Behavioral/Systems/Cognitive

Distributed Neural Representation of Expected Value

Brian Knutson,[1] Jonathan Taylor,[2] Matthew Kaufman,[1] Richard Peterson,[1] and Gary Glover[3]
Departments of [1]Psychology, [2]Statistics, and [3]Radiology, Stanford University, Stanford, California 94305

Anticipated reward magnitude and probability comprise dual components of expected value (EV), a cornerstone of economic and psychological theory. However, the neural mechanisms that compute EV have not been characterized. Using event-related functional magnetic resonance imaging, we examined neural activation as subjects anticipated monetary gains and losses that varied in magnitude and probability. Group analyses indicated that, although the subcortical nucleus accumbens (NAcc) activated proportional to anticipated gain magnitude, the cortical mesial prefrontal cortex (MPFC) additionally activated according to anticipated gain probability. Individual difference analyses indicated that, although NAcc activation correlated with self-reported positive arousal, MPFC activation correlated with probability estimates. These findings suggest that mesolimbic brain regions support the computation of EV in an ascending and distributed manner: whereas subcortical regions represent an affective component, cortical regions also represent a probabilistic component, and, furthermore, may integrate the two.

Key words: **gain; loss; reward; probability; monetary; striatum; accumbens; mesial prefrontal cortex; fMRI; human**

Introduction

All organisms, from bees to bridge builders, must forecast the future to decide what to do next (Montague and Berns, 2002). To choose optimally, they must consider both the potential value of different courses of action and the probability (PRB) that each will lead to a desired outcome. Blaise Pascal formalized the product of these two considerations in the concept of "expected value" (EV), which has played a prominent role in both economic and psychological theory (von Neumann and Morgenstern, 1944; Rotter, 1972; Bandura, 1977). Although EV is often inferred from choice behavior, behavioral evidence suggests that people do not always act to maximize EV (Kahneman and Tversky, 1984). Nonetheless, expectations about the magnitude (MAG) and probability of valued outcomes probably constitute important considerations in decision making. Currently, the extent to which neural mechanisms that represent these variables are unitary or separable remains unclear. Additionally, although computation of EV has traditionally been conceptualized as involving reflective acts of deliberation, recent evidence suggests that affective reactions may also figure into the equation (Loewenstein et al., 2001; Slovic et al., 2002).

The goal of this study was to identify neural substrates that support the computation of EV and to characterize their psychological functions. By definition, the brain must compute EV before incentive outcomes. The spatial and temporal resolution of event-related functional magnetic resonance imaging (fMRI) now offers researchers an opportunity to ask where and when these computations might occur. Although previous fMRI research has examined brain activity during anticipation of incentives (Breiter et al., 2001; Knutson et al., 2001a; O'Doherty et al., 2002), investigators have not yet simultaneously varied expected incentive valence (VAL), magnitude, and probability and observed their combined effect on anticipatory brain activation. Overall, we predicted that EV-computing regions should show increased activity during anticipation of large-magnitude, high-probability gains.

Comparative theorists have proposed that regions innervated by mesolimbic dopamine projections play a critical role in the computation of EV (Gallistel, 1986; Shizgal, 1997); therefore, specific predictions focused on these mesolimbic regions. Because fMRI studies of humans indicate that anticipation of monetary gains proportionally increases blood oxygen level-dependent contrast (hereafter, "activation") in the subcortical nucleus accumbens (NAcc) (Knutson et al., 2001a), we predicted that anticipation of large-magnitude gains (but not losses) would increase NAcc activation. Because fMRI studies also indicate that gain outcomes instead activate the mesial prefrontal cortex (MPFC) (Delgado et al., 2000; Elliott et al., 2000; Knutson et al., 2001b) and because outcomes involve a collapse of probability, we predicted that increased gain probability might increase MPFC activation, even during anticipation. Because ventral tegmental area midbrain dopamine neurons that project to the NAcc and MPFC have also been implicated in the computation of EV (Schultz et al., 1997), we also examined anticipatory midbrain activity. Finally, we investigated whether activity in these regions covaried with individual differences in positive aroused affect (e.g., feelings of excitement) evoked by gain cues (Knutson et al., 2001a; Bjork et al., 2004), as well as the perceived probability of obtaining gains.

Received Sept. 28, 2004; revised March 25, 2005; accepted April 9, 2005.

This work was funded by National Institute of Mental Health Grant R03-MH066923 and a National Alliance for Research on Schizophrenia and Depression Young Investigator award (B.K.), as well as National Center for Research Resources Grant RR-09784 (G.G.).We thank Jamil Bhanji, Anna Chen, and Christopher Smith for assistance in collecting and analyzing data, as well as Antonio Rangel, Piotr Winkielman, and two anonymous reviewers for comments on drafts of this manuscript.

Correspondence should be addressed to Brian Knutson, Stanford University, Building 420, Jordan Hall, Stanford, CA 94305-2130. E-mail: knutson@psych.stanford.edu.

DOI:10.1523/JNEUROSCI.0642-05.2005

 0270-6474/05/254806-07$15.00/0

Materials and Methods

Fourteen healthy volunteers (eight women; right-handed; mean age, 22) participated in the study. Before enrolling, volunteers were screened for physical and mental disorders (including neurological damage and abnormal cardiac function) via a medical interview. All subjects gave written informed consent, and the experiment was approved by the Institutional Review Board of the Stanford University Medical School. Before entering the scanner, subjects received a verbal description of the task, completed a practice version (10 min) to minimize learning effects in the scanner, and were tested on what each of the incentive cues indicated. Subjects were also shown the money that they could earn by performing the task successfully in the scanner, and all reported believing that they would receive cash contingent on their performance at the end of the experiment. Once in the scanner, anatomical scans were acquired. Subjects then engaged in four 10 min blocks of the incentive task during functional scan acquisition. After the scan, subjects rated their affective reactions to the different cues on seven-point Likert scales (i.e., valence from "bad" to "good" and arousal from "not at all aroused" to "highly aroused"), estimated the probability of success for each cue type, and were tested a second time to ensure that they understood the meaning of each cue.

