would not receive the food item, but would get to keep the \$5, and that otherwise they would have to pay the amount generated by the computer and would get to consume the food item.

In each session, participants first went through a "contingency learning" phase, in which each action was sampled 10 times with associated food rewards occurring with respective probabilities. Only one action was available on each trial in this phase, to ensure complete sampling, and outcome occurrences were generated using predetermined sequences such that if, for example, the probability of an outcome was 0.2, that outcome was presented on exactly 2 of the 10 trials. Furthermore, participants were instructed that they would not actually earn any of the food rewards produced by the actions in this phase, but that it was simply meant to expose them to action–outcome relationships. They then proceeded to a "choice phase" in which, on each of 48 trials, they chose between two of the four action alternatives (Fig. 1*A*). They were instructed that, at the end of the experiment, three trials would be drawn from this phase, and that they would be allowed to consume any rewards earned on those trials upon completion of the task. Following the choice phase, participants provided judgments of the existence and strength of each action–outcome relationship. Active scanning was only performed during the choice phases.

Computational learning model. We implemented a model-based RL learner, which uses experience with state transitions to update a matrix, $T(\mathbf{s},a,\mathbf{s'})$, of state transition probabilities, where each element of $T(\mathbf{s},a,\mathbf{s'})$ holds the current estimate of the probability of transitioning from state $\mathbf{s}$ to $\mathbf{s'}$ given action a. In our task, as illustrated in Figure 1*A*, on each trial participants were presented with a choice screen displaying two available actions together with the food outcomes potentially produced by those two actions. Thus, each initial state was defined by the particular available actions and their potential outcomes. The two available actions were drawn from a total of four; consequently, in the choice phase of each session, there were six distinct initial states, repeatedly encountered across 48 trials. The state transitions were initialized to the preprogrammed distributions from the contingency learning phase. In each step, leaving state $\mathbf{s}$ and arriving in state $\mathbf{s'}$ having taken action a, the FORWARD learner computes a state prediction error[3] (SPE): $\delta_{SPE} = 1 - T(\mathbf{s},a,\mathbf{s'})$, and updates the probability $T(\mathbf{s},a,\mathbf{s'})$ of the observed transition via: $T(\mathbf{s},a,\mathbf{s'}) = T(\mathbf{s},a,\mathbf{s'}) + \eta\delta_{SPE}$ where η is a free parameter controlling the learning rate. Estimated transition probabilities are used together with the rewards at the end states, $r(\mathbf{s'})$ (the magnitude of which were based on pleasantness ratings and taken as given, since potential rewards were displayed together with the actions) to compute state-action values, $Q(s,a)$ as the expectation over the value of the successor state. This is done by defining the state-action values at each level in terms of reward anticipated at the next level: $Q(s,a) = \Sigma_{\mathbf{s'}} T(\mathbf{s},a,\mathbf{s'}) * r(s')$.

The model additionally assumes that participants select actions stochastically using probabilities generated by a softmax distribution, such that $P(s, a) = \frac{\exp(\tau \times Q(s, a))}{\Sigma_{b=1}^{n} \exp(\tau \times Q(s, b))}$, where the free "inverse temperature" parameter τ controls the degree to which choices are biased toward the highest valued action. To account for the difference in salience between food pictures and the "no outcome" display (which simply consisted of a white line), we used separate learning rate parameters, η and η', on rewarded and nonrewarded trials, respectively. Free parameters were fit to behavioral data by minimizing the negative log-likelihood, $-\Sigma\log(P(s,a))$, of obtained choices for each individual.

Information theoretic variables. We computed the JS divergence of the outcome probability distributions for the two actions available on a given trial. A finite and symmetrized version of the Kullback–Leibler divergence, JS divergence specifies the distance between probability distributions M and N as follows:

$$JSD = \frac{1}{2}\sum_i ln\left(\frac{M(i)}{P(i)}\right) M(i) + \frac{1}{2}\sum_i ln\left(\frac{N(i)}{P(i)}\right) N(i) \text{ where } P = \frac{1}{2}(M + N).$$

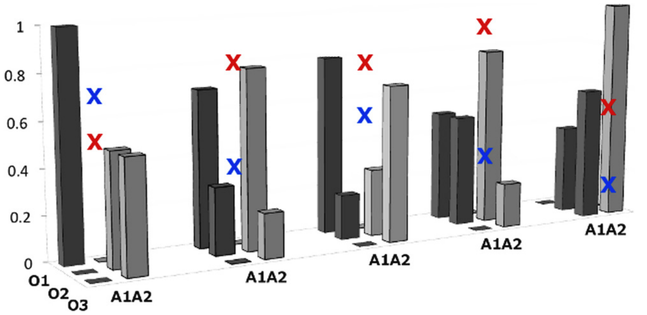

Figure 2. A representative set of cases used in the experiment. Each case consists of two probability distributions, one for each available action, across three potential outcome states (01, 02, and 03) including the no-food outcome. A1 and A2 indicate the two actions available on a given trial, drawn from a total of four possible action alternatives. X indicates the levels of JS divergence (blue) and entropy (red) for each case.

It is worth noting that, while we have used JS divergence here to quantify the degree to which alternative actions differ with respect to contingent transitions in environmental states, it is not the only computational variable that captures this conceptual point. For example, mutual information (between actions and outcomes) is a highly related information theoretic measure that would be identical to JS divergence for our current purposes, as would be the χ^2 divergence. A particularly compelling aspect of JS divergence is its remarkable generality: it applies to nominal and numerical, discrete and continuous random variables, and it intuitively generalizes to an arbitrary number of probability distributions. The applicability to multiple distributions (Lin, 1991) is especially important as it eliminates the need for a complex, and presumably computationally costly, process of comparing multiple available actions in a pairwise fashion.

