Overlapping Prediction Errors in Dorsal Striatum During Instrumental Learning With Juice and Money Reward in the Human Brain

Vivian V. Valentin and John P. O'Doherty

J Neurophysiol 102:3384-3391, 2009. First published 30 September 2009; doi:10.1152/jn.91195.2008

Journal of Neurophysiology publishes original articles on the function of the nervous system. It is published 12 times a year (monthly) by the American Physiological Society, 9650 Rockville Pike, Bethesda MD 20814-3991. Copyright © 2009 by the American Physiological Society. ISSN: 0022-3077, ESSN: 1522-1598. Visit our website at http://www.the-aps.org/.

J Neurophysiol 102: 3384–3391, 2009.
First published September 30, 2009; doi:10.1152/jn.91195.2008.

Overlapping Prediction Errors in Dorsal Striatum During Instrumental Learning With Juice and Money Reward in the Human Brain

Vivian V. Valentin[1] **and John P. O'Doherty**[1,2]*

[1]*Division of Humanities and Social Sciences, and Computation and Neural Systems Program, California Institute of Technology, Pasadena, California; and* [2]*Trinity College Institute of Neuroscience, and School of Psychology, Trinity College, University of Dublin, Dublin, Ireland*

Submitted 10 November 2008; accepted in final form 28 September 2009

Valentin VV, O'Doherty JP. Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *J Neurophysiol* 102: 3384–3391, 2009. First published September 30, 2009; doi:10.1152/jn.91195.2008. Prediction error signals have been reported in human imaging studies in target areas of dopamine neurons such as ventral and dorsal striatum during learning with many different types of reinforcers. However, a key question that has yet to be addressed is whether prediction error signals recruit distinct or overlapping regions of striatum and elsewhere during learning with different types of reward. To address this, we scanned 17 healthy subjects with functional magnetic resonance imaging while they chose actions to obtain either a pleasant juice reward (1 ml apple juice), or a monetary gain (5 cents) and applied a computational reinforcement learning model to subjects' behavioral and imaging data. Evidence for an overlapping prediction error signal during learning with juice and money rewards was found in a region of dorsal striatum (caudate nucleus), while prediction error signals in a subregion of ventral striatum were significantly stronger during learning with money but not juice reward. These results provide evidence for partially overlapping reward prediction signals for different types of appetitive reinforcers within the striatum, a finding with important implications for understanding the nature of associative encoding in the striatum as a function of reinforcer type.

INTRODUCTION

Accumulating evidence from neurophysiological recording studies in nonhuman primates and rodents implicates the phasic responses of dopamine neurons in encoding prediction errors for reward (Mirenowicz and Schultz 1994; Morris et al. 2006; Roesch et al. 2007; Schultz 1998). These neurons demonstrate a response profile that resembles the difference between expected and actual rewards associated with the presentation of particular stimuli and/or performance of specific actions (Mirenowicz and Schultz 1994; Montague et al. 1996; Roesch et al. 2007; Schultz 1998). Consistent with the animal data, human imaging studies have revealed evidence for reward prediction error signals in target areas of dopamine neurons in the human brain, most prominently in the ventral and dorsal striatum (Abler et al. 2006; McClure et al. 2003; O'Doherty et al. 2003, 2004; Pessiglione et al. 2006; Tanaka et al. 2004). Prediction error signals have been reported in humans during learning with a number of different kinds of reward, including money, liquid food rewards, water, and attractive faces (Bray and O'Doherty 2007; Kim et al. 2006; McClure et al. 2003; O'Doherty et al. 2003; Pessiglione et al. 2006).

