Behavioral/Systems/Cognitive

Neural Correlates of Instrumental Contingency Learning: Differential Effects of Action–Reward Conjunction and Disjunction

Mimi Liljeholm,[1] **Elizabeth Tricomi,**[2] **John P. O'Doherty,**[1,3] **and Bernard W. Balleine**[4,5]

[1]Division of the Humanities and Social Sciences, Computation and Neural Systems Program, California Institute of Technology, Pasadena, California 91125, [2]Department of Psychology, Rutgers University, Newark, New Jersey 07102, [3]Institute of Neuroscience, Trinity College Dublin, Dublin 2, Ireland, [4]Department of Psychology, and the Brain Research Institute, University of California, Los Angeles, California 90024, and [5]Brain and Mind Research Institute, University of Sydney, New South Wales 2050, Australia

Contingency theories of goal-directed action propose that experienced disjunctions between an action and its specific consequences, as well as conjunctions between these events, contribute to encoding the action–outcome association. Although considerable behavioral research in rats and humans has provided evidence for this proposal, relatively little is known about the neural processes that contribute to the two components of the contingency calculation. Specifically, while recent findings suggest that the influence of action–outcome conjunctions on goal-directed learning is mediated by a circuit involving ventromedial prefrontal, medial orbitofrontal cortex, and dorsomedial striatum, the neural processes that mediate the influence of experienced disjunctions between these events are unknown. Here we show differential responses to probabilities of conjunctive and disjunctive reward deliveries in the ventromedial prefrontal cortex, the dorsomedial striatum, and the inferior frontal gyrus. Importantly, activity in the inferior parietal lobule and the left middle frontal gyrus varied with a formal integration of the two reward probabilities, ΔP, as did response rates and explicit judgments of the causal efficacy of the action.

Introduction

The capacity for goal-directed action depends critically on our ability to detect and represent the causal relationship between actions and their consequences. Evidence suggests that, while such judgments are biased by conjunctions, or pairings, of an action with its specific consequences, they are also highly sensitive to disjunctions; behavioral studies have found that judgments regarding the causal status of actions vary with the likelihood of the outcome occurring noncontingently (i.e., in the absence of the action and unsignaled) (Shanks and Dickinson, 1991). However, there has been little research investigating the neural bases of the influence of noncontingent outcomes on the encoding of the action–outcome relationship.

Instrumental contingency theory formalizes the integration of response-contingent and noncontingent rewards by representing the strength of the action–reward relationship as the difference between the following two conditional probabilities: the probability of gaining a target reward (r) given that a specific action (a) is performed and the probability of gaining the reward in the absence of that action ($\sim a$) [i.e., $\Delta P = P(r|a) - P(r|\sim a)$ (Hammond, 1980). Hence, according to this view, when the two probabilities are equal, the net action–reward contingency, and so the causal status of the action, is nil, regardless of the number of experienced action–reward conjunctions. The sensitivity of goal-directed actions to the instrumental contingency has now been convincingly demonstrated in both humans (Chatlosh et al., 1985; Shanks and Dickinson, 1991) and rats (Hammond, 1980; Balleine and Dickinson, 1998). Numerous studies have found evidence of a selective decrease in the performance of an action as the contingency is degraded by increasing $P(r|\sim a)$ while keeping $P(r|a)$ constant (Balleine and Dickinson, 1998), and, in humans, explicit causal judgments vary with the instrumental contingency across variations in both conditional probabilities (Chatlosh et al., 1985).

At a neural level, studies in rats suggest that the influence of contingency degradation is mediated by a circuit involving the prelimbic prefrontal cortex and the dorsomedial striatum (DMS) (Balleine and Dickinson, 1998; Corbit and Balleine, 2003; Yin et al., 2005). Consistent with these results, in a human imaging study, Tanaka et al. (2008) found that activity in the medial prefrontal cortex (mPFC), the medial orbitofrontal cortex (mOFC), and the DMS increased with an increase in the contingency between pressing a button and receiving monetary reward. Activity in mPFC also scaled linearly with explicit causal judgments, which, in turn, were significantly correlated with the instrumental contingency. However, Tanaka et al. (2008) only assessed changes in P($r|a$) and did not manipulate P($r|\sim a$), which re-

Table 1. Programmed conditional probabilities and the resulting ΔP for each of the six blocks making up a session with $P(r \sim a)$ specified per second

Blocks	1	2	3	4	5	6
$P(r\,a)$	0.00	0.06	0.12	0.18	0.18	0.18
$P(r \sim a)$	0.00	0.00	0.00	0.06	0.12	0.18
DP	0.00	0.06	0.12	0.12	0.06	0.00
Mean $P(r\,a)$	0 (0)	0.06 (0.05)	0.12 (0.04)	0.17 (0.08)	0.18 (0.10)	0.18 (0.15)
Mean $P(r \sim a)$	0 (0)	0 (0)	0 (0)	0.05 (0.06)	0.12 (0.10)	0.15 (0.13)
Mean CJ	9.1 (14.0)	24.6 (23.7)	35.7 (25.3)	38.8 (25.7)	32.1 (30.7)	30.1 (27.1)