Another important information theoretic variable that can be extracted from the state-transition matrix, and that has been previously shown to profoundly affect decision making (Paulus et al., 2002; Feinstein et al., 2006), is the uncertainty, or entropy, of outcome states. Whereas JS divergence reflects the distance between outcome probability distributions, entropy reflects the degree of uncertainty in an outcome, which is greatest when the probability distribution over outcomes is uniform (i.e., all outcomes are equally likely) and smallest when the probability of a particular outcome is 1 or 0. We computed the Shannon entropy of the outcome variable X conditional on the action variable Y, defined as $H(X|Y) = \sum_{x\epsilon X,\ y\epsilon Y} p(x,y) log \frac{p(y)}{p(x, y)}$, both for the case where the (chosen) action is known ($p(y) = [1,0]$) and for the case where the two available actions are equally likely ($p(y) = [0.5,0.5]$). To illustrate the relationship between outcome probabilities and information theoretic variables, a representative set of probability distributions, with corresponding levels of JS divergence and entropy, are shown in Figure 2.

Imaging procedure and analysis. A 3 T scanner (MAGNETOM Trio; Siemens) was used to acquire structural T1-weighted images and T2*-weighted echoplanar images (repetition time = 2.65 s; echo time = 30 ms; flip angle = 90°; 45 transverse slices; matrix = 64 × 64; field of view = 192 mm; thickness = 3 mm; slice gap = 0 mm) with blood oxygenation level-dependent (BOLD) contrast. To recover signal loss from dropout in the medial orbitofrontal cortex (O'Doherty et al., 2002), each horizontal section was acquired at 30° to the anterior commissure–posterior commissure axis. Image processing and statistical analyses were performed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm). The first four volumes of images were discarded to avoid T1 equilibrium effects. All remaining volumes were corrected for differences in the time of slice acquisition, realigned to the first volume, spatially normalized to the Montreal Neurological Institute (MNI) echoplanar imaging template, and spatially smoothed with a Gaussian kernel (8 mm, full-width at half-maximum). We used a high-pass filter with a cutoff of 128 s.

For each subject, we constructed an fMRI design matrix, merged across the three sessions, with two regressors modeling the distinct time periods of each trial. The first, choice-period regressor, modeled a BOLD response from the onset of each trial until the chosen action was performed and a second stick function modeled the onset of the feedback screen. For the choice-period regressor, we entered as parametric modulators, in order, the expected value of the chosen action, the sum of and the absolute difference between the expected values of the two available actions, the utility of the outcome potentially produced by the chosen action, and the sum and absolute difference in utility of the two potential outcomes depicted on the screen. Absolute differences, rather than that between chosen and unchosen, were used to minimize regressor redundancies. Finally, for this trial period, we entered as modulators the entropy conditional on the chosen action, the entropy conditional on both available actions, and the JS divergence of outcome probability distributions of available actions. For the outcome regressor, we entered as modulators, in order, the RPE, the SPE, and the utility of the received outcome. Orthogonalization was applied according to order such that each parametric modulator was orthogonalized to all preceding modulators associated with the same onset regressor. To rule out motor-execution components, the response time on each trial was added as a regressor of no interest, as were two regressors indicating the three sessions and six regressors accounting for the residual effects of head motion. All regressors were convolved with a canonical hemodynamic response function. Group-level random-effects statistics were generated by entering contrasts of parameter estimates for the different modulators into a between-subjects analysis.

We specifically looked for neural effects in areas previously shown to be involved in the implementation of our modeled decision variables. First, in a recent neuroimaging study, Gläscher et al. (2010) assessed neural correlates of SPEs, finding effects in the lateral prefrontal cortex (LPFC) and intraparietal sulcus (IPS). We predicted that activity in these areas would likewise correlate with SPEs in the current study. We also predicted that the dorsolateral prefrontal cortex (DLPFC) would encode the entropy of outcome probability distributions: activity in this area has been shown to scale with the amount of uncertainty associated with a decision, to predict risk aversion (Weber and Huettel, 2008) and, when disrupted by transcranial magnetic stimulation, to increase selection of risky option (Knoch et al., 2006). With respect to our motivational variables, studies of goal-directed performance, which emphasize the casual relationships between actions and outcomes (Tanaka et al., 2008; Liljeholm et al., 2011) or higher-order relational structures (Hampton et al., 2006), have implicated the ventromedial prefrontal cortex (VMPFC) in the encoding of action values. In contrast, activity in the medial orbitofrontal cortex (mOFC), the insula, and ventral striatum (VS), has been shown to correlate with the utility, or value, of a stimulus (Hare et al., 2008; Abler et al., 2009; Schmidt et al., 2012). The anterior insula has also been implicated in the affective evaluation of food pictures (Pelchat et al., 2004; Wang et al., 2004) and in anticipation and experience of appetitive tastes (O'Doherty et al., 2001, 2002). Finally, The VS has been shown, by myriad studies, to encode a RPE (O'Doherty et al., 2003, 2004).

Small volume corrections (SVCs) were performed on a priori regions of interest (ROIs), using a 10 mm sphere with center coordinates obtained by averaging across relevant studies (coordinates are listed in Table 2). All other effects were reported at $p < 0.05$, using cluster size thresholding (CST) to adjust for multiple comparisons (Forman et al., 1995). AlphaSim, a Monte Carlo simulation (AFNI) was used to determine cluster size and significance. Using an individual voxel probability threshold of $p = 0.001$ indicated that using a minimum cluster size of 134 MNI transformed voxels resulted in an overall significance of $p < 0.05$. To eliminate nonindependence bias for plots of parameter estimates, a leave-one-subject-out (Esterman et al., 2010) approach was used, in which 22 general linear models (GLMs) were run with one subject left out in each, and with each GLM defining the voxel cluster for the left out subject. Using rfxplot (Gläscher, 2009), mean β-weights were extracted from spheres (10 mm) centered on the LOSO peaks (identified within ROIs for SVCs) and were averaged across subjects to plot overall effect sizes.