An outstanding question is whether prediction error signals are distinct depending on the type of reward involved. Addressing this question is important for developing a better understanding of the computational mechanisms used by the brain to learn reward predictions. The presence of a generic reward prediction error signal that does not distinguish between different reinforcer types (whether biological compared with abstract rewards or visual versus flavor rewards) might be used to learn a nonspecific reward prediction signal that can indicate whether a particular stimulus or action is "rewarding" but nonetheless carries no information about the specific nature of the reward to be expected. On the other hand, if distinct prediction error signals are found to be recruited for different reinforcer types, then this could imply that the reward predictions trained by such signals might encode reinforcer specific information, thereby permitting motivationally specific modulation of predictive representations. Understanding the nature of prediction error signaling for different types of reinforcers is also relevant for evaluating theories of economic decision making predicated on the notion of a "common utility," whereby representations for the value or utility of different kinds of reinforcers are suggested to be encoded on a common scale to facilitate comparisons between actions leading to different types of reinforcers (Montague and Berns 2002; O'Doherty 2007). One way such common utility representations might be learned for different kinds of reinforcers is through a nondiscriminative generic reward prediction error signal. However, it is important to note that a nondiscriminative prediction error signal is only one of several mechanisms by which a common currency could be implemented. An alternative possibility is that reward-predictions for each reinforcer type could be learned separately via distinct prediction error signals and that these value predictions get converted into a common currency at the time of choice (i.e., independently of learning). Nevertheless if in the present study we find overlapping prediction error signals during learning, this could provide evidence in support of the possibility that predictive representations for different reinforcer types are integrated at the point of learning and not merely at the time of choice.

The degree to which reward prediction error signals distinguish between different reinforcer types has not been systematically addressed in either single-unit recording studies in animals or in functional imaging studies in humans. In direct recordings from dopamine neurons in animals, the range of reinforcers used has been limited to liquid foods and water, which does leave open the

Address for reprint requests and other correspondence: J. P. O'Doherty, Trinity College Institute of Neuroscience, Lloyd Building, Trinity College, University of Dublin, Dublin 2, Ireland (E-mail: odoherjp@tcd.ie).

0022-3077/09 $8.00 Copyright © 2009 The American Physiological Society

question whether firing patterns of dopamine neurons might distinguish between more diverse types of reinforcer. In humans, while brain regions such as ventral and dorsal striatum have been found to correlate with prediction error signals in response to a range of different reinforcers, no study to date has systematically compared and contrasted prediction error responses to different reinforcer types, leaving open the question whether prediction error representations are common or distinct as a function of reinforcer type.

The goal of the present study was to compare and contrast prediction error signals elicited while subjects learned to perform actions to obtain one of two distinct reinforcer types: a liquid food reward (juice), and monetary reward (monetary gain). Subjects were scanned with functional magnetic resonance imaging (fMRI) while performing a task during which they had to choose actions denoted by distinct stimulus pairs, which in one case led to either a high or low probability of obtaining a pleasant juice reward (1 ml of apple juice) and in another case a high or low probability of obtaining a monetary outcome (gaining 5 cents). We tested for regions showing significant correlations with prediction error signals separately during choices involving money and juice rewards to establish the extent to which prediction error signals elicited during learning with money and juice engage distinct or overlapping representations.

METHODS

Subjects

Seventeen healthy right-handed individuals [5 females, 12 males; mean age: 25 ± 1.7 (mean ± SD); range: 19–40] participated in the experiment. The subjects were preassessed to exclude those with a prior history of neurological or psychiatric illness. Prior to participation in the experiment, the subjects were prescreened to ensure that they found apple juice to be highly pleasant.

Subjects were asked to fast for ≥6 h prior to 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.

Stimuli

The liquid-food reward was apple juice, and its control was an affectively neutral tasteless solution that consists of the main ionic components of human saliva (25 mM KCl and 2.5 mM $NaHCO_3$). The liquids were delivered by means of separate electronic syringe pumps (1 for each liquid) positioned in the scanner control room. These pumps transferred the liquid stimuli to the subject via ~10-m-long polyethylene plastic tubes (6.4 mm diam), the other end of which were held between the subject's lips like a straw while they lay supine in the scanner. The monetary reward was indicated with a picture of a nickel to mean that five cents was added to the accumulating winnings, and its control was a scrambled picture of a nickel to mean that there was no change in income.

Task

The task consisted of four trial types: money, scrambled-money, juice, or neutral-solution, the occurrence of which was fully randomized throughout the experiment (see Fig. 1). On each trial, subjects were faced with the choice between two possible actions. Each trial type had unique pairs of arbitrary, affectively neutral, fractal stimuli representing those actions. The action of choosing one of the stimuli delivered the respective reward with a probability of 0.6, and the other delivered the same reward with a probability of 0.3. Subjects could choose a given action by selecting one of two button presses on a response pad corresponding to the left or the right location on the screen. The location assignment of the two images was fully counterbalanced across trials. The assignment of the images to each trial type and outcome probability was fully counterbalanced across subjects. The subjects' task on each trial was to choose one of the two possible available images. If a response was not registered before 1.5 s, a response omission was indicated to the subject, and the trial was aborted. When an image had been selected, it increased in brightness, and 4 s later the screen was cleared. Immediately following this, depending on the condition, the outcome was either the delivery of 1 ml apple juice or a picture of a nickel designating a 5-cent increase in the accumulating income; the delivery of 1 ml of affectively neutral control tasteless solution or a picture of a scrambled nickel designating no change to income; or else no stimulus was delivered or presented (according to the reward schedule associated with the particular stimulus chosen). The outcome lasted 1 s and was followed by a jittered intertrial interval drawn from a Poisson distribution with a mean of 4 s.