Rows 1–3 show the values of the programmed conditional probabilities (i.e., probability of reward, r, given the presence vs absence of the action) and the resulting ΔP, for each of the six blocks making up a session with $P(r \sim a)$ specified per second. Rows 4 and 5 show the mean experienced conditional probabilities respectively, while row 6 shows mean causal judgment (CJ) (SD).
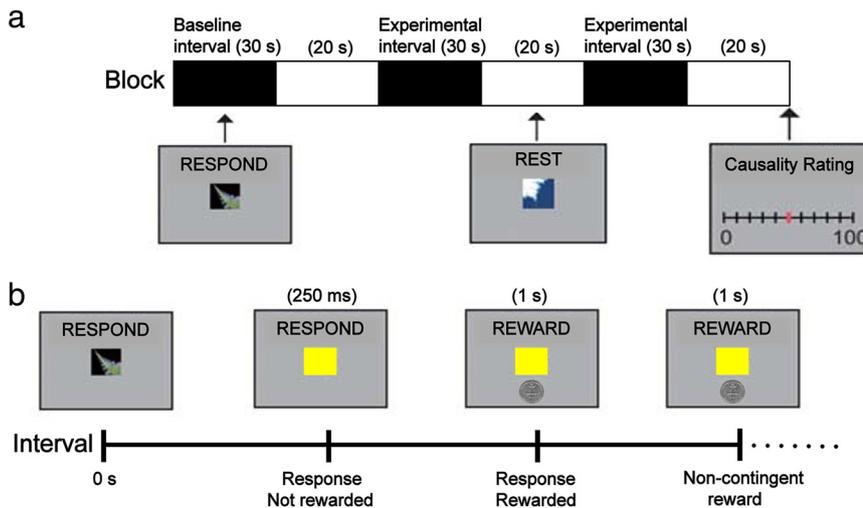

Figure 1. Illustration of the task structure. **a**, A block within a session, corresponding to a column in Table 1. Each block consisted of three 30 s respond intervals, one baseline interval, and two experimental intervals corresponding to one of the conditions listed in Table 1 (see text). The response intervals were interleaved with three 20 s rest intervals. Causal ratings were collected at the end of the block. **b**, Example of events within a respond interval. Each event triggered the display of a yellow square, which remained on the screen for 250 ms for nonrewarded key presses and for 1 s for reward deliveries. Response-contingent rewards were delivered immediately upon pressing the key. In addition, a running total of the amount of money earned within a block was continuously displayed in the upper right corner of the screen (data not shown in figure).

mained constant (at zero) across conditions. Thus, it is unknown how activity in these areas relates to the representation of noncontingent reward probabilities and their integration with response-contingent ones.

The goal of the current study was, therefore, to assess the neural basis of contingency detection in humans sampling across a broad contingency space in which we systematically varied both conditional probabilities, $P(r|a)$ and $P(r|\sim a)$, across blocks of training. Together with changes in neural activation, we assessed the effects of these manipulations behaviorally, both on changes in performance and in explicit causal judgment.

Materials and Methods

Subjects. Nineteen healthy right-handed volunteers (25 ± 4 years old, 8 females) participated in the study. The volunteers were preassessed to exclude those with a previous history of neurological or psychiatric illness. All subjects gave informed consent, and the study was approved by the Institutional Review Board of the California Institute of Technology.

Experimental procedures. At the beginning of the experiment, subjects were informed that they would be given the opportunity to earn a 25 cent reward by pressing a key but that, while they were free to press the key as often as they liked, each press would cost them 1 cent. They were further instructed that the relationship between pressing the key and receiving the 25 cent reward would vary across blocks and that, in some blocks, the 25 cent reward might be delivered in the absence of a response. To ensure

some familiarity with the task, subjects completed a single 80 s training block, in which $P(r|a)$ equaled 0.18 and $P(r|\sim a)$ equaled 0 (i.e., $\Delta p = 0.18$), before they entered the scanner.

Once in the scanner, each subject was presented with six different contingency conditions; the values of the two conditional probabilities for each of these conditions are listed in Table 1, together with the resulting ΔP (rows 1–3). Due to methodological constraints imposed by the functional magnetic resonance imaging (fMRI) method, each block of training was divided into three 30 s "respond" intervals, during which responding was unconstrained, interleaved with three 20 s "rest" intervals, during which subjects were instructed to not respond (Fig. 1). To ensure sufficient sampling of the response in each condition, and to avoid any carryover of response suppression from low-contingency blocks, the first respond interval in each block, henceforth the baseline interval, always had the same, relatively high, contingency as that used during prescanning practice (i.e., $\Delta p = 0.18$). The subsequent two respond intervals within a block, henceforth the experimental intervals, shared one of the condition-specific contingencies shown in Table 1. For each subject, the entire set of six contingency blocks was presented in each of three sessions, separated by a 5 min break during which the subject remained in the scanner, with the order of the blocks randomized within sessions. Our task included a cost for responding to encourage subjects to regulate their performance according to the instrumental contingency.

Each time the participant pressed the key or received a reward, a yellow fractal appeared. For nonrewarded responses, the duration of this fractal was 250 ms, whereas, whenever a reward was delivered, the yellow fractal remained on the screen for 1 s, together with a depiction of a quarter and the text "Reward, You Win!" (Fig. 1). While the yellow fractal, and associated information, was displayed, no other events were generated; consequently, these brief periods imposed a constraint on how closely a noncontingent reward could occur to a response, in addition to the time bin used to implement $P(r|\sim a)$. Finally, throughout each response interval, a running total of cents accumulated within the block was displayed in the top right corner of the screen.