Table 2. Center coordinates for SVC, averaged across local maxima and studies

		MNI coord.	
Area	Function	*x*, *y*, *z*	Sources
VMPFC	Chosen action value	4, 55, −7	Liljeholm et al. (2011) Tanaka et al., (2008) Hampton et al. (2006)
mOFC	Stimulus utility	0, 29, −14	Hare et al. (2008) Plassmann et al. (2010)
Lateral VS	Stimulus utility/RPE	±26, 7, −6	O'Doherty et al. (2004)
Medial VS		6, 14, −2	
Insula	Stimulus utility	L/R anatomical	WFU pickatlas
LPFC	SPE	±45, 11, 32	Gläscher et al. (2009)
IPS	SPE	−27/39, 54, 42	
Right DLPC	Entropy	36, 6, 52	Weber et al. (2008)

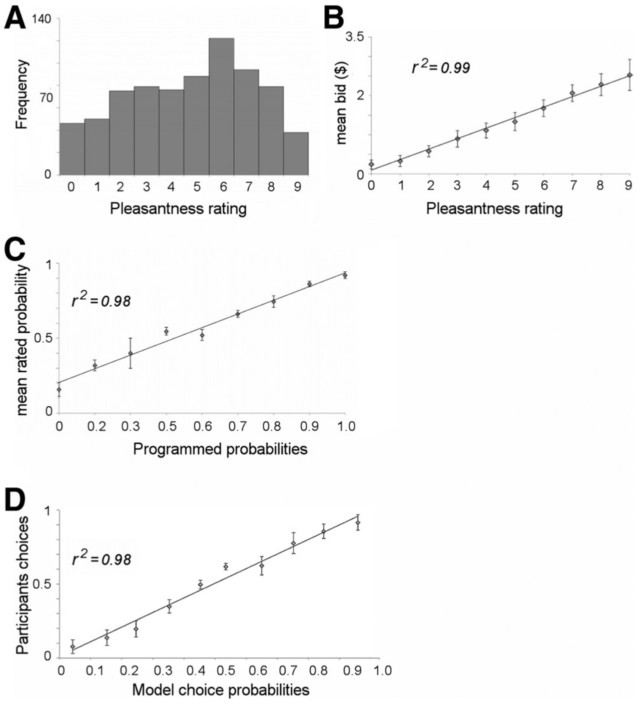

Figure 3. Behavioral results: ***A***, Frequencies of pleasantness ratings for 36 food stimuli across participants. ***B***, Scatter plot showing the mean bid in the BDM, across participants and food stimuli, as a function of pleasantness ratings. ***C***, Scatter plot showing the mean rated action probability, across participants, as a function of programmed action probabilities (binned). ***D***, Scatter plot showing participants mean choices as a function of the model-generated choice probabilities (binned). Error bars in indicate SEM.

Results

Behavioral data and model fits

Pleasantness ratings were fairly evenly distributed across the scale, and were highly correlated ($r^2 = 0.99$) with bids in the BDM auction (Fig. 3*A*). Moreover, participants' judgments of action–outcome relationships, collected at the end of each session, were close to the programmed contingencies ($r^2 = 0.98$; Fig. 3*B*). Finally, a comparison of the model-derived choice probabilities with participants' actual choices suggested that the model matches behavior well ($r^2 = 0.98$; Fig. 3*C*).

Neuroimaging results

All results described below are corrected for multiple comparisons at $p < 0.05$ using either CST across the whole brain, or SVC based on coordinates averaged across previous studies reporting effects of relevant decision variables (see Materials and Methods

Table 3. Coordinates and significance values for imaging contrasts

Contrast	Area	MNI coord. *x*, *y*, *z*	Cluster size at $p < 0.005$
JS Divergence	IPL	51, −25, 40	398
	SMA	9, −10, 58	
	Precentral	54, −10, 40	
Entropy of chosen action	DLPFC	45, 20, 34	68
Summed utility of potential outcomes	Anterior insula	−39, 14, 7	138
	Ventral putamen	−27, 11, −14	
Q value of chosen action	VMPFC	0, 56, −11	33
	Dorsal SMA	−12, −10, 73	536
	Paracentral	−24, −28, 73	
RPE	Medial frontal cortex	−6, 62, 19	1263
	VS	9, 11, −5	50
	Occipital	39, −88, 10	3356
SPE	PCC	3, −31, 31	388
	IPS	−36, −43, 46	198
	LPFC	42, 20, 37	108