Experimental design

Before starting the experimental task, we collected pleasantness ratings of the visual cue stimuli (−5, very unpleasant; +5, very pleasant) based on subjective preference, and after the experiment we collected pleasantness ratings of the cues based on their learned associated outcomes. Subjects underwent two 25-min scanning sessions—two training sessions, each with different images (new learning) consisting of 160 trials (40 trials per condition: money, juice, scrambled, and neutral). There was a break in between the two sessions to allow subjects to rest. All four trial types were pseudorandomly intermixed throughout both of the sessions. Prior to each training session, subjects were told that there were four pairs of stimuli, and on each trial, one of these pairs would be displayed. They were instructed to select one of the possible visual stimuli on each trial by pressing the left or right response button. They were told that following their choices they could receive 1 ml of apple juice, 1 ml of a neutral solution, an image of a nickel designating winning 5 cents, a scrambled image designating no change, or nothing. They were not told which stimulus was associated with which particular outcome, but they were told that one of each pair of stimuli was associated with a higher probability of obtaining an outcome than the other. Subjects were instructed to learn to choose the stimuli that gave them the most reward.

fMRI data acquisition

The functional imaging was conducted by using a Siemens 3.0 Tesla Trio MRI scanner to acquire gradient echo T2* weighted echo-planar (EPI) images with blood-oxygenation-level-dependent (BOLD) contrast. To optimize functional sensitivity in orbitofrontal cortex (OFC), we used a tilted acquisition in an oblique orientation of 30° to the AC-PC line (Deichmann et al. 2003). In addition, we used an eight-channel phased array coil which yields a 40% signal increase in signal in the medial OFC over a standard head coil. Each volume comprised 32 axial slices. A total of 750 volumes (25 min) were collected during the experiment in an interleaved-ascending manner. The imaging parameters were: echo time, 30 ms; field of view, 192 mm; in-plane resolution and slice thickness, 3 mm; TR, 2 s. Whole-brain high resolution T1-weighted structural scans (1 × 1 × 1 mm) were acquired from the 17 subjects and co-registered with their mean EPI images and averaged together to permit anatomical localization of the functional activations at the group level. Image analysis was performed using SPM2 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). Temporal normalization was applied to the scans, each slice being centered to the middle of the scan (TR/2). To correct for subject motion, the images were realigned to the first volume, spatially normalized to a standard T2* template with a resampled voxel size of 3 mm, and spatially smoothed