Probabilistic monetary incentive delay task. The probabilistic variant of the monetary incentive delay (MID) task included 288 8 s trials. During each trial, subjects saw 1 of 18 shapes formed by varying three features (cue, 2000 ms), focused on a fixation cross (X) while waiting for a variable anticipatory delay period (anticipation, 2000–2500 ms), and then responded with a button press to the appearance of a solid white square (target, 160–310 ms). After a short delay, subsequent feedback notified subjects if they had won or lost money, as well as their cumulative total (outcome, 2000 ms) (Fig. 1). Each of 18 trial types was presented 16 times in an individually randomized order.

Cue features were counterbalanced and signaled the incentive value of each trial: shape (circle or square) indicated valence (potential gain or loss), position of a vertical line (left, middle, or right) indicated magnitude ($0.00, $1.00, or $5.00), and position of a horizontal line (high, middle, or low) indicated probability of success (high, medium, or low). Expected probability was manipulated by altering the average target duration via an adaptive timing algorithm that followed subjects' performance, such that they would succeed on ~80% of high-probability trials, 50% of medium-probability trials, and 20% of low-probability trials overall. Before playing, subjects were told which cue features indicated that they had a high, medium, or low chance of success but were not given numerical percentage estimates. fMRI volume acquisitions were time locked to cue onsets and thus coincided with each phase of each trial.

fMRI acquisition. Imaging was performed using a 1.5 T General Electric (Milwaukee, WI) MRI scanner with a standard quadrature head coil. Twenty-four 4-mm-thick slices (in-plane resolution, 3.75 × 3.75 mm; no gap) extended axially from the midpons to the top of the skull, which provided adequate spatial resolution of subcortical regions of interest (e.g., midbrain, ventral striatum) and omitted only the base of the cerebellum or crown of the skull in some subjects. Functional scans of the entire brain were acquired every 2 s [repetition time (TR), 2 s] with a T2*-sensitive in/out spiral pulse sequence [echo time (TE), 40 ms; flip, 90°] specifically designed to minimize signal dropout at the base of the brain (Glover and Law, 2001). High-resolution structural scans were subsequently acquired using a T1-weighted spoiled grass sequence (TR, 100 ms; TE, 7 ms; flip, 90°), which facilitated subsequent localization and coregistration of functional data.

fMRI analysis. Analyses focused on changes in activation specifically during anticipatory delay periods for both hit and miss trials (i.e., after subjects saw cues but before they responded to targets). All analyses were conducted using analysis of functional neural images software (Cox, 1996). For preprocessing, voxel time series were sinc interpolated to correct for nonsimultaneous slice acquisition within each volume, concatenated across runs, and corrected for three-dimensional motion. Visual inspection of motion correction estimates confirmed that no subject's head moved >2.0 mm in any dimension from one volume acquisition to the next. Data were preprocessed via bandpass filtering

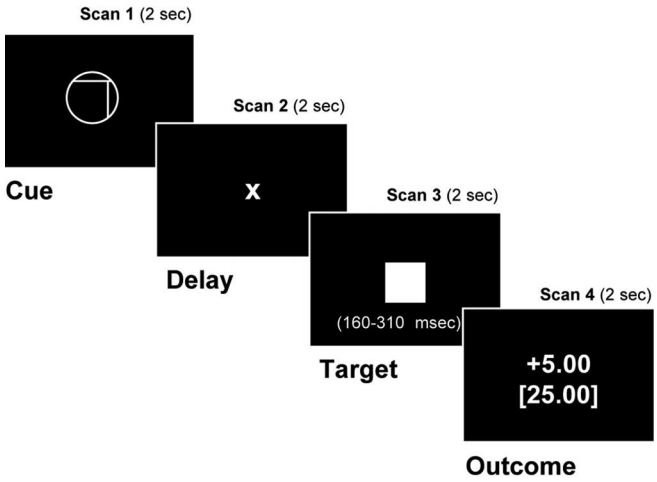

Figure 1. Probabilistic monetary-incentive delay-task trial structure.

(admitting frequencies from 8 to 90 s) and computation of percentage signal change (calculated with respect to the mean activation over the entire experiment in each voxel).

Preprocessed data were submitted to regression models, which included six regressors indexing residual motion, 12 regressors modeling baseline, linear, and quadratic trends for each of the four runs, one regressor of noninterest, which modeled general anticipation effects by unit weighting each anticipatory period, and a second regressor of noninterest, which modeled motor preparation by weighting each anticipatory period with its subsequent associated reaction time. Both group analyses indicated that the motor preparation regressor correlated positively with foci in the left putamen [$Z = 3.72$; Talairach coordinates (TCs), $-13, 4, -6$ (x, y, z)], supplementary motor area ($Z = 4.30$; TCs, 0, $-8, 50$), and motor cortex ($Z = 4.77$; TCs, $-42, -22, 48$) and correlated negatively with a focus in the right inferior frontal gyrus ($Z = -3.69$; TCs, 48, 18, 4).

Group analyses consisted of two types: localization and decomposition. The goal of the localization analysis was to identify candidate regions whose activity correlated with a linear model of EV, whereas the goal of the decomposition analysis was to verify which of these regions maximally correlated with independent components of EV. For the localization analysis, the regressor of interest was a linear model of cued EV (i.e., the product of cued incentive valence, magnitude, and probability). For the decomposition analysis, fully orthogonalized regressors of interest contrasted cued (1) positive versus negative VAL, (2) $5.00 versus $0.00 MAG, (3) 80 versus 20% PRB, (4) VAL by MAG, (5) VAL by PRB, (6) MAG by PRB, and (7) VAL by MAG by PRB (EV). Because of the hypothetical focus on anticipation and a lack of statistical power, brain activation related to incentive outcomes was not analyzed.