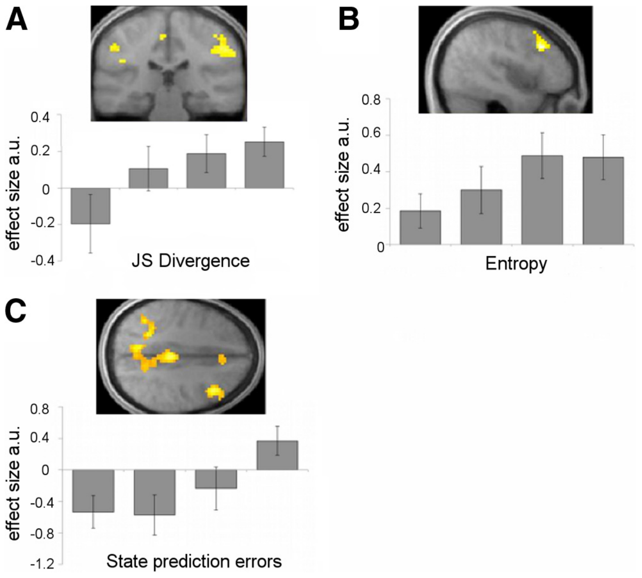

Figure 4. Imaging effects of state-transition variables. ***A***, Maps of the *t* statistics for tests of neural modulation by JS divergence, showing effects in the right supramarginal gyrus of the IPL. ***B***, Map of the *t* statistics for tests of neural modulation by the entropy of the outcome probability distribution for the chosen action, showing effects in the DLPFC. ***C***, Map of the *t* statistics for tests of neural modulation by SPEs during the feedback period of each trial, showing effects in the LPFC, IPS, and PCC. Bar plot shows responses to SPEs in the LPFC. Bar plots showing mean β-weights across variable values are binned into the 25th, 50th, 75$^{\text{th}}$, and 100th percentiles. Error bars indicate SEM. a.u., arbitrary units.

for details of multiple-comparison correction strategy). Coordinates and cluster sizes for all the activated areas described below are reported in Table 3.

State-transition variables

An exploratory test, for areas in which activity during the choice period (i.e., from the trial onset until a response was performed) correlated with the distance between the outcome probability distributions of the two available actions (i.e., with JS divergence) yielded effects in right anterior supramarginal gyrus of the inferior parietal lobule (IPL; Fig. 4*A*) as well as in the supplementary motor area (SMA) extending into the right precentral gyrus, surviving CST correction. Critically, these effects also emerged when no orthogonalization was applied, ruling out the possibility that activity in these areas was selectively modulated by the orthogonalized component. Neural activity correlating with the entropy of the chosen action during the same period emerged in the right DLPFC (SVC) (Fig. 4*B*). At the time of outcome delivery, activity in both the LPFC and IPS was significantly modulated by the SPE (SVC), as was activity extending throughout middle and posterior cingulate cortex(PCC; CST) (Fig. 4*C*).

Simpler representations of outcome probabilities

A possible alternative explanation for our effects of JS divergence is that the IPL and SMA are encoding simpler representations of outcome probabilities; for example, if participants did not attend to the sensory-specific or potential motivational differences between food outcomes but instead encoded, for each action, the probability of obtaining any food reward, activity in the IPL or SMA might be scaling with the simple difference between or sum of these probabilities across available actions. As illustrated in Figure 1*B*, JS divergence can deviate quite dramatically from the difference between reward probabilities, with the former being relatively high and the latter being zero in this example. Indeed, the difference between reward probabilities was not strongly correlated with JS divergence in our task and, moreover, weak correlations were in different directions across subjects (mean absolute value of $r = 0.18$, SEM = 0.03). Our task also included several instances for which divergence varied independently of the sum of reward probabilities; for example, again using Figure 1*B*, consider a case in which the probabilities associated with the Twix bar are shifted to the no-food outcome state and vice versa; this would yield the same level of JS divergence but a dramatic reduction in the sum of reward probabilities. Nonetheless, the sum of reward probabilities across available actions was strongly positively correlated with JS divergence for each subject (mean $r = 0.78$, SEM = 0.05).

To empirically assess the neural effects of simpler representations of outcome probabilities relative to those of JS divergence we specified two additional GLMs that were identical to our original model except for the replacement of the JS divergence modulator with a regressor modeling the difference between or the sum of reward probabilities, respectively, in the second and third model. Separate models were specified for two main reasons: first, to avoid excessive colinearity of regressor variables and, second, to rule out the possibility that neural correlates varied only with those components of JS divergence that are orthogonal to linear representations of probabilities. No significant effects of JS divergence, the difference between, or sum of reward probabilities emerged when a single GLM was used to model all three variables (in addition to those previously specified) suggesting that there was indeed too much shared variance between these regressors. We also found no effects of either the difference or the sum of reward probabilities at our threshold of statistical significance when using separate GLMs, although effects did emerge at an uncorrected threshold of $p < 0.05$. To formally determine which of the three variables provided the best account of neural activity in the SMA and IPL, we performed a Bayesian model selection analysis. Specifically, we used the first-level Bayesian estimation procedure in SPM8 to compute a voxelwise whole-brain log-model evidence map for every subject and each model (Penny et al., 2005). Then, to model inference at the group level, we applied a random effects approach (Rosa et al., 2010) at every voxel of the log evidence data falling within anatomical masks of the right supramarginal gyrus and SMA, constructing an exceedance posterior probability (EPP) map for each model and for each anatomical area.

We found that the difference between reward probabilities did indeed provide the best account of neural activity in the largest

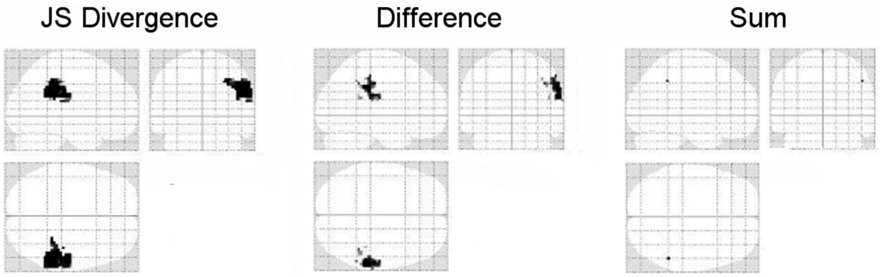

Figure 5. Results of a Bayesian model selection analysis. EPP maps in an anatomical mask of the right supramarginal gyrus, generated based on three GLMs that were identical except for the inclusion of either JS divergence (left), the difference between reward probabilities (middle), or the sum of reward probabilities (right). The EPP maps are thresholded at 0.333 with probabilities of 1.00 indicated by black.