The default time bin for generating noncontingent rewards was 1000 ms; however, this time bin was modified for each subject based on the average rate of responding. Specifically, for each subject and in each session, the time bin in each block, except the first, was equal to the average inter-response interval in the previous block, as long as this bin was not <500 or >2000 ms, in which case the default of 1000 ms was used. Although in some cases this generated experienced reward probabilities that deviated substantially from programmed ones, it ensured that even subjects who responded at very high rates received noncontingent rewards. In addition to our primary behavioral measure of response rate, judgments of the causal relationship between pressing the key and receiving the 25 cent reward were collected at the end of each block, on a scale ranging from 0 (pressing the key never caused the reward to occur) to 100 (pressing the key always caused the reward to occur).

Imaging procedures. A 3 tesla scanner (MAGNETOM Trio; Siemens) was used to acquire structural T1-weighted images and T2*-weighted echo-planar 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) blood-oxygen-level-dependent (BOLD) contrast. To recover signal loss from dropout in the mOFC (O'Doherty et al., 2002), each horizontal section was acquired at 30° to the anterior commissure–posterior commissure axis.

Behavioral data analysis. The conditional probabilities acted as parameters for a software probability generator and, consequently, the actual

values of $P(r|a)$ and $P(r|{\sim}a)$ sometimes deviated from those listed in Table 1. To equate temporal assumptions across subjects and sessions, we specified a sampling period of 1 s, coding each of the relevant events (i.e., responses, response-contingent rewards, and noncontingent rewards) as either present or absent in each such period. We then computed ΔP based on the resulting event frequencies, collapsing across the two experimental intervals making up each block. Response rates associated with these objective contingency values were computed as presses per second, correcting for the time consumed by reward-deliveries (i.e., 1 s per reward). Finally, the six blocks were ranked according to ΔP values, and response rates and causal ratings for the ranked blocks were entered into a contingency × session repeated-measures ANOVA. For a more fine-grained analysis, we also computed contingencies and response rates for each 10 s period, across blocks and sessions, and assessed the correlation between these variables for each subject.

Imaging data analysis. Image processing and statistical analyses were performed using SPM5 (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 slice acquisition, realigned to the first volume, spatially normalized to the Montreal Neurological Institute (MNI) echo-planar imaging template, and spatially smoothed with a Gaussian kernel (8 mm, full width at half-maximum). We used a high-pass filter with cutoff at 128 s. Our imaging analysis focused on the following two basic questions: (1) is there a neural signal that maps onto the instrumental contingency across variations in both conjunctive and disjunctive reward probabilities?; and (2) are the two components of the contingency correlated with distinct patterns of neural activity? Following Tanaka et al. (2008) to assess changes in neural activity over time as a function of local fluctuations in the relevant variables, we constructed two sets of subject-specific fMRI design matrices, each with an onset regressor modeling a BOLD response over 10 s periods. In one model, the two conditional probabilities, computed for each period, were entered as parametric modulators, and in the other ΔP was entered as a single modulator. In both models, response rates and reward deliveries associated with each 10 s period were convolved with a canonical hemodynamic response function and entered, without orthogonalization, as regressors of no interest together with six additional regressors accounting for the residual effects of head motion. All regressors of interest 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.

Small volume corrections (SVCs) were performed on three a priori regions of interest using a 10 mm sphere; the center coordinates were obtained by averaging across several studies (Table 2). All of these areas have been identified in highly relevant studies assessing goal-directed instrumental action-selection, as follows: (1) medial orbitofrontal cortex $(x, y, z = 3, 33, -19)$; (2) medial prefrontal cortex $(x, y, z = 7, 51, -5)$ (O'Doherty et al., 2003; Hampton et al., 2006; Valentin et al., 2007; Tanaka et al., 2008; Gläscher et al., 2009); and (3) right $(x, y, z = 13, 11, 11)$ and left $(x, y, z = -9, 7, 4)$ caudate nucleus (CN) (O'Doherty et al., 2003; Tricomi et al., 2004; Tanaka et al., 2008). All other areas were reported at $p < 0.05$, using cluster size thresholding (CST) to adjust for multiple comparisons (Forman et al., 1995). Monte Carlo simulations, implemented by the Analysis of Functional NeuroImaging AlphaSim program, were 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 separate effects due to the processing of conditional probabilities from those reflecting encoding of ΔP, we conducted an exclusion analysis, masking the contrasts of conditional probabilities with the ΔP contrast; specifically, with a positive contrast for $P(r|a)$ and a negative contrast for $P(r|{\sim}a)$. Because this analysis involves accepting the null hypothesis that neural activity does not correlate with ΔP, we used a very liberal threshold of 0.1 for these masking contrasts.