For group analyses, regressors of noninterest and interest were convolved with a γ-variate function that modeled a canonical hemodynamic response before regression (Cohen, 1997). Maps of *t*-statistics for regressors of interest were transformed into *Z* scores, coregistered with structural maps, spatially normalized by warping to Talairach space, slightly spatially smoothed (full width at half maximum, 4 mm) to minimize the effects of anatomical variability, resampled at 2 mm^3, and combined into a group map using a meta-analytic formula [average $Z \times \sqrt{(n)}$] (Knutson et al., 2000). For the decomposition analysis, foci activation thresholds were determined by false discovery rate ($Z > 3.24$; $p < 0.05$, corrected) (Genovese et al., 2002) and required a minimum cluster of four contiguous voxels. For the more exploratory localization analysis, foci activation thresholds were set an order of magnitude higher ($Z > 3.88$; $p < 0.0001$, uncorrected) and required a minimum cluster of four contiguous voxels.

Individual analyses were also conducted to verify the results of the decomposition analysis within each identified region. For individual analyses, volumes of interest (VOIs) were specified by imposing 8-mm-

Table 1. Regressor of interest *Z* scores and Talairach coordinates for peak activation foci

	Regressor				
Region	Valence	Magnitude	VAL by MAG	Probability	VAL by MAG by PRB
Left orbital frontal cortex		5.94 (−22, 41, −11)			
Right orbital frontal cortex		5.75 (22, 41, −10)			
Left dorsolateral PFC	*−4.71 (38, 22, 46)*	10.22 (−38, 52, 8)	4.41 (−42, 46, 1)		
Right dorsolateral PFC		10.75 (37, 53, 10)			
Left mesial PFC		*−3.75 (−1, 46, −11)*	5.00 (−4, 48, −3)	4.49 (−4, 52, −6)	
Right mesial PFC		*−3.31 (4, 53, −10)*	4.60 (4, 45, 0)	5.08 (4, 52, −10)	
Left anterior cingulate		14.66 (0, 22, 34)	4.59 (−4, 38, 16)	*−4.17 (0, 22, 42)*	4.36 (−4, 45, 27)
Right anterior cingulate			4.61 (1, 22, 35)		4.07 (1, 42, 27)
Supplementary motor area	4.42 (0, −8, 45)	18.94 (0, 0, 46)	5.02 (−4, 0, 42)	*−4.64 (−2, −12, 50)*	
Left insula		10.65 (−30, 19, 1)		*−3.90 (−30, 18, 4)*	
Right insula		12.00 (30, 19, 8)			
Left NAcc	4.51 (−11, 11, −2)	21.46 (−11, 8, 0)	5.16 (−15, 11, 0)		
Right NAcc	4.74 (11, 11, −2)	21.62 (11, 11, 0)	5.89 (14, 14, 0)		
Left caudate		21.30 (−7, 7, 4)	4.61 (−12, −7, 20)	*−5.02 (−11, 8, 5)*	
Right caudate		24.45 (7, 7, 4)	4.66 (7, −7, 20)	*−6.14 (15, 12, 5)*	3.46 (15, 0, 23)
Left putamen			5.65 (−19, 10, 0)		
Right putamen			5.89 (16, 14, 0)	*−5.16 (19, 8, −4)*	
Thalamus		22.91 (0, −14, 16)	5.68 (0, −15, 16)		
Parahippocampal gyrus				4.16 (−22, −12, −10)	
Midbrain		13.79 (0, −18, −10)			
Posterior cingulate				4.49 (1, −48, 24)	

Documented foci extend to $A = -55$. Foci passed a $Z > 3.24$ threshold ($n = 14$; $p < 0.05$; false detection rate); minimum cluster, four contiguous voxels. Targeted foci are underlined; negative foci are italicized.

diameter spheres at peak activation foci in regions of interest in the midbrain, NAcc, MPFC, and anterior cingulate (Table 1). Activation time courses were extracted and averaged from these VOIs by trial type, after ensuring that anticipatory activation did not differ by hit-versus-miss outcomes with paired comparisons. Peak activation values at a 4 s lag were then analyzed with a 2 (VAL) by 3 (MAG) by 3 (PRB) ANOVA for each VOI, and main effects of MAG and PRB were tested for both linear and quadratic trends. Significant effects were verified by comparing incentive with nonincentive conditions matched for both valence and probability using paired comparisons ($p < 0.005$, corrected for nine comparisons).

Behavior and affect. Hits were calculated as percentage of correct responses per condition (i.e., the button press occurred during target presentation). Ratings of cue-elicited valence and arousal were mean-deviated within subjects and plotted within a Euclidean two-dimensional space. These dimensions were then rotated by 45° to derive measures of positive arousal (PA) [i.e., PA = arousal/$\sqrt{(2)}$ + valence/$\sqrt{(2)}$] and negative arousal (NA) [NA = arousal/$\sqrt{(2)}$ − valence/$\sqrt{(2)}$] (Watson et al., 1999). Actual hit rate, estimated hit rate, cue-elicited PA, and cue-elicited NA for each trial type were analyzed with a 2 (VAL) by 3 (MAG) by 3 (PRB) repeated-measures ANOVA. Significant main effects or interactions were verified with Tukey's honestly significant difference (HSD) tests ($p < 0.05$, corrected). Cue-elicited affect and estimated probability of success were also correlated with activation in VOIs across individuals.