portion of SMA, generating EPPs >0.33 in 413 voxels, followed by JS divergence with EPPs >0.33 in 325 voxels. Meanwhile, the sum of reward probabilities only generated EPPs >0.33 in 14 voxels. In contrast to the SMA, as shown in Figure 5, JS divergence provided a better account of neural activity in a dramatically greater portion of the right IPL than did both the difference between and sum of reward probabilities, with EPPs >0.33 in 342 voxels for JS divergence, in 123 voxels for the difference between probabilities, and in only 2 voxels for the sum. It is of course entirely feasible that more than one computation is being performed in a large anatomical area, and not necessarily the case that the variable that shows superiority in the largest number of voxels is the most essential; in particular when the difference in cluster size is relatively small, and when only one variable generates significant effects using a classical analysis, as is the case here with JS divergence. Nonetheless, as we cannot completely rule out the difference between reward probabilities as the source of our effects in the SMA, we refrain from any further discussion of this area.

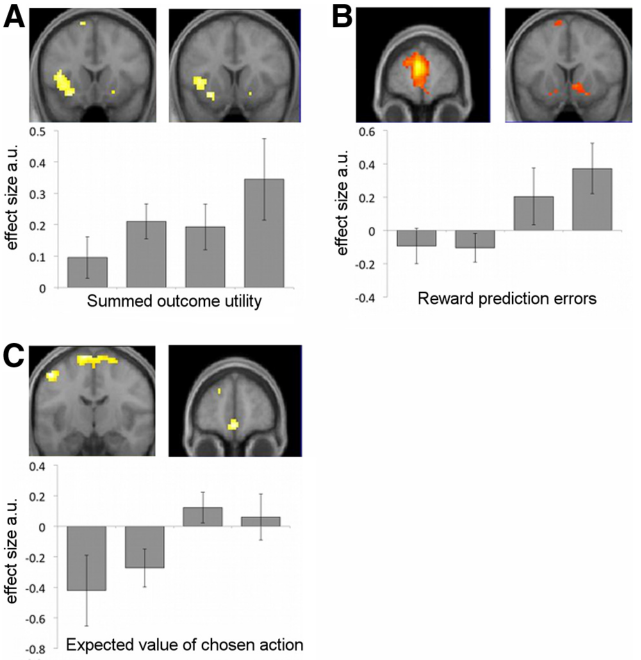

Figure 6. Imaging effects of motivational variables. ***A***, Maps of the *t* statistics for tests of neural modulation by the summed utility of food outcomes obtainable on a given trial, showing effects in the anterior insula and lateral VS. Bar plot shows responses in the left anterior insula. ***B***, Map of the *t* statistics for tests of neural modulation by RPEs during the feedback period of each trial, showing effects throughout the medial PFC and in the medial VS. Bar plot shows responses to RPEs in the VS. ***C***, Map of the *t* statistics for tests of neural modulation by the *Q* value of the chosen action, showing effects in the dorsal SMA and VMPFC. Bar plots show responses in the VMPFC. Bar plots showing mean β-weights across variable values are binned into the 25th, 50th, 75th, and 100th percentiles. Error bars indicate SEM. a.u., arbitrary units.

Stimulus utility and RPEs

During the choice period, the summed utilities (i.e., pleasantness ratings) of the two food rewards that could potentially be obtained on a given trial correlated with activity in the left anterior insula (SVC) and in the left lateral VS (SVC) (Fig. 6*A*). Weaker effects also emerged in the right anterior insula and right lateral VS at $p < 0.005$ uncorrected. No other effects of stimulus utility emerged during this trial period. At the time of outcome delivery, activity in the medial VS was significantly correlated with the RPE (SVC), as was activity throughout the medial frontal and visual cortex (CST; Fig. 6*B*). Notably, there was no effect of the utility of the delivered outcome; however, this lack of result is likely due to the high correlation between this variable and the RPE ($r = 0.7$). To verify that this was the case, we orthogonalized the RPE to outcome utility, rather than the other way around, giving outcome utility the explanatory power. Using this model, a test for the utility of the delivered outcome yielded significant effects in the mOFC (SVC) and throughout the lingual gyrus and calcarine sulcus (CST).

Action values

The expected value of the chosen action during the choice period was significantly correlated with activity in VMPFC (SVC) (Fig. 6*C*), as well as with activity in dorsal SMA, extending throughout the paracentral lobule and into adjacent frontal and parietal areas (CST). No effects were found for the sum of, or difference between, the expected values of available actions.

Discussion

Despite their conceptual appeal and recent popularity, very little is yet known about how well model-based RL theories capture the neural computations underlying goal-directed behavior. In particular, the formalization of Tolman's cognitive map as a probability distribution over instrumental actions and outcomes has remained largely untested. Here, we found that activity in the supramarginal gyrus of the IPL correlated with the divergence of outcome probability distributions associated with available actions. In contrast, activity in the DLPFC varied with the entropy of outcome distributions for chosen actions, while activity in the IPS and LPFC reflected the error-based updating of state transitions. Importantly, these effects were dissociable from those of motivational variables, such as the utility of potential outcomes, which elicited activity in the insula and VS, and the expected (*Q*) value of the chosen action, which correlated with activity in VMPFC. Our findings complement recent data suggesting that animals develop and maintain a rich internal model of their environment (den Ouden et al., 2009; Gläscher et al., 2010; Abe and Lee, 2011).