To eliminate nonindependence bias for plots of parameter estimates, a leave-one-subject-out (LOSO) (Esterman et al., 2010) approach was used in which 19 general linear models (GLMs) were run with one subject

Table 2. Coordinates for ROIs

ROI	*x, y, z*	References
Ventromedial PFC		
Medial OFC	−2, 30, −20	Tanaka et al. (2008)
	−6, 24, −21	Glaescher et al. (2009)
	9, 27, −12	
	−3, 36, −24	Valentin et al (2007)
	12, 36, −18	O'Doherty et al. (2003)
	9, 42, −12	
	0, 33, −24	Hampton et al (2006)
Medial PFC	−6, 52, −10	Tanaka et al. (2008)
	3, 54, −3	Glaescher et al. (2009)
	6, 30, −9	
	24, 45, −6	Valentin et al (2007)
	9, 66, 6	O'Doherty et al. (2003)
	6, 57, −6	Hampton et al (2006)
Caudate nucleus		
Right CN	6, 10, 20	Tanaka et al. (2008)
	9, 16, 4	Tricomi et al. (2004)
	21, 0, 18	O'Doherty et al. (2003)
	15, 18, 3	
Left CN	−12, 11, 8	Tricomi et al. (2004)
	−6, 3, 0	O'Doherty et al. (2003)

Table 3. Coordinates and significance levels for contrasts

| Region | MNI coordinates | | | |
	X	y	z	*p* value
Negative correlation with ΔP				
Left SPL	−42	−51	57	*
Left IPL	−51	−54	18	*
MFG	−27	0	54	*
Positive correlation with $P(r{\sim}a)$				
Left MFG	−48	21	36	*
Right MFG	45	24	33	*
Medial FG	6	24	54	*
Right pSTG	54	−39	15	*
Right IPL	54	−60	48	*
Left pCN	−9	0	15	*
Red Nucleus	3	−24	−3	*
Lingual Gyrus	−6	−84	−6	*
Right IFG	27	24	−12	*
Positive correlation with $P(r\,a)$				
Right mPFC	12	57	−6	**
Right aCN	15	9	15	**

pSTG, Posterior superior temporal gyrus. *$p < 0.05$, CST; **$p < 0.05$, SVC.

left out in each and with each GLM defining the voxel cluster for the subject that was left out. The relevant local maxima were then used to extract β weights for a range of modulator values for each subject and session, and these were averaged to plot overall effect sizes. Extreme modulator values for which >70% of data points were missing (≤ -0.5 and >0.5 for the contingency modulator, and >0.5 for the two conditional probabilities) were excluded from the plots.

Results
Behavioral results
The mean causal ratings and mean objective conditional probabilities (based on actually experienced event frequencies) are shown for each programmed condition in Table 1. Note that, for conditions across which $P(r|{\sim}a)$ varies while $P(r|a)$ remains constant and high, mean ratings also remain high and relatively unchanged, suggesting a bias toward $P(r|a)$. However, these mean causal ratings more likely reflect individual differences in the objective variable values. Specifically, objective values of $P(r|a)$ and $P(r|{\sim}a)$ in individual blocks differed substantially from the mean objective values computed across subjects and sessions, as
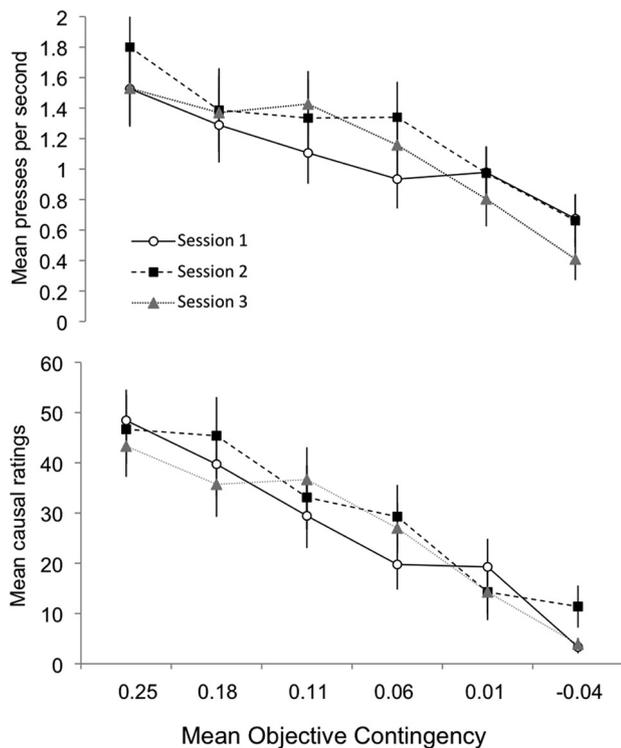
Figure 2. Mean presses per second (top) and mean causal ratings (bottom) across blocks sorted in descending order by objective contingency. Blocks were sorted separately for each subject and session before the mean contingency was computed. Thus, while all subjects experienced one block in each session for which both objective conditional probabilities (and consequently contingency) were zero, the position of this block in the descending order of objective contingency values varied across subjects and sessions, resulting in a nonzero mean for each block listed on the x-axis. Error bar = SEM.

well as from the programmed values. Furthermore, while large deviations in the objective values sometimes resulted in negative contingencies, the rating scale did not allow participants to indicate a preventive causal relationship, thus biasing judgments in a positive direction. Consistent with this interpretation, when mean causal ratings were computed solely based on blocks in which both objective conditional probabilities were close to the programmed values (within a 0.05 deviation), mean ratings equaled 46.2, 32.5, and 18.8, respectively, for the conditions listed in columns 4 to 6 of Table 1, suggesting a much stronger influence of $P(r|\sim a)$. All of our statistical analyses, and all subsequently reported descriptive statistics, are based on objective, rather than programmed, variable values.