Manipulation checks established that cues had their intended influence on behavior and affect. Behaviorally, hit rate showed a main effect for PRB ($F_{(2,26)} = 1348.57$; $p < 0.001$), establishing the efficacy of the adaptive timing manipulation. Mean hit rates (±SEM) were 23.23% (0.66) for 20% trials, 52.01% (0.80) for 50% trials, and 76.67% (0.58) for 80% trials. Subjects' hit-rate estimates also showed a main effect of PRB ($F_{(2,26)} = 104.42$; $p < 0.001$), indicating that they were explicitly aware of cue meaning. Mean hit-rate estimates (±SEM) were 16.39% (3.22) for 20% trials, 47.5% (3.66) for 50% trials, and 73.7% (4.18) for 80% trials. Additionally, reaction time for hits did not differ between gain and loss trials of equal magnitude, as illustrated by main effects of MAG ($F_{(2,26)} = 8.22$; $p < 0.005$) and PRB ($F_{(2,20)} = 16.99$; $p < 0.001$), but not of VAL, on reaction time.

In contrast, the combination of cue VAL × MAG elicited differential affective responses, as predicted, and consistent with midbrain and NAcc activation patterns (see Results). Cue-elicited positive arousal showed main effects of VAL ($F_{(1,13)} = 83.29$; $p < 0.001$) and MAG ($F_{(2,26)} = 55.34$; $p < 0.001$) with linear and quadratic trends, qualified by the predicted interaction of VAL by MAG ($F_{(2,20)} = 34.68$; $p < 0.001$), as well as a separate interaction of MAG by PRB ($F_{(4,52)} = 6.65$; $p < 0.005$). Tukey's HSD tests indicated that subjects responded with more positive arousal to gain and loss cues than to nonincentive cues and with more positive arousal to gain cues than to loss cues of equivalent magnitude. Cue-elicited negative arousal showed main effects of VAL ($F_{(1,13)} = 99.26$; $p < 0.001$), MAG ($F_{(2,26)} = 44.10$; $p < 0.001$) with linear and quadratic trends, and PRB ($F_{(2,26)} = 26.01$; $p < 0.001$) with a linear trend, qualified by the predicted interaction of VAL by MAG ($F_{(2,20)} = 40.92$; $p < 0.001$). Tukey's HSD tests indicated that subjects reacted with greatest negative arousal to large- and small-magnitude loss cues relative to all other cues and to cues indicating a low probability of success relative to medium- and high-probability cues.

Results

Group localization analyses indicated that the predicted mesolimbic regions showed activity correlated with a linear model of EV. Specifically, foci in the midbrain ($Z = 4.76$; TCs, 0, −15, −11), bilateral NAcc (right, $Z = 5.83$, TCs, 11, 11, −3; left, $Z = 4.09$, TCs, −8, −11, −1), and left MPFC ($Z = 5.63$; TCs, −4, 50, 0) showed activation correlated with linear EV (Fig. 2). Additionally, subcortical foci in the thalamus ($Z = 5.23$; TCs, 7, −15, 16), bilateral putamen (right, $Z = 6.06$, TCs, 14, 7, −7; left, $Z = 5.93$, TCs, −15, 7, −3), right caudate ($Z = 4.54$; TCs, 48, 11, 5), and right insula ($Z = 4.54$; TCs, 48, 11, 5) showed activation correlated with linear EV. Cortical foci in the right orbitofrontal cortex ($Z = 4.26$; TCs, 22, 33, −11), left dorsolateral prefrontal cortex ($Z = 4.99$; TCs, −18, 44, 34), left anterior cingulate cortex ($Z = 4.44$; TCs, −8, 30, 27), and supplementary motor area ($Z = 4.37$; TCs, −3, 0, 42) also showed activation correlated with linear EV. Although the localization analysis established that activation in these regions correlated with the linear model of EV, it could not establish which components of EV correlated with regional activity.

Decomposition analyses did, however, indicate that distinct mesolimbic regions preferentially responded to different components of EV (Table 1, underlined regions). Specifically, the main effect of VAL correlated with activation foci in the left anterior

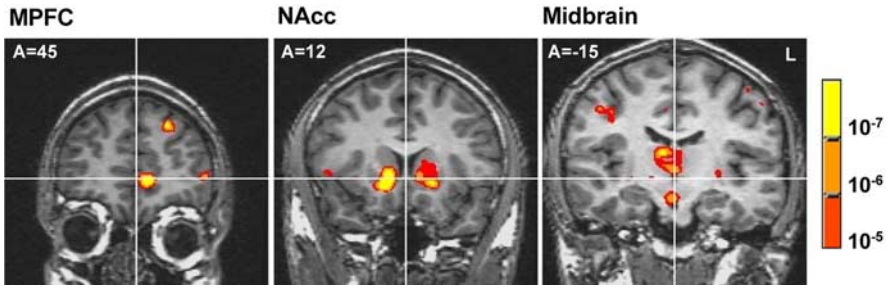

Figure 2. Group maps of regions whose activation correlates with the linear model of expected value. Warm colors signify activation, whereas cool colors signify deactivation (threshold, $p < 10^{-5}$). A, Anterior; L, left.

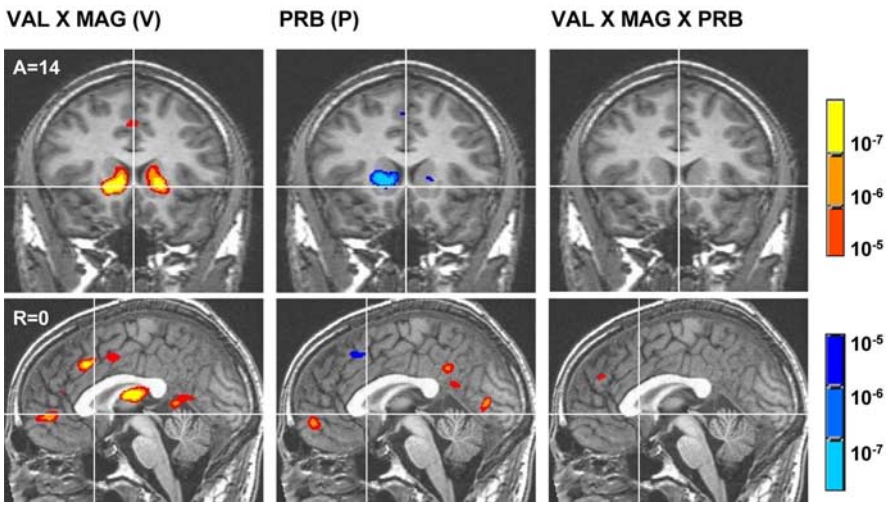

Figure 3. Group maps of contrasts related to distinct terms of expected value. Maps illustrate the interaction of VAL by MAG (V), the main effect of PRB (P), and the interaction of VAL by MAG by PRB (EV). Warm colors signify activation, whereas cool colors signify deactivation (threshold, $p < 10^{-5}$). A, Anterior; R, right.