Although tasks similar to ours have previously been used to address action–outcome learning, our specific hypothesis–that outcome divergence may capture how multiple instrumental relationships are simultaneously contrasted in a meaningful way–is to our knowledge a novel proposal. Our results suggest that the IPL, an area previously implicated in the planning, execution, and observation of goal-directed actions (Fincham et al., 2002; Liljeholm et al., 2011, 2012), as well as in the experience of agency (Chaminade and Decety, 2002; Farrer et al., 2008; Sperduti et al., 2011), implements a comparison of instrumental probability distributions. This finding has broad implications, potentially generalizing to other types of predictive relationships, and providing a means of linking action–outcome learning to more abstract features of goal-directed performance, such as agency and intent attribution.

A closely related topic is how the brain represents various outcome identities (Hamilton and Grafton, 2006, 2008; Stalnaker et al., 2010; Abe and Lee, 2011; Klein-Flügge et al., 2013), over which instrumental divergence can be defined. In a recent neuroimaging study, Klein-Flügge et al. (2013) assessed repetition suppression of BOLD responses to cues that signaled either the same or different food rewards, essentially yielding a low versus high level of outcome divergence. They found that the identities of food outcomes were encoded by the mOFC, with no such effects emerging in the IPL. Notably, Klein-Flügge et al. (2013) eliminated any effects of stimuli that predicted neutral events, to show that mOFC encodes only the identities of rewarding outcomes. In contrast, the anterior IPL has been shown to exhibit repetition suppression of BOLD responses to neutral outcome identities (Hamilton and Grafton, 2006, 2008). Here, we investigate a valence-neutral "cognitive map" of action–outcome contingencies, treating nonreward and rewarding outcome states as equivalent in our computation of divergence. Although several other factors differed across Klein-Flügge et al.'s (2013) task and ours (e.g., our use of probabilistic and instrumental contingencies), we suspect that their exclusion of any areas in which the identities of both neutral and rewarding outcomes were encoded might account for the differences in neural results.

It should be noted that the IPL has been strongly implicated in visuospatial attention and salience; however, such effects tend to emerge in a much more posterior region of the inferior parietal cortex than that identified here (Müri et al., 1996; Gottlieb et al., 1998; Kastner et al., 1999; Corbetta and Shulman, 2002; Husain and Rorden, 2003; Mevorach et al., 2006; Buschman and Miller, 2007; Arcizet et al., 2011; Leathers and Olson, 2012). Indeed, the anterior portion of the supramarginal gyrus identified in the current study has been anatomically established as clearly distinct from more posterior parietal regions (Mars et al., 2011), and has been functionally implicated in the representation of action outcomes with paradigms that largely rule out attentional confounds (Hamilton and Grafton, 2006, 2008).

A few previous neuroimaging studies have used economic decision tasks to separate information theoretic from motivational variables (Luhmann et al., 2008; Abler et al., 2009; Smith et al., 2009; Symmonds et al., 2011). The current experiment differs from such studies in several critical respects: First, in our study, outcome probabilities were acquired through trial-by-trial exposure to action–outcome contingencies, rather than being verbally or graphically instructed–substantial behavioral evidence suggests that decisions based on descriptive information can differ quite dramatically from those based on direct experience (Hertwig, 2012). Second, we used instrumental contingencies, whereas in gambling studies decisions are stimulus based, with stimuli being randomly assigned to particular actions on each trial. Finally, unlike previous studies addressing the entropy, risk, or probability associated with a single decision, we assessed neural encoding of the divergence of outcome distributions associated with simultaneously available action alternatives.

Our approach yields unique insights: to our knowledge, the activity currently observed in the IPL is quite different from parietal activity emerging in gambling paradigms, which has been more posterior, and which has not been unambiguously attributable to probability magnitudes versus entropy or risk (Ernst et al., 2004; Weber and Huettel, 2008; Smith et al., 2009; Symmonds et al., 2011). Our effects in the more anterior portion of the IPL may reflect the use of instrumental contingencies: Gläscher et al. (2009) found that effects in this area were stronger for action-based than for stimulus-based decisions. Another novel contribution of the current study is the dissociation of goal and action values. Both Abler et al. (2009) and Smith et al. (2009) found that activity in the insula increased with reward magnitude, consistent with our effects in this area of the utility of potential food outcomes. However, whereas neither previous study reported any areas of activation for action values beyond those observed for reward magnitude, we found that VMPFC activity increased with the value of the chosen action. This discrepancy between previous studies and ours may reflect selective encoding of experienced over instructed information: Fitzgerald et al. (2010) found greater VMPFC activation for experientially acquired than for described value signals.

Other previous work has directly investigated neural correlates of model-based versus model-free RL. However, these studies focus primarily on value signals (e.g., Daw et al., 2005), without exploring the possibility that valence-neutral state transitions are represented independently of associated motivational features. One notable exception is a study by Simon and Daw (2011) in which the degree of model-based branching, essentially the complexity of forward planning, was defined as the number of choices available in the current state, as well as the expected number of choices in subsequent states. They found positive neural correlates of these measures in the lateral precentral cortex, the anterior insula, and the anterior cingulate/SMA, but not in the IPL. There are, however, two critical differences between their analyses and ours: First, they were modeling the number of available actions, rather than contingent outcome states; this distinction is particularly important given previous findings that the right anterior IPL distinguishes between outcome identities, but not between action kinematics (Hamilton and Grafton, 2006, 2008). Second, Simon and Daw modeled the number of expected future options summed across, rather than as a divergence across, currently available options. It is not surprising, therefore, that the results differed substantially from those obtained here.