As can be seen in the top panel of Figure 2, mean response rates clearly decrease with a decrease in mean objective contingency ($F_{(1,18)} = 38.8$, $p < 0.001$). Simple contrasts revealed that all differences between adjacent contingency levels were significant ($p < 0.05$), except for that between the second and third levels ($p = 0.07$). The mean response rate and mean objective contingency for the baseline intervals were both high (1.4 and 0.22, respectively) relative to those in the majority of experimental conditions. Comparable results were observed with respect to the explicit causal judgments (Fig. 2, bottom), which also decreased linearly as a function of objective contingency ($F_{(1,18)} = 47.5$, $p < 0.001$), with all differences between adjacent contingency levels reaching significance ($p < 0.05$), except for that between the first and second levels ($p = 0.20$). There was no significant effect of session, nor any significant interactions, for either response rates or causal judgments (all F values <1.0). The

mean difference, across subjects and sessions, between the two experimental intervals within each block was 0.32 with an SD of 0.36. Notably, such variations in response rate were likely due to the fact that the experienced contingency also varied across these two intervals; indeed, differences in response rates across the intervals within a block were highly correlated, across subjects and sessions, with concomitant differences in contingency ($p < 0.001$). Finally, the correlations between ΔP and response rates, computed across 10 s bins for each subject, were highly significant for the vast majority of subjects ($p < 0.001$ for 16 of 19 subjects; $p < 0.05$ for 2 of the remaining 3 subjects). In summary, our results replicate those of Shanks and Dickinson (1991) and Chatlosh et al. (1985); both response rates and explicit causal judgments showed a systematic decline with a decrease in objective contingency.

Imaging results: ΔP

Coordinates and significance levels for all contrasts assessing parametric modulation are shown in Table 3. We first tested for areas showing changes in activity related to the instrumental contingency. Our results indicated that three distinct areas tracked contingency values: the left middle frontal gyrus and the left superior and inferior parietal lobules. Interestingly, we found that activity in these areas correlated negatively with this parametric modulator (CST, all p values < 0.05) (Fig. 3a,b), such that their activity was greatest when subjects experienced low (including negative) contingencies and was weakest when subjects experienced high contingences. Bilateral activity was seen in all three areas at an uncorrected threshold of 0.005. No voxels survived our statistical threshold for the reverse contrast, testing for areas in which activity correlated positively with contingency.

Imaging results: conditional probabilities

We next tested for areas showing changes in activity related to $P(r|\sim a)$ and found significant responses throughout the lateral frontal cortex bilaterally, the medial frontal cortex, the right posterior superior temporal gyrus, the right posterior intraparietal sulcus, and the left posterior caudate (Levitt et al., 2002) (CST, all p values < 0.05). Activity emerged bilaterally in all of these areas, except for that in the posterior caudate nucleus, at an uncorrected threshold of 0.005. Activity also emerged in the medial prefrontal cortex [$(x, y, z = 12, 42, 3)$], although this did not quite reach significance ($p < 0.001$, uncorrected). Moreover, only the effects found in the right inferior frontal cortex (IFG) and the left posterior CN (pCN) survived masking with an exclusive, negative contrast for ΔP, thresholded at 0.1, suggesting that activity in these areas is specific to a representation of $P(r|\sim a)$ (Fig. 4a). To rule out overall reward rate as the source of observed neural activity, we also conducted an additional analysis in which we included this variable in the design matrix and found that the significant effects still emerged at the corrected threshold of 0.05. The test for areas showing changes in activity related to $P(r|a)$ revealed significant effects in the right mPFC (SVC, $p < 0.05$) and the right anterior CN (aCN) (SVC, $p < 0.05$). An uncorrected threshold of 0.005 did not render these effects bilateral, and only the effects found in the caudate survived masking with an exclusive positive contrast for ΔP, again thresholded at 0.1 (Fig. 4b).

Imaging results: additional region-of-interest analyses

To further assess the effects found for ΔP in the inferior parietal lobule, we anatomically defined this region, in each hemisphere, using WFU PickAtlas (Maldjian et al., 2003). We then used Mars-Bar (Brett et al., 2002) to perform region-of-interest (ROI) anal-

yses of a (low-high) contrast and found significant effects in both the left and right inferior parietal lobules (IPLs) (both p values < 0.05), confirming our previous finding that activity in this area decreases as the contingency increases. We also performed ROI analyses contrasting high with low values for the two conditional probabilities within our caudate ROIs, as well as within an ROI defined across the ventromedial prefrontal cortex (vmPFC) (consisting of the mOFC and adjacent mPFC regions defined in Table 2). For $P(r|\sim a)$, these analyses yielded significant effects in the left caudate nucleus ($p < 0.05$). For $P(r|a)$, we found significant effects in both the left and right caudate, as well as in the vmPFC (all p values <0.05).

Discussion

When trying to determine how effective an action is in producing some reward, it is important to consider the following two conditional probabilities: the probability that the action is followed by that reward, $P(r|a)$; and the probability that the reward occurs in the absence of that action, $P(r|\sim a)$. The behavioral influence of both response-contingent and noncontingent rewards on free operant responding has been convincingly demonstrated in both humans and rats, and has been central to claims about the role of causal knowledge in the performance of goal-directed actions (Balleine and Dickinson, 1998). However, while there has been extensive research on the neural processes underlying the influence of response-contingent rewards on action selection, what we know about the neural bases of processing noncontingent rewards and their integration with response-contingent ones is limited to the results of a relatively small body of studies in rodents (Yin et al., 2005). The current study used fMRI to investigate the neural substrates of action–outcome contingency learning in humans, with a focus on identifying areas responsible for integrating information about response-contingent and noncontingent reward probabilities. Consistent with previous results (e.g., Tanaka et al., 2008), we found that neural activity in the vmPFC and the right aCN encoded the probability with which an action would be followed by reward. In contrast, information about the probability of noncontingent reward was processed by two separate neural circuits: activity in the IFG and the left pCN was found to vary with the probability of receiving reward in the absence of any action. Finally, activity in the inferior and superior parietal lobules (I/SPLs) and the middle frontal gyrus (MFG) varied with instrumental contingency, a formal integration of the two reward probabilities, as did response rates and subjective causal judgments.