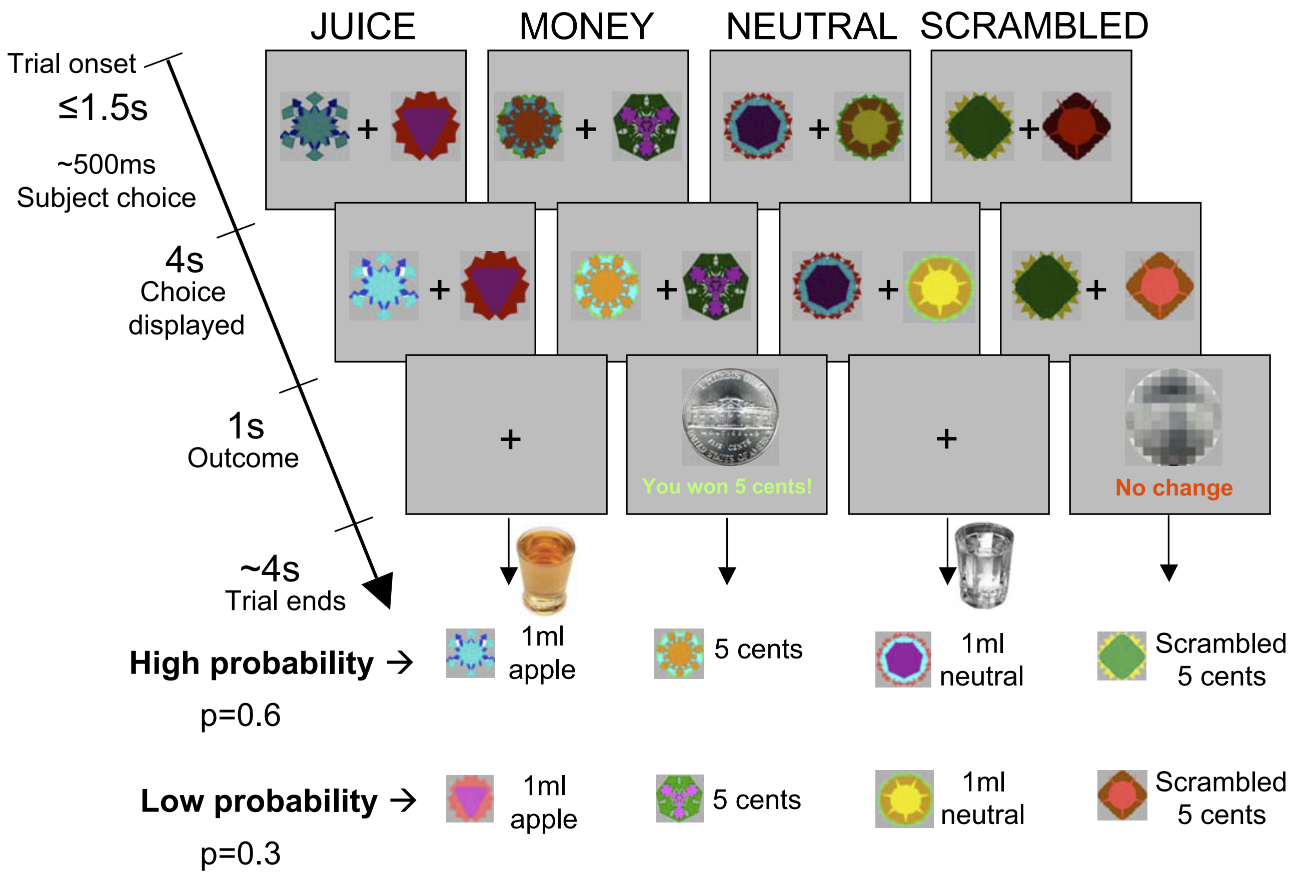

FIG. 1. Instrumental task illustration. Each of the 4 conditions, juice, money, neutral, and scrambled, were signified by different pairs of arbitrary stimuli. On each trial, subjects had to choose between 2 possible images, 1 leading to a high probability of outcome (0.6) and the other a low probability (0.3). Depending on the condition, the outcome was 1 ml apple juice, a picture of a nickel, 1 ml neutral control solution, or a picture of a scrambled nickel. The chosen stimulus was illuminated, and 4 s later the outcome was delivered.

using a Gaussian kernel with a full-width at half-maximum (FWHM) of 8 mm. Intensity normalization and high pass temporal filtering (using a filter width of 128 s) were also applied to the data (Friston et al. 1995).

Reinforcement learning model

We used a simple reinforcement learning model to learn action values Qa and Qb for each pair of actions a,b available for each of the four separate trial types (juice, money, and their corresponding control conditions). The action values were updated using a Rescorla-Wagner learning rule, following selection of that action on a given trial t

$$Q_{\mathrm{chosen}}(t+1) = Q_{\mathrm{chosen}}(t) + \alpha\delta(t)$$

Where $\delta(t)$ is a prediction error computing the difference between predicted and actual reward obtained on that trial

$$\delta(t) = \mathrm{r}(t) - Q_{\mathrm{chosen}}(t)$$

and α is a learning rate parameter.

We set $r(t)$ to 1 or 0 to denote receipt of a liquid-food/money outcome or no outcome, respectively. The probability of choosing a given choice option given the learned values is determined using a logistic sigmoid

$$p_{\mathrm{a}}(t) = \frac{\exp[\beta Q_{\mathrm{a}}(t)]}{\exp[\beta Q_{\mathrm{a}}(t)] + \exp[\beta Q_{\mathrm{b}}(t)]}$$