Goal-directed performance is characterized by a sensitivity to changes in the instrumental contingency that has been reliably demonstrated in rodents as well as humans (Hammond, 1980; Shanks and Dickinson, 1991; Balleine and Dickinson, 1998; Liljeholm et al., 2011). In a previous study, using a free-operant task, in which the rate of executing a single rewarded action is self-paced, we found that activity in the IPL correlated with changes in instrumental contingency, formalized as the difference between probabilities of reward in the presence versus absence of an action (Liljeholm et al., 2011). Unlike outcome divergence, instrumental contingency conflates the probabilities and values of outcomes. Moreover, instrumental contingency is signed, reflecting the relative advantage of performing or not performing a particular action (indeed, in our previous study, activity in the left IPL was found to correlate negatively with instrumental contingency, perhaps reflecting the relative advantage of withholding a response). Nonetheless, outcome divergence can be characterized as a general, symmetric, extension of instrumental contingency to the case of multiple actions and outcomes. As such, our current demonstration of a role for the IPL in encoding divergence is consistent with our previous work implicating this area in the probabilistic integration of action alternatives.

In conclusion, we show modulation of the IPL by the divergence of outcome probability distributions associated with alternative actions, and dissociate this decision variable from both stimulus utilities and action values. As applied here, JS divergence reflects the extent to which discrimination between available actions has any impact on the occurrence, and predictability, of future states. Conversely, this information theoretic measure captures the attributability of a current environment to distinct antecedent actions. As such, it is likely to play a central role in goal-directed encoding of action–outcome contingencies, a suggestion that is supported by the current findings.

References

Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron 70:731–741. CrossRef Medline

Abler B, Herrnberger B, Grön G, Spitzer M (2009) From uncertainty to reward: BOLD characteristics differentiate signaling pathways. BMC Neurosci 10:154. CrossRef Medline

Arcizet F, Mirpour K, Bisley JW (2011) A pure salience response in posterior parietal cortex. Cereb Cortex 21:2498–2506. CrossRef Medline

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419. CrossRef Medline

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. Behav Sci 9:226–232. CrossRef Medline

Blodgett HC (1929) The effect of the introduction of reward upon the maze performance of rats. University of California Publications in Psychology 4:113–134.

Buschman TJ, Miller EK (2007) Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science 315: 1860–1862. CrossRef Medline

Chaminade T, Decety J (2002) Leader or follower? Involvement of the inferior parietal lobule in agency. Neuroreport 13:1975–1978. CrossRef Medline

Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201–215. Medline

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

den Ouden HE, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A dual role for prediction error in associative learning. Cereb Cortex 19: 1175–1185. Medline

Doya K, Samejima K, Katagiri K, Kawato M (2002) Multiple model-based reinforcement learning. Neural Comput 14:1347–1369. CrossRef Medline

Ernst M, Nelson EE, McClure EB, Monk CS, Munson S, Eshel N, Zarahn E, Leibenluft E, Zametkin A, Towbin K, Blair J, Charney D, Pine DS (2004) Choice selection and reward anticipation: an fMRI study. Neuropsychologia 42:1585–1597. CrossRef Medline

Esterman M, Tamber-Rosenau BJ, Chiu YC, Yantis S (2010) Avoiding non-independence in fMRI data analysis: leave one subject out. Neuroimage 50:572–576. CrossRef Medline

Farrer C, Frey SH, Van Horn JD, Tunik E, Turk D, Inati S, Grafton ST (2008) The angular gyrus computes action awareness representations. Cereb Cortex 18:254–261. Medline

Feinstein JS, Stein MB, Paulus MP (2006) Anterior insula reactivity during certain decisions is associated with neuroticism. Soc Cogn Affect Neurosci 1:136–142. CrossRef Medline

Fincham JM, Carter CS, van Veen V, Stenger VA, Anderson JR (2002) Neural mechanisms of planning: a computational analysis using event-related fMRI. Proc Natl Acad Sci U S A 99:3346–3351. CrossRef Medline

Fitzgerald TH, Seymour B, Bach DR, Dolan RJ (2010) Differentiable neural substrates for learned and described value and risk. Curr Biol 20:1823–1829. CrossRef Medline

Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn Reson Med 33:636–647. CrossRef Medline

Garner DM, Olmsted MP, Bohr Y, Garfinkel PE (1982) The eating attitudes test: psychometric features and clinical correlates. Psychol Med 12:871–878. CrossRef Medline

Gläscher J (2009) Visualization of group inference data in functional neuroimaging. Neuroinformatics 7:73–82. 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

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595. CrossRef Medline

Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual salience in monkey parietal cortex. Nature 391:481–484. CrossRef Medline

Hamilton AF, Grafton ST (2006) Goal representation in human anterior intraparietal sulcus. J Neurosci 26:1133–1137. CrossRef Medline

Hamilton AF, Grafton ST (2008) Action outcomes are represented in human inferior frontoparietal cortex. Cereb Cortex 18:1160–1168. Medline

Hammond LJ (1980) The effect of contingency upon the appetitive conditioning of free-operant behavior. J Exp Anal Behav 34:297–304. CrossRef Medline

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. CrossRef Medline

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. CrossRef Medline

Hertwig R (2012) The psychology and rationality of decisions from experience. Synthese 187:269–292. CrossRef

Husain M, Rorden C (2003) Non-spatially lateralized mechanisms in hemispatial neglect. Nat Rev Neurosci 4:26–36. CrossRef Medline

Kastner S, Pinsk MA, De Weerd P, Desimone R, Ungerleider LG (1999) Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22:751–761. CrossRef Medline

Klein-Flügge MC, Barron HC, Brodersen KH, Dolan RJ, Behrens TE (2013) Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. J Neurosci 33:3202–3211. CrossRef Medline