cingulate and bilateral NAcc and negatively correlated with an activation focus in the left dorsolateral prefrontal cortex. The main effect of MAG correlated with several activation foci in the bilateral orbitofrontal cortex, dorsolateral prefrontal cortex, anterior cingulate, supplementary motor area, insula, NAcc, medial caudate, thalamus, and midbrain. The main effect of MAG also negatively correlated with activation foci in the bilateral MPFC. The critical interaction of VAL by MAG correlated with activation foci in the bilateral NAcc as predicted, as well as the bilateral MPFC, left dorsolateral prefrontal cortex, bilateral anterior cingulate, bilateral medial caudate, bilateral putamen, dorsomedial thalamus, and supplementary motor area, replicating previous findings (Knutson et al., 2001a).

In contrast, the main effect of PRB correlated with activation foci in the bilateral MPFC as predicted, as well as the left parahippocampal gyrus and posterior cingulate. Additionally, the main effect of PRB negatively correlated with activation foci in the left anterior cingulate, left insula, bilateral medial caudate, and right putamen. The interaction of VAL by PRB did not correlate with activation in any of these regions, whereas the interaction of MAG by PRB was only correlated with activation foci in the bilateral medial caudate and right putamen (data not listed in Table 1). Finally, the interaction of VAL by MAG by PRB correlated only with activation foci in the bilateral anterior cingulate and right tail of the caudate (Fig. 3).

For individual analyses, percentage signal change was extracted from VOIs centered on mesolimbic foci evident in both localization and decomposition analyses. These included activation foci in the midbrain (corresponding to MAG), NAcc (corresponding to VAL by MAG), MPFC (corresponding to PRB), and anterior cingulate (corresponding to VAL by MAG by PRB). Analysis of peak activation in the midbrain VOI yielded main effects of VAL ($F_{(1,13)} = 6.15$; $p < 0.05$) and MAG ($F_{(2,26)} = 26.06$; $p < 0.001$) with a linear trend, qualified by an interaction of VAL by MAG ($F_{(2,26)} = 11.89$; $p < 0.001$). Pairwise comparisons indicated that midbrain activation was greater for anticipation of all large gains and 50% large losses than for matched nonincentive cues. Thus, for gains, the midbrain showed activation consistent with representation of the value term of EV but not the probability term (Fig. 4).

Analysis of peak activation in the left NAcc VOI yielded main effects for VAL ($F_{(1,13)} = 7.74$; $p < 0.05$) and MAG ($F_{(2,26)} = 32.12$; $p < 0.001$) with a linear trend, qualified by the predicted interaction of VAL by MAG ($F_{(2,26)} = 11.89$; $p < 0.001$). Analysis of peak activation in the right NAcc VOI also yielded main effects of VAL ($F_{(1,13)} = 6.22$; $p < 0.05$) and MAG ($F_{(2,26)} = 49.05$; $p < 0.001$) with a linear trend, qualified by an interaction of VAL by MAG ($F_{(2,26)} = 13.96$; $p < 0.001$). Paired comparisons indicated that bilateral NAcc activation was greater for anticipation of all large gains and 20% large losses than for matched nonincentives. Thus, for gains, the bilateral NAcc also showed activity consistent with representation of the value term of EV but not the probability term.

Analysis of peak activation in the left MPFC VOI (associated with PRB) yielded the predicted main effect of PRB ($F_{(2,26)} = 6.18$; $p < 0.01$) with a linear trend, as well as a separate interaction of VAL by MAG ($F_{(2,26)} = 9.36$; $p < 0.01$). Analysis of peak activation in the right MPFC VOI also yielded a main effect of PRB ($F_{(2,26)} = 3.38$; $p < 05$) with a linear trend, as well as a main effect of VAL ($F_{(1,13)} = 6.52$; $p < 0.05$), qualified by an interaction of VAL by MAG ($F_{(2,26)} = 5.46$; $p < 0.05$). Pairwise comparisons indicated that bilateral MPFC activation was greater only for anticipation of 80% large gains relative to matched nonincentives. Thus, the bilateral MPFC showed activity consistent with representation of the probability as well as the value terms of EV.

Analysis of peak activation in the left anterior cingulate VOI yielded only the predicted interaction of VAL by MAG by PRB ($F_{(4,52)} = 5.68$; $p < 0.001$). Analysis of peak activation in the right anterior cingulate VOI also yielded the interaction of VAL by MAG by PRB ($F_{(4,52)} = 2.62$; $p < 0.05$). Pairwise comparisons indicated that bilateral anterior cingulate activation was greater for anticipation of 80% large gains than for matched nonincentives. Thus, the bilateral anterior cingulate did not show activity consistent with representation of the value or probability terms of EV, but activation in this region might relate instead to integration of these terms.