where β is an inverse temperature that determines the ferocity of the competition. To find optimal model parameters, we calculated the log likelihood fit of the actual choices made by subjects according to Rescorla-Wagner learning, for a variety of learning rates (α) and inverse temperature parameters (β). A single set of model parameters were fit to the group as a whole. These optimal parameters were obtained separately for different trial types and separately across runs: $\alpha = 0.06$ and $\beta = 10$ for money in run 1; $\alpha = 0.2$ and $\beta = 4.6$ for money in run 2; $\alpha = 0.02$ and $\beta = 3.8$ for scrambled money in run 1; $\alpha = 0.02$ and $\beta = 10$ for scrambled money in run 2; $\alpha = 0.06$ and $\beta = 10$ for juice in run 1; $\alpha = 0.1$ and $\beta = 7$ for juice in run 2; $\alpha = 0.02$ and $\beta = 5.6$ for neutral in run 1; $\alpha = 0.02$ and $\beta = 10$ for neutral in run 2. The above parameters were used to generate the actual regressors for the fMRI data analysis for each subject.

fMRI data analysis

The event-related fMRI data were analyzed by creating regressors composed of sets of delta (stick) functions. For each trial type, we approximated a full temporal difference prediction error signal (with cue and outcome responses) by using a regressor as a parametric modulator in which the signal at the time of cue presentation was set to the value of the chosen action $Q_{\mathrm{chosen}}(t)$, while the activity at the time of outcome corresponded to $\delta(t)$. Additional regressors modeled activity corresponding to receipt of reward at the time of the outcome (which was set to 1 if a reward was delivered, 0 otherwise). Separate regressors modeled activity for each trial type. Regressors for omitted trials and onsets of swallowing events were also created. All of these regressors were convolved with a hemodynamic response function. In addition, the six scan-to-scan motion parameters produced during realignment were included to account for residual effects of movement. A swallowing-motion parameter was also recorded via a motion detection coil placed on each subjects' neck. These regressors were then entered into a regression analysis against the fMRI data for each individual subject. Linear contrasts of regressor coefficients were computed at the single subject level to enable comparison between the juice, neutral, money, and scrambled trial types. The results from each subject were taken to a random effects level by including the contrast images from each single subject into a one-way ANOVA with no mean term. The main contrasts reported in this study were between the prediction error signals of juice minus neutral and money minus scrambled. To find commonalities, we took the average (main effect of prediction error) and the conjunction of these contrasts. The conjunction was computed by applying a statistical threshold to one of the prediction error (PE) contrasts ([juice – neutral]) and then applying the other PE contrast ([money – scrambled]) as an inclusive mask using the same threshold (e.g., at $P < 0.001$). To find differences, we looked for an interaction between these contrasts: [money – scrambled] – [juice – neutral]. We also tested for regions responding following receipt of the different outcomes, computing conjunctions and differences in the same way as described in the preceding text for the prediction error contrasts. We also modeled a tonic value signal set to $Q_{\text{chosen}}(t)$, which began at the cue presentation and ended at the time of outcome presentation.

The structural T1 images were co-registered to the mean functional EPI images for each subject and normalized using the parameters derived from the EPI images. Anatomical localization was carried out by overlaying the *t*-maps on a normalized structural image averaged across subjects and with reference to an anatomical atlas. When testing for prediction error signals, we focused in particular on the ventral and dorsal striatum and report results in this area using an uncorrected threshold of $P < 0.001$ and an extent threshold of five contiguous voxels. Activities in all other regions are reported only if surviving correction for multiple comparisons at whole brain level using family-wise error at $P < 0.05$.

RESULTS

Behavioral results

EFFECTS OF TRAINING DURING INSTRUMENTAL CONDITIONING. Figure 2*A* shows averaged learning curves for the high probability actions associated with apple juice, neutral, money, and scrambled outcomes over the course of the two training sessions and across the 17 subjects. In the last 10-trial block of training (averaged from each session), subjects chose the high probability action significantly more often than the low probability action in both the juice [$t(16) = 3.82$, $P < 0.001$, 1-tailed] and the money [$t(16) = 5.30$, $P < 0.001$, 1-tailed] conditions. This indicates that subjects learned to choose the instrumental action associated with the most reward in both conditions. On the contrary, subjects did not learn to choose the high probability action more than the low probability action in the last block of the neutral condition [$t(16) = 0.13$, $P = 0.90$, 2-tailed], and the scrambled condition [$t(16) = 0.71$, $P = 0.49$, 2-tailed], indicating that subjects were indifferent as to whether they obtained the affectively neutral control stimuli. In addition, we found no significant difference in the number of high probability action choices made in the juice and money conditions in the last 10 trials [$t(16) = -0.31$, $P = 0.76$, 2-tailed].