Knoch D, Gianotti LR, Pascual-Leone A, Treyer V, Regard M, Hohmann M, Brugger P (2006) Disruption of right prefrontal cortex by low-frequency repetitive transcranial magnetic stimulation induces risk-taking behavior. J Neurosci 26:6469–6472. CrossRef Medline

Leathers ML, Olson CR (2012) In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. Science 338:132–135. CrossRef Medline

Liljeholm M, Tricomi E, O'Doherty JP, Balleine BW (2011) Neural correlates of instrumental contingency learning: differential effects of action-reward conjunction and disjunction. J Neurosci 31:2474–2480. CrossRef Medline

Liljeholm M, Molloy CJ, O'Doherty JP (2012) Dissociable brain systems mediate vicarious learning of stimulus-response and action-outcome contingencies. J Neurosci 32:9878–9886. CrossRef Medline

Lin J (1991) Divergence measures based on the Shannon entropy. IEEE Trans Inform Theory 37:145–151.

Luhmann CC, Chun MM, Yi DJ, Lee D, Wang XJ (2008) Neural dissociation of delay and uncertainty in intertemporal choice. J Neurosci 28: 14459–14466. CrossRef Medline

Mars RB, Jbabdi S, Sallet J, O'Reilly JX, Croxson PL, Olivier E, Noonan MP, Bergmann C, Mitchell AS, Baxter MG, Behrens TE, Johansen-Berg H, Tomassini V, Miller KL, Rushworth MF (2011) Diffusion-weighted imaging tractography-based parcellation of the human parietal cortex and comparison with human and macaque resting-state functional connectivity. J Neurosci 31:4087–4100. CrossRef Medline

Mevorach C, Humphreys GW, Shalev L (2006) Opposite biases in salience-based selection for the left and right posterior parietal cortex. Nat Neurosci 9:740–742. CrossRef Medline

Müri RM, Iba-Zizen MT, Derosier C, Cabanis EA, Pierrot-Deseilligny C (1996) Location of the human posterior eye field with functional magnetic resonance imaging. J Neurol Neurosurg Psychiatry 60:445–448. CrossRef Medline

O'Doherty JP, Deichmann R, Critchley HD, Dolan RJ (2002) Neural responses during anticipation of a primary taste reward. Neuron 33:815–826. CrossRef Medline

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ (2003) Temporal difference models and reward-related learning in the human brain. Neuron 38:329 –337. CrossRef Medline

O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F (2001) Representation of pleasant and aversive taste in the human brain. J Neurophysiol 85:1315–1321. Medline

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

Paulus MP, Hozack N, Frank L, Brown GG (2002) Error rate and outcome predictability affect neural activation in prefrontal cortex and anterior cingulate during decision-making. Neuroimage 15:836 – 846. CrossRef Medline

Pelchat ML, Johnson A, Chan R, Valdez J, Ragland JD (2004) Images of desire: food-craving activation during fMRI. Neuroimage 23:1486 –1493. CrossRef Medline

Penny WD, Trujillo-Barreto NJ, Friston KJ (2005) Bayesian fMRI time series analysis with spatial priors. Neuroimage 24:350 –362. CrossRef Medline

Plassman H, O'Doherty JP, Rangel A (2010) Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. J Neurosci 30:10799 –10808.

Rosa MJ, Bestmann S, Harrison L, Penny W (2010) Bayesian model selection maps for group studies. Neuroimage 49:217–224. CrossRef Medline

Schmidt L, Lebreton M, Cléy-Melin ML, Daunizeau J, Pessiglione M (2012) Neural mechanisms underlying motivation of mental versus physical effort. PLoS Biol 10:e1001266. CrossRef Medline

Shanks DR, Dickinson A (1991) Instrumental judgment and performance under variations in action-outcome contingency and contiguity. Mem Cognit 19:353–360. CrossRef Medline

Simon DA, Daw ND (2011) Neural correlates of forward planning in a spatial decision task in humans. J Neurosci 31:5526 –5539.

Smith BW, Mitchell DG, Hardin MG, Jazbec S, Fridberg D, Blair RJ, Ernst M (2009) Neural substrates of reward magnitude, probability, and risk during a wheel of fortune decision-making task. Neuroimage 44:600 – 609. CrossRef Medline

Sperduti M, Delaveau P, Fossati P, Nadel J (2011) Different brain structures related to self- and external-agency attribution: a brief review and meta-analysis. Brain Struct Funct 216:151–157. CrossRef Medline

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2010) Neural correlates of stimulus-response and response-outcome associations in dorsolateral versus dorsomedial striatum. Front Integr Neurosci 4:12. Medline

Symmonds M, Wright ND, Bach DR, Dolan RJ (2011) Deconstructing risk: separable encoding of variance and skewness in the brain. Neuroimage 58:1139 –1149. CrossRef Medline

Tanaka SC, Balleine BW, O'Doherty JP (2008) Calculating consequences: brain systems that encode the causal effects of actions. J Neurosci 28: 6750 – 6755. CrossRef Medline

Thorndike EL (1933) A proof of the Law of Effect. Science 77:173–175. CrossRef Medline

Tolman EC (1948) Cognitive maps in rats and men. Psychol Rev 55:189 –208. CrossRef Medline

Tolman EC, Honzik CH (1930) "Insight" in rats. University of California, Publications in Psychology 4:215–232.

Wang GJ, Volkow ND, Telang F, Jayne M, Ma J, Rao M, Zhu W, Wong CT, Pappas NR, Geliebter A, Fowler JS (2004) Exposure to appetitive food stimuli markedly activates the human brain. Neuroimage 21: 1790 –1797. CrossRef Medline

Weber BJ, Huettel SA (2008) The neural substrates of probabilistic and intertemporal decision making. Brain Res 1234:104 –115. CrossRef Medline