Our finding that neural activity in the right aCN and mPFC varied with the probability of response-contingent reward deliv-
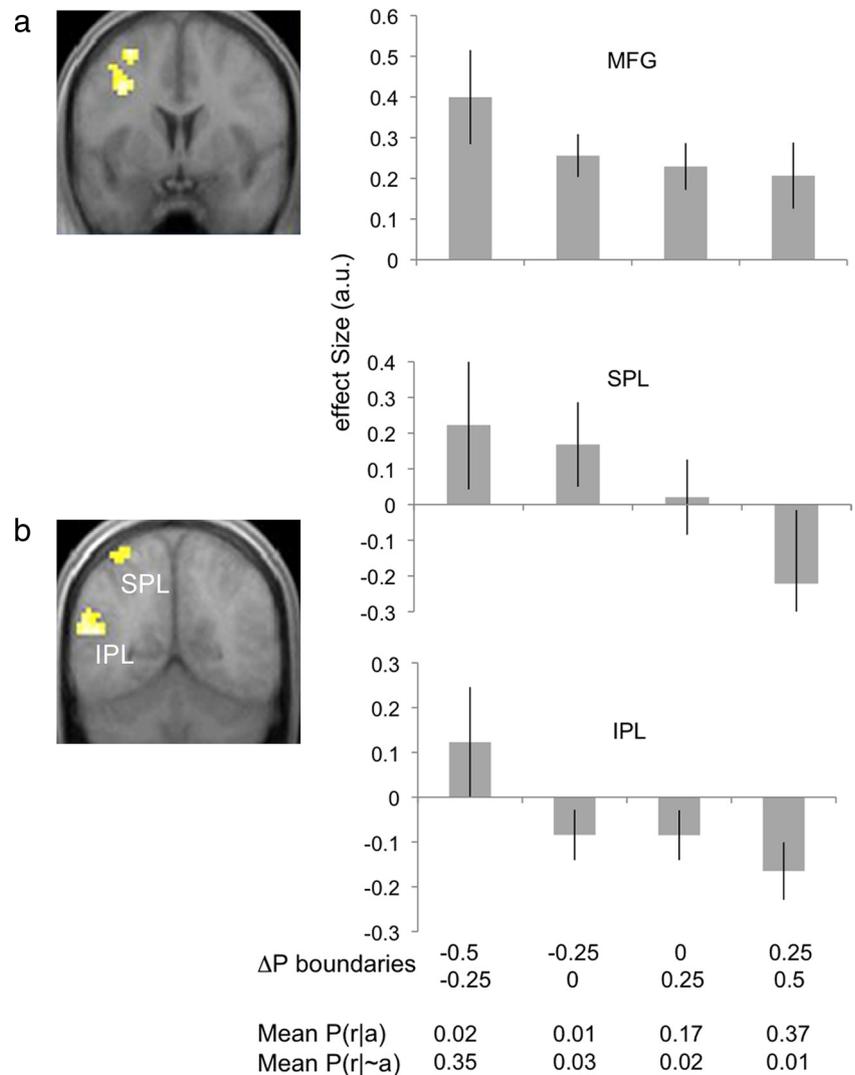

Figure 3. Activation related to instrumental contingency. *a, b*, Voxels showing significant negative correlation with the instrumental contingency were found in the left MFG ($x, y, z = -30, 3, 57; p < 0.05$, corrected) (*a*), and the SPL ($x, y, z = -42, -51, 57; p < 0.05$, corrected) and IPL ($x, y, z = -51, -57, 18; p < 0.05$, corrected) (*b*). Bar plots show mean βs (y-axes) estimated at each LOSO peak voxel and averaged across subjects and sessions. The first two rows on the bottom x-axis list the boundaries of the range of contingency values for each bin, with the top row indicating the lower boundary and the bottom row indicating the upper, inclusive boundary. The third and fourth rows list the mean experienced values for the two conditional probabilities respectively in each bin. Error bar = SEM.

ery, $P(r|a)$, is consistent with that of Tanaka et al. (2008) and supports the suggestion that these structures are functionally homologous to the rodent DMS and prelimbic cortex, respectively (Balleine and O'Doherty, 2009). Notably, the mPFC effects did not survive masking by the contingency contrast, suggesting that, rather than just encoding $P(r|a)$, this area might contribute to contingency computations; indeed, activity in mPFC was also found for $P(r|\sim a)$, albeit below our threshold for statistical significance. Additional evidence for a role of mPFC in reward integration comes from studies showing that activity this area correlates with the average value of distinct stimuli (Wunderlich et al., 2010), with the subjective valuation of delayed rewards (Kable and Glimcher, 2007), and with the relative decision value between monetary and social rewards (Smith et al., 2010).