Effects of training on the affective evaluation of visual stimuli

Subjective pleasantness ratings for the eight different stimuli (images associated with the low and high probability outcomes for each of the 4 trial types) were taken before and after both sessions of training and their average from the two sessions are plotted in Fig. 2*B*. The subjective pleasantness of the stimuli associated with high probability of money or juice outcomes increased from before to after conditioning [money-high: $t(16) = -1.95$, $P < 0.05$, 1-tailed; juice-high: $t(16) = -1.85$, $P < 0.05$, 1-tailed], whereas the pleasantness of the images associated with low probability of money or juice outcomes decreased [money-low: $t(16) = 2.73$, $P < 0.01$, 1-tailed; juice-low: $t(16) = 2.75$, $P < 0.01$, 1-tailed]. These results indicate that subjects' subjective affective evaluations for the visual stimuli associated with a high probability of reward were increased from before to after conditioning for both food and money rewards. On the other hand, for the nonrewarding control trials, the subjective pleasantness of the stimuli decreased from before to after, regardless of the probability of outcome with which that stimulus was associated [neutral-high: $t(16) = 1.53$, $P = 0.07$, 1-tailed; scrambled-high: $t(16) = 3.41$, $P < 0.01$, 1-tailed; neutral-low: $t(16) = 2.31$, $P < 0.05$, 1-tailed; scrambled-low: $t(16) = 4.03$, $P < 0.001$, 1-tailed]. Furthermore, the increase in subjective pleasantness of the stimuli associated with high probability of money was not

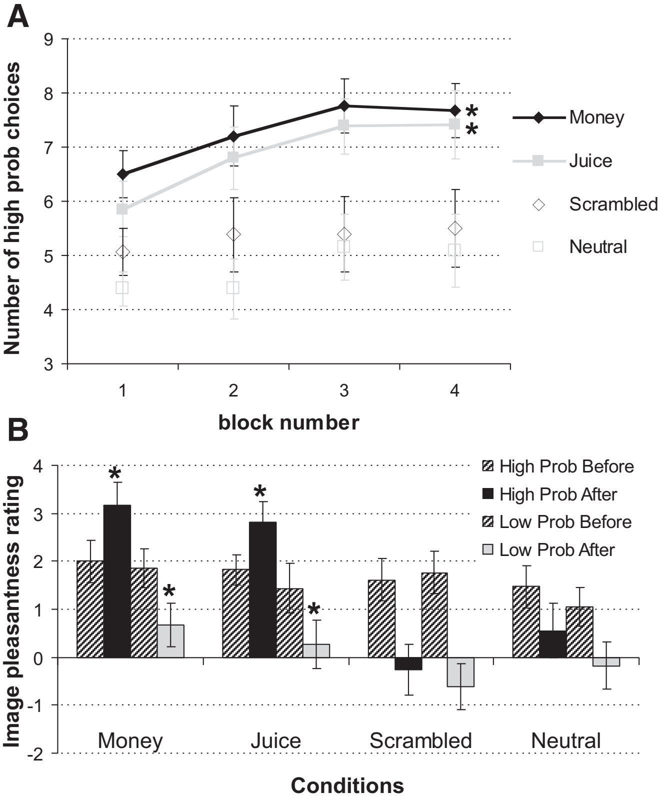

FIG. 2. *A*: learning curves. Total number of high probability action choices over 4 10-trial blocks shown averaged across 17 subjects and 2 learning sessions. Over the course of training, subjects increasingly favored the high probability images associated with apple juice or money over their low probability counterparts, but this was not the case for the neutral and scrambled condition where subjects were indifferent between the high and low probability actions (*, $P < 0.001$, 1-tailed). *B*: subjective pleasantness ratings of images on a scale of −5 (very unpleasant) to +5 (very pleasant), before and after training, averaged across 17 subjects and two learning sessions. The ratings for the high probability reward conditions significantly increased, and the low probability reward conditions significantly decreased, but this was not the pattern in the control conditions (*, $P < 0.05$, 1 tailed).

significantly different from that of juice [$t(34) = 0.74$, $P = 0.46$].