Psychologically, if the NAcc supports a more affective compu-

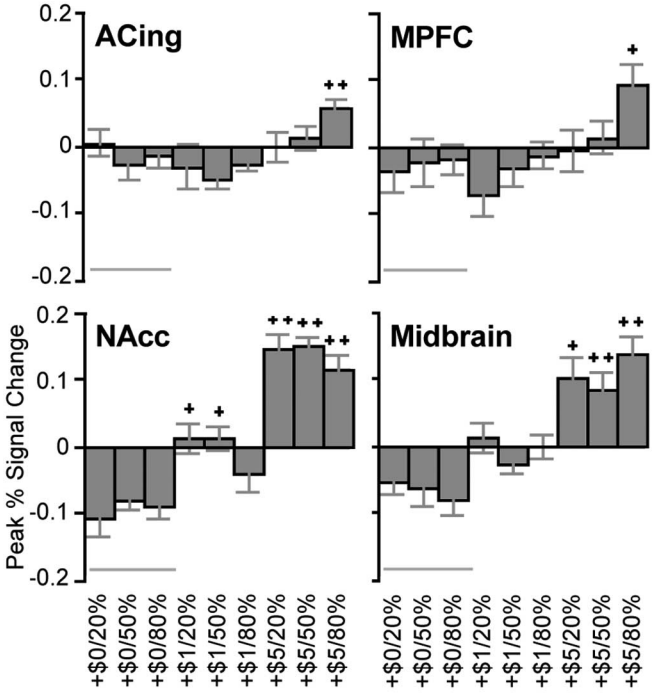

Figure 4. Peak percentage signal change by gain trial type for right VOIs. Bars represent means ± SEM ($n = 14$). Trial alignment (left to right) reflects expected value (low to high). Symbols indicate significant difference from the low (+, >$0.00) and middle (++, >$1.00 and >$0.00) magnitude incentive conditions, matched for valence and probability (within-subject pairwise comparisons; $p < 0.005$, corrected for 9 comparisons). ACing, Anterior cingulate. Error bars represent SEM.

tation related to the anticipated value of an outcome, individuals who show the strongest NAcc activation to large gain cues (which elicited significant activation) should also report experiencing the most positive aroused reactions. In contrast, if the MPFC supports a more probabilistic computation related to the perceived likelihood of an outcome, individuals who show the greatest discrepancy in MPFC activation for high- versus low-probability gain cues should instead report the largest disparity in their estimated probability of success.

Individual difference analyses of VOI data supported both hypotheses. Analysis of peak activation in the midbrain and anterior cingulate VOIs did not significantly correlate with cue-elicited affect or estimated probability. However, individuals with greater NAcc activation to +$5.00 cues also reported greater cue-elicited positive arousal (left Nacc, $r = 0.48$; right Nacc, $r = 0.53$; $p < 0.05$; one-tailed), replicating previous findings (Knutson et al., 2001a; Bjork et al., 2004), whereas NAcc activation was not significantly correlated with estimated probability. In contrast, individuals who showed a greater disparity in MPFC activation in response to high- versus low-probability +$5.00 cues also estimated a greater difference in the probability of succeeding on high- versus low-probability trials (left MPFC, 0.39, not significant; right MPFC, $r = 0.51$; $p < 0.05$; one-tailed) (Fig. 5), whereas MPFC activation was not significantly correlated with positive arousal.

Discussion

In this initial investigation of the neural correlates of EV, mesolimbic regions (including the midbrain, NAcc, and MPFC) showed activation correlated with a linear model of EV. However, additional analyses suggested that distinct regions preferentially responded to different components of EV. Specifically, whereas NAcc activity correlated only with anticipated gain magnitude, MPFC activity also correlated with anticipated gain probability. Individual analyses confirmed that, although NAcc activation was associated with positive arousal elicited by large gain cues, MPFC activation was associated with the perceived probability of obtaining large gains. Together, these findings suggest that distinct mesolimbic regions play different roles in EV computation (Shizgal, 1997).

Novel features of the present design enabled us to address alternative accounts of mesolimbic activation. First, event-related trial decomposition allowed temporally specific interrogation of anticipatory activation. Although the short length of individual trials (8 s) made it difficult to separate neural activation during cue presentation and subsequent response from activation during anticipation, identical group analyses timed for cue presentation (i.e., 2 s earlier) yielded similar but weaker results, whereas analyses timed for target response (i.e., 2 s later) failed to yield predicted results. Time course analyses additionally suggested that activation peaks coincided with the anticipatory delay and that anticipatory activation did not differ as a function of outcome. Thus, analyses targeted anticipatory activation. Whether anticipatory activation can be sustained over intervals longer than 2–3 s remains to be determined.

Second, inclusion of valence as a factor enabled exclusion of sensory stimulation, arousal, salience, attention, and motor preparation interpretations, because behavioral analysis indicated that all of these demands were equated across gain and loss conditions of equal magnitude. Although some regions (including the NAcc) showed increased activation during anticipation of both large-magnitude gains and large-magnitude losses, all mesolimbic regions showed significantly greater increases in activation during anticipation of gains versus losses, as documented in previous research using both monetary and juice rewards (Knutson et al., 2001a; O'Doherty et al., 2002). Thus, brain activation related to magnitude may reflect widespread increases in oxygen use during anticipation of large incentives but cannot account for asymmetrical findings related to incentive valence.

Third, inclusion of multiple levels of incentive probability enabled examination of whether activation in mesolimbic regions would correlate with probability (i.e., 80 vs 20%, a linear trend) or uncertainty (i.e., 50 vs 80 and 20%, a quadratic trend), as suggested by electrophysiological recordings of midbrain dopamine neurons in monkeys (Fiorillo et al., 2003). Importantly, activation in volumes of interest (specifically, the MPFC) showed linear but not quadratic trends, supporting probabilistic rather than uncertainty-based interpretations. However, the monkey research did not include an anticipatory delay between cue presentation and outcome, which may preclude direct comparisons. Also, the blood oxygen level-dependent signal indexed by fMRI may reflect a number of synaptic events, including, but not limited to, increased dopamine release (e.g., changes in local glutamate activity) (Logothetis and Wandell, 2004).