We also found that activity in the left pCN, but not the aCN, varied with the probability of noncontingent rewards, indicating that distinct striatal areas may support estimation of the respec-
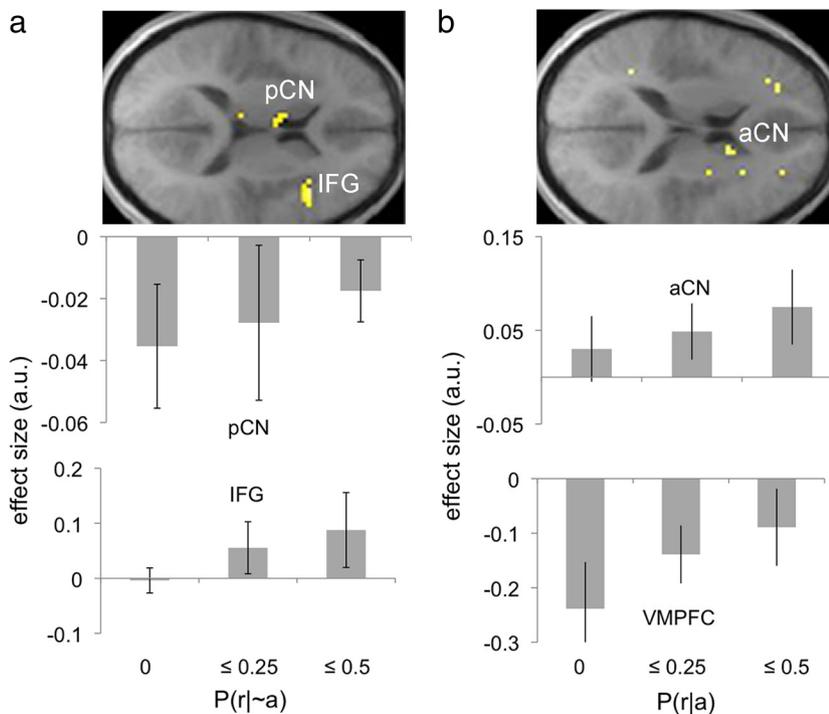
Figure 4. Activations related to the two conditional probabilities. ***a***, Results from a contrast testing for modulation by $P(r|\sim a)$ with exclusive masking by the negative contrast for ΔP. Effects were found in the left posterior caudate ($x, y, z = -9, 0, 12; p < 0.05$, CST) and the IFG ($x, y, z = 42, 18, 12$); $p < 0.05$, corrected). ***b***, Results from a contrast testing for modulation by $P(r|a)$, with exclusive masking by the positive contrast for ΔP. Significant activation was found in the right anterior caudate ($x, y, z = 15, 9, 15$; $p < 0.05$) SVC with a 10 mm sphere centered at 6, 10, and 21 (Tanaka et al., 2008). Significant activation was also found in the vmPFC ($x, y, z = 12, 57, -6; p < 0.05$) SVC with a 10 mm sphere centered at 6, 57, and −6 (Hampton et al., 2006)], but this did not survive masking. Bar plots show mean βs (y-axes) estimated at each LOSO peak voxel and averaged across subjects and sessions for the range of modulator values listed on the bottom x-axes. Error bar = SEM.

tive reward probabilities. Importantly, a similar dissociation between anterior and posterior dorsomedial striatum has previously been demonstrated in rodents; Yin et al. (2005) found that inactivation of the posterior DMS abolished sensitivity to contingency degradation (i.e., to the delivery of noncontingent rewards) while inactivation of the anterior DMS had no effect. Likewise, Corbit and Janak (2010) found that the posterior DMS was critical for the acquisition of both response–outcome and stimulus–outcome relationships, while the anterior DMS appeared to be needed only for response–outcome encoding. The current results suggest that a comparable heterogeneity might exist along the anterior–posterior axis of the human caudate nucleus (i.e., across the head and body of the caudate), providing converging evidence for the proposal that brain systems responsible for the modulation of goal-directed actions based on variations in instrumental contingency are highly conserved across species (Balleine and O'Doherty, 2010). Consistent with previous work (Shanks and Dickinson, 1991), our behavioral results, depicted in Figure 2, show a clear decrease in response rates with a decrease in the difference between the probabilities of response-contingent and noncontingent rewards (i.e., with instrumental contingency). To explore the neural basis of this response modulation, we tested for regions correlating with the local contingency computed over 10 s intervals. We found that activity increased with a decrease in contingency in the I/SPL and MFG. Although several recent neuroimaging studies have implicated the I/SPL and MFG in the representation of action–reward contingencies (Delgado et al., 2005; Schlund and Cataldo, 2005; Koch et al., 2008; Schlund and Ortu, 2010), they have primarily

explored the role of these areas in reward predictability. For example, Delgado et al. (2005) and Koch et al. (2008) found that activity in the IPL increased as the probability of being correct, and thus of reward, given one of two alternative actions decreased from 1.0 to 0.5 across stimulus conditions. They interpreted these results as reflecting a recruitment of areas responsible for controlled cognitive processes due to the decrease in reward predictability. Without additional assumptions, it is difficult to apply the concept of reward predictability to the distinction between response-contingent and noncontingent rewards that is the focus of the current study. Nonetheless, we note that this account fails to explain our finding that activity in the I/SPL appears to decrease as $P(r|a)$ increased toward 0.5, while $P(r|\sim a)$ remained relatively unchanged. In other words, that activity in these areas decreased with a decrease in the predictability of response-contingent reward (Fig. 3*b*, last two bars and bottom rows of the effect-size plots).