Neuroimaging results

COMMONALITIES IN PREDICTION ERROR CODING BETWEEN REINFORCERS. To identify striatal regions correlating with prediction error for both juice and money, we first averaged the prediction error contrasts: juice-neutral and money-scrambled. This analysis revealed significant effects in a number of striatal subregions: the nucleus accumbens [15, 12, −6 mm, $z = 3.90$, $P < 0.001$], and dorsal striatum (caudate nucleus) [−9, 3, 15 mm, $z = 4.22$, $P < 0.001$; see Fig. 3*A*]. Figure 3,*B* and *C,* also shows the separate results from the individual juice-neutral and the money-scrambled prediction error contrasts respectively.

We also tested for striatal areas showing significant effects in the conjunction of juice-neutral and money-scrambled prediction error contrasts. This analysis revealed significant effects in the dorsal striatum, specifically anterior caudate nucleus [-9, 0, 6 mm; $z = 3.34$, $P < 0.001$]. These results are tabulated in Supplementary Table S1, and time course plots from this region are shown in Supplementary Fig. S1.[1]

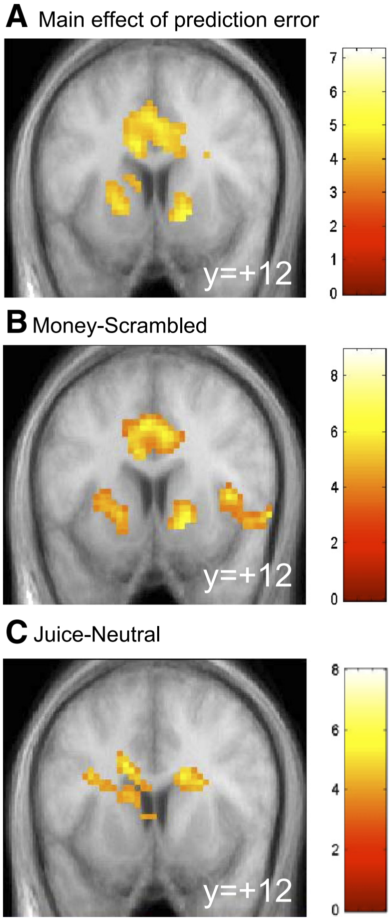

FIG. 3. Area of striatum showing prediction error responses to the different reinforcers. *A*: average prediction error (PE) signal combined across money and juice trials (from the contrasts of juice PE − neutral PE and money PE − scrambled PE) showing activity in both ventral and dorsal striatum. *B*: regions of striatum correlating with reward prediction errors during money trials (from the contrast of money PE − scrambled PE). *C*: regions of ventral and dorsal striatum correlating with reward prediction errors during juice trials (from the juice PE − neutral PE contrast). Images are shown with threshold set at $P < 0.001$.

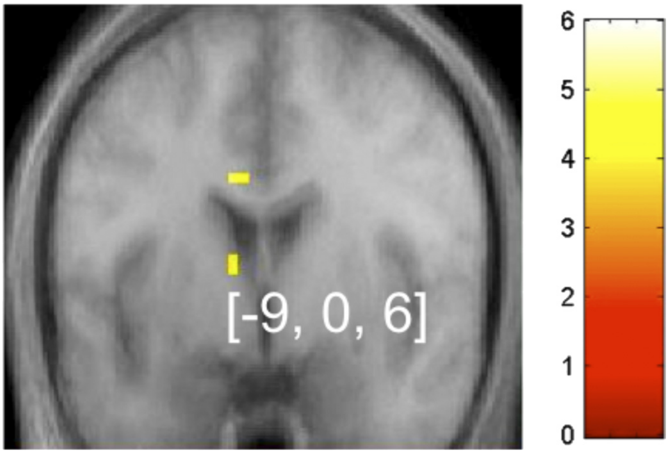

FIG. 4. Region of dorsal striatum showing common prediction error coding for juice and money reward in a conjunction of prediction error contrasts: (juice PE − neutral PE) and (money PE − scrambled PE). Image is shown with threshold set at $P < 0.001$.

Differences in prediction error coding between reinforcers

Next we looked for regions that were specific to a reinforcer type (Fig. 4). To test for areas responding more during money than juice prediction errors, we tested for the following contrast: (money PE − scrambled PE) − (juice PE − neutral PE). This analysis revealed significant effects in a region of the right nucleus accumbens [15, 9, −6 mm, $z = 3.58$, $P < 0.001$], suggesting that this region is significantly more correlated with prediction errors in the money than the juice conditions (see Fig. 5; Table S2). No areas within striatum showed significantly stronger effects for juice compared with money PEs.