Fourth, use of an adaptive timing algorithm ensured that cued probabilities of success matched each individual's performance throughout the experiment. To maximize engagement, all conditions required a contingent response, raising the possibility that anticipatory activations resulted from motor preparation. However, covariance of a regressor modeling motor preparation did not diminish any of the observed effects. Furthermore, although premotor regions (e.g., the supplementary motor area) showed increased activation during anticipation of low-probability outcomes as might be predicted by a motor preparation account, the

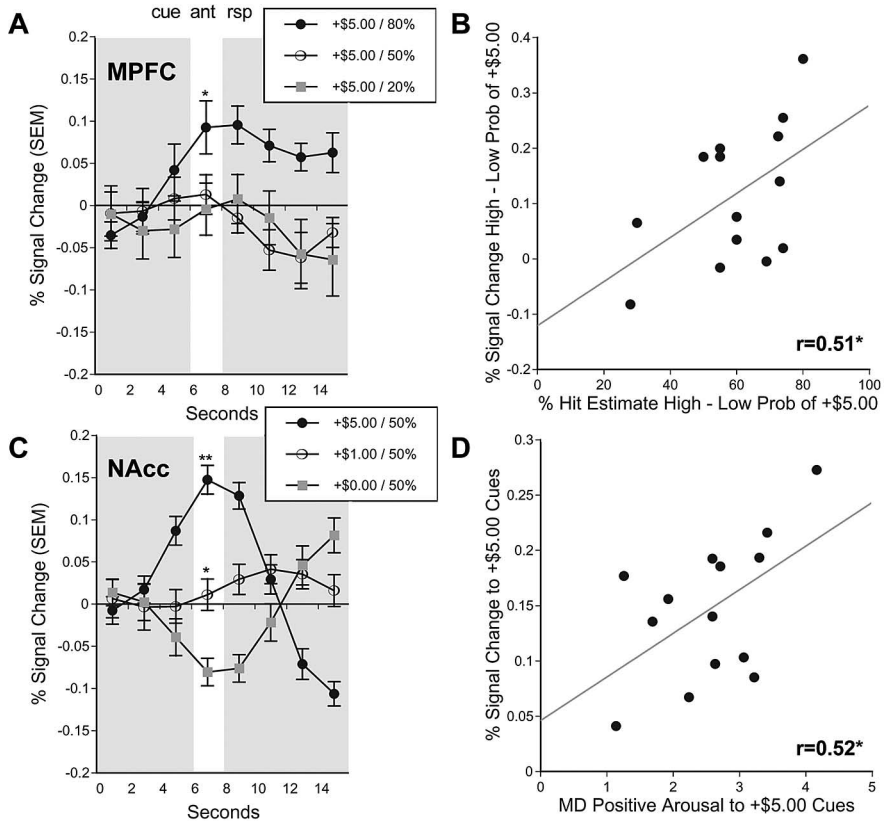

Figure 5. VOI time courses and functional correlations. ***A***, Activation time courses taken from the right MPFC VOI for +$5.00 gain trials of varying probabilities (*, 80 vs 20% probability of success; $p < 0.05$; two-tailed *t* test). ***B***, Correlation of percentage hit estimate for high- versus low-probability trials with percentage signal change in the right MPFC VOI for high- versus low-probability +$5.00 gain trials (lag, 4 s; $r = 0.51$; $p < 0.05$; one-tailed). ***C***, Activation time courses taken from the right NAcc VOI for gain trials of varying magnitudes (**, +$5.00 vs +$1.00 and +$0.00; *, $1.00 vs $0.00; $p < 0.05$; two-tailed *t* test). ***D***, Correlation of self-reported positive arousal in response to +$5.00 cues with percentage signal change in the right NAcc VOI in response to +$5.00 cues (lag, 4 s; $r = 0.52$; $p < 0.05$; one-tailed). ant, Anticipation; rsp, response; Prob, probability; MD, mean-deviated. Error bars represent SEM.

MPFC instead showed the opposite pattern of increased activation during anticipation of high-probability outcomes. Other experiments have also demonstrated ventral–striatal activation during anticipation of rewards using noncontingent tasks (O'Doherty et al., 2002; Ramnani et al., 2004; Wittmann et al., 2005).

Although group analyses suggested that NAcc and MPFC activation correlated with different components of EV, VOI analyses suggested that these associations were more robust for gains than for losses. This finding corresponds with animal research indicating that anticipated gains and losses may be processed by distinct subcortical mechanisms (Panksepp, 1998). In the realm of gains (Kahneman and Tversky, 1984), the findings raise the interesting possibility that, although the NAcc is a viable candidate for representing the value term, the MPFC is a better candidate for representing the probability term of EV. This interpretation is consistent with previous findings that the MPFC does not activate during gain anticipation when outcome probabilities are held constant (Knutson et al., 2001b, 2003). Combined with the activation seen in these regions, the anterior cingulate showed a more complex pattern, suggestive of an integrative weighting function that flattens anticipation of nonincentives and sharpens the anticipation of large incentives. Such a weighting function might be consistent with the postulated role of the anterior cingulate in attention and performance monitoring (Fig. 4) (Ridderinkhof et al., 2004).

In the economic literature, EV is often inferred from choice behavior (e.g., observed choices between gambles). Nonetheless, to avoid circularity and truly predict choice, the definition of EV should not include the behavior that it predicts. Brain-imaging techniques now offer the possibility of substituting neural signals as more proximal predictors of choice. To isolate brain activity specifically associated with anticipation of incentives, we sought to control subsequent choice in this study. Thus, these findings cannot address whether EV predicts choice behavior (or which components do so). However, the findings do identify and characterize neural targets for future research that endeavors to predict individual choice behavior (Erk et al., 2002; Paulus and Frank, 2003; McClure et al., 2004).