In the current task, the decision to respond or not is based on the relative probability of reward in the presence and the absence of the action. Note that this is akin to choosing between two alternative actions based on their respective reward probabilities; in both cases, an integration of distinct sources of information is required in order for optimal response strategies to develop. In a recent neurophysiology study, Seo et al. (2009) updated the individual values of two alternative actions using reinforcement learning theory (Sutton and Barto, 1998) and modeled the difference between action value functions as a decision variable in a free-choice task. Recording the activity of neurons in a subregion of the IPL (the lateral intraparietal cortex) in rhesus monkeys, they found that a substantial percentage of these neurons changed their activity according to the difference between the two action value functions. Interestingly, using a similar task but recording from the monkey striatum, Samejima et al. (2005) found that a greater number of striatal neurons were selective to reward expectancies associated with one but not the other action than were tuned to the difference between action values. To our knowledge, the current results provide the first simultaneous regions respectively. Further studies are needed to clarify the exact role of the implicated parietostriatal circuit in goal-directed action selection, and how it relates to the frontostriatal network that has been the focus of rodent lesion studies of instrumental contingency learning.

Although contingency computations are considered central to goal-directed learning, a couple of additional factors known to strongly influence instrumental performance likely contributed to the current findings. For example, it is possible that the immediacy of response-contingent reward delivery (i.e., the strong action–reward contiguity) used here played a role in the estimation of $P(r|a)$ (Shanks and Dickinson, 1991), and consequently in the observed correlation between this variable and activity in the anterior caudate. It is also important to note that instrumental

contingency is closely related to the utility of performing an action; recall that, in the current study, whereas noncontingent rewards were free, response-contingent rewards (of equal magnitude) were associated with a small monetary cost and, presumably, also with an effort-based cost. The IPL has been previously implicated in the integration of reward and risk (Ernst et al., 2004) and in response selection based on reward maximization (Bush et al., 2002); it is possible, therefore, that the currently reported effects in this area reflect the incorporation of noncontingent rewards into a cost–benefit analysis, rather than the computation of contingency per se. Further research is needed to determine how neural correlates of instrumental contingency learning relate to the encoding of action–reward contiguity and to the estimation of utility.

References

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

Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: cortico-striatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35:48–69.

Brett M, Anton J, Valabregue R, Poline J (2002) Region of interest analysis using an SPM toolbox. Paper presented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, June 2–6, 2002. [Available on CD-ROM in Neuroimage 16(2).]

Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, Rosen BR (2002) Dorsal anterior cingulate cortex: a role in reward-based decision making. Proc Natl Acad Sci U S A 99:523–528.

Chatlosh DL, Neunaber DJ, Wasserman EA (1985) Response-outcome contingency: behavioral and judgmental effects of appetitive and aversive outcomes with college students. Learn Motiv 16:1–34.

Corbit LH, Balleine BW (2003) The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 146:145–157.

Corbit LH, Janak PH (2010) Posterior dorsomedial striatum is critical for both instrumental and Pavlovian reward learning. Eur J Neurosci 31:1312–1321.

Delgado MR, Miller MM, Inati S, Phelps EA (2005) An fMRI study of reward-related probability learning. Neuroimage 24:862–873.

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.

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.

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.

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.

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

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.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10:1625–1633.

Koch K, Schachtzabel C, Wagner G, Reichenbach JR, Sauer H, Schlösser R (2008) The neural correlates of reward-related trial-and-error learning: an fMRI study with a probabilistic learning task. Learn Mem 15:728–732.

Levitt JJ, McCarley RW, Dickey CC, Voglmaier MM, Niznikiewicz MA, Seidman LJ, Hirayasu Y, Ciszewski AA, Kikinis R, Jolesz FA, Shenton ME (2002) MRI study of caudate nucleus volume and its cognitive correlates in neuroleptic-naive patients with schizotypal personality disorder. Am J Psychiatry 159:1190–1197.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH (2003) An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19:1233–1239.

O'Doherty J, Critchley H, Deichmann R, Dolan RJ (2003) Dissociating value of outcome from behavioral control in human orbital and ventral prefrontal cortices. J Neurosci 23:7931–7939.

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

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

Schlund MW, Cataldo MF (2005) Integrating functional neuroimaging and human operant research: brain activation correlated with presentation of discriminative stimuli. J Exp Anal Behav 84:505–519.

Schlund MW, Ortu D (2010) Experience-dependent changes in human brain activation during contingency learning. Neuroscience 165:151–158.

Seo H, Barraclough DJ, Lee D (2009) Lateral intraparietal cortex and reinforcement learning during a mixed-strategy game. J Neurosci 29:7278–7289.

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

Smith DV, Hayden BY, Truong TK, Song AW, Platt ML, Huettel SA (2010) Distinct value signals in anterior and posterior ventromedial prefrontal cortex. J Neurosci 30:2490–2495.

Sutton RS, Barto AG (1998) Reinforcement learning. Cambridge, MA: MIT.

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

Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. Neuron 41:281–292.

Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019–4026.

Wunderlich K, Rangel A, O'Doherty JP (2010) Economic choices can be made using only stimulus values. Proc Natl Acad Sci U S A 107:15005–15010.

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 22:513–523.