Overall, the findings suggest a componential model of EV computation, in which the magnitude of potential gains is represented subcortically, whereas correction for their future likelihood occurs cortically. This scheme is consistent with the proposition that the NAcc computes gain prediction, whereas the MPFC computes gain prediction errors (Knutson et al., 2003). Such a model has anatomical feasibility, because tract-tracing studies indicate an ascending spiral of connectivity running from the ventral striatum to the ventral prefrontal cortex and back to more dorsal aspects of the striatum and prefrontal cortex, before terminating in motor pathways (Haber, 2003; Voorn et al., 2004). This componential model of EV computation implies that organisms with impaired frontal function, such as members of other species and very young or old humans, might have no problem anticipating gains but might find it more difficult to adjust when outcomes betray their expectations. Indeed, data from adults with selective MPFC lesions suggest that they do not show deficits in anticipating gains but rather in framing this information in the context of other possible outcomes (Camille et al., 2004), as illustrated by the famous example of acquired impulsivity in the frontal-lesioned patient Phineas Gage (Bechara et al., 2000; Camille et al., 2004).

The findings are broadly consistent with the notion that subcortical circuits can generate affective reactions, whereas cortical circuits support more probabilistic considerations (MacLean, 1990; Panksepp, 1998). Indeed, individual difference analyses indicated that subjects who showed the greatest NAcc activation also experienced the most positive arousal during gain anticipation, whereas subjects who showed the greatest disparity in MPFC activation estimated the largest difference in the probability of succeeding on high- versus low-probability gain trials. Together, these findings imply that Pascal's historic formula may include both affective and cognitive computations and take neuroscientists a step closer to understanding how the brain integrates these considerations to guide economic and social choices.

References

Bandura A (1977) Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Bechara A, Damasio H, Damasio AR (2000) Emotion, decision making and the orbitofrontal cortex. Cereb Cortex 10:295–307.

Bjork JM, Knutson B, Fong GW, Caggiano DM, Bennett SM, Hommer DW (2004) Incentive-elicited brain activation in adolescents: similarities and differences from young adults. J Neurosci 24:1793–1802.

Breiter HC, Aharon I, Kahneman D, Dale A, Shizgal P (2001) Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron 30:619–639.

Camille N, Coricelli G, Sallet J, Pradat-Diehl P, Duhamel JR, Sirigu A (2004) The involvement of the orbitofrontal cortex inn the experience of regret. Science 304:1167–1170.

Cohen MS (1997) Parametric analysis of fMRI data using linear systems methods. NeuroImage 6:93–103.

Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance images. Comput Biomed Res 29:162–173.

Delgado MR, Nystrom LE, Fissell C, Noll DC, Fiez JA (2000) Tracking the hemodynamic response to reward and punishment in the striatum. J Neurophysiol 84:3072–3077.

Elliott R, Friston KJ, Dolan RJ (2000) Dissociable neural responses in human reward systems. J Neurosci 20:6159–6165.

Erk S, Spitzer M, Wunderlich AP, Galley L, Walter H (2002) Cultural objects modulate reward circuitry. NeuroReport 13:2499–2503.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299:1898–1902.

Gallistel CR (1986) The role of the dopaminergic projections in MFB self-stimulation. Behav Brain Res 22:97–105.

Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging data using the false discovery rate. NeuroImage 15:870–878.

Glover GH, Law CS (2001) Spiral-in/out BOLD fMRI for increased SNR and reduced susceptibility artifacts. Magn Reson Med 46:515–522.

Haber SN (2003) The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat 26:317–330.

Kahneman D, Tversky A (1984) Choices, values, and frames. Am Psychol 39:341–350.

Knutson B, Westdorp A, Kaiser E, Hommer D (2000) FMRI visualization of brain activity during a monetary incentive delay task. NeuroImage 12:20–27.

Knutson B, Adams CM, Fong GW, Hommer D (2001a) Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci 21:RC159(1–5).

Knutson B, Fong GW, Adams CM, Varner JL, Hommer D (2001b) Dissociation of reward anticipation and outcome with event-related fMRI. NeuroReport 12:3683–3687.

Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D (2003) A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related FMRI. NeuroImage 18:263–272.

Loewenstein CK, Loewenstein GF, Weber EU, Welch N (2001) Risk as feelings. Psychol Bull 2:267–286.

Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Annu Rev Neurosci 66:735–769.

MacLean PD (1990) The triune brain in evolution. New York: Plenum.

McClure SM, Li J, Tomlin D, Cypert KS, Montague LM, Montague PR (2004) Neural correlates of behavioral preferences for culturally familiar drinks. Neuron 44:379–387.

Montague PR, Berns GS (2002) Neural economics and the biological substrates of valuation. Neuron 36:265–284.

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

Panksepp J (1998) Affective neuroscience: the foundations of human and animal emotions. New York: Oxford UP.

Paulus MP, Frank LR (2003) Ventromedial prefrontal cortex activation is critical for preference judgments. NeuroReport 14:1311–1315.

Ramnani N, Elliott R, Athwal BS, Passingham RE (2004) Prediction error for free monetary reward in the human prefrontal cortex. NeuroImage 23:777–786.

Ridderinkhof KR, Ullsperger M, Crone EA, Nieuwenhuis S (2004) The role of the medial frontal cortex in cognitive control. Science 306:443–447.

Rotter JB (1972) Applications of a social learning theory of personality. New York: Holt.

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.

Shizgal P (1997) Neural basis of utility estimation. Curr Opin Neurobiol 7:198–208.

Slovic P, Finucane M, Peters E, MacGregor D (2002) The affect heuristic. In: Heuristics and biases: the psychology of intuitive judgment (Gilovich T, Griffin D, Kahneman D, eds), pp 397–420. New York: Cambridge UP.

von Neumann J, Morgenstern O (1944) Theory of games and economic behavior. Princeton, NJ: Princeton UP.

Voorn P, Vanderschuren LMJ, Groenewegen HJ, Robbins TW, Pennartz CMA (2004) Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 27:468–474.

Watson D, Wiese D, Vaidya J, Tellegen A (1999) The two general activation systems of affect: structural findings, evolutionary considerations, and psychobiological evidence. J Pers Soc Psychol 76:820–838.

Wittmann BC, Schott BH, Guderian S, Frey JU, Heinze HJ, Duzel E (2005) Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. Neuron 45:459–467.