Separate prediction errors at the time of cue and outcomes

We also ran an additional analysis in which we split the full PE signal into two components, one at the time of cue presentation and the other at the time of outcome presentation. For the outcome PE regressors, we obtained broadly similar results in the striatum for the conjunction and difference contrasts to that seen in the full PE case, albeit only at a trend level ($P < 0.01$), whereas for the cue PE regressors, we did not find such effects. These results suggest that the effects of the full PE analysis

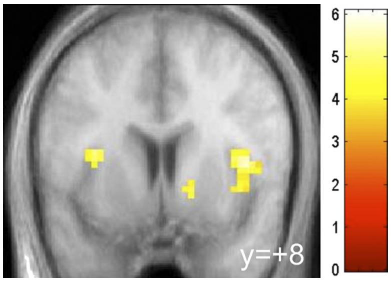

FIG. 5. Region of nucleus accumbens showing differential prediction error coding between reinforcers. Right nucleus accumbens was significantly more correlated with money prediction errors than with juice prediction errors. Image is shown with threshold set at $P < 0.001$.

[1] The online version of this article contains supplemental data.

may be especially driven by the prediction error signal at the time of outcome, although a combined analysis of the error signals across both time points is necessary to reach our threshold for significance in the striatum.

Testing for the contribution of outcome responses to prediction error effects

To rule out the contribution of responses to outcomes independently of prediction errors on the above findings, we ran a further analysis in which the receipt of juice, money and neutral and scrambled outcomes were modeled as separate regressors. The prediction error regressors were then orthogonalized with respect to those outcome regressors ensuring that the outcome regressors were assigned all the variance common to prediction errors and outcomes. We then performed a conjunction analysis across the money and juice outcome regressors: i.e., (money − scrambled) *and* (juice − neutral). We did not find any significant effects in dorsal striatum in this contrast at $P < 0.001$ or even when the conjunction between both contrasts was tested at $P < 0.01$.

We also tested for a difference between responses to the money and juice outcomes: (money − scrambled) − (juice − neutral). Again we did not find any significant effects in ventral striatum (nucleus accumbens) at $P < 0.001$ (or at $P < 0.01$) in this analysis as had been found for the prediction error analysis. Taken together these results indicate that differences in the sensory properties of outcomes (whether between money and scrambled or juice and neutral, or between money and juice) are unlikely to account for the prediction error findings described earlier. However, because this conclusion is based on a null result, the contribution of sensory differences in outcome coding to the prediction error results cannot be completely ruled out.

DISCUSSION

Here we provide evidence with fMRI that prediction error signals during learning with food and money reward are at least partially overlapping in humans. Specifically within the striatum, a part of dorsal striatum (caudate nucleus) was found to exhibit prediction error signals during learning with both money and juice reward. In addition to our finding of overlapping prediction error representations in some areas, we also found evidence for partially distinct prediction error representations in other areas. In particular, a region of right nucleus accumbens, part of the ventral striatum, was found to be engaged by prediction errors during learning with money but not juice reward.

These results have important implications for understanding the computational mechanisms underlying learning of reward associations in humans. It is notable that dorsal striatum should turn out to be the locus of common prediction error responses given that this region is thought to contribute toward guiding action selection for reward. According to actor-critic models of reinforcement learning, action selection for reward is mediated via the "actor," which learns a policy through the formation of stimulus-response associations. These associations (described as habitual in the animal learning literature) are, once formed, insensitive to the current incentive value of the outcomes used to stamp in such associations in the first place (Barto 1992, 1995; Daw et al. 2005; Dayan and Balleine 2002; Montague et al. 2006; O'Doherty et al. 2004). Although here we do not test whether the instrumental actions being performed by subjects are under goal-directed or habitual control (Balleine and Dickinson 1998), one plausible interpretation of the present results is that the overlapping prediction error signals we found in dorsal striatum correspond to the site of learning of habitual stimulus-response associations in which there is no explicit representation of the outcome (Daw et al. 2005).

Another feature of actor-critic models of reinforcement learning is the "critic," which learns about the expected future rewards arising from being in a particular context (usually denoted by specific stimulus configurations). The critic has been proposed to learn Pavlovian stimulus-outcome associations. Unlike habitual stimulus-response associations in the actor (Balleine and Dickinson 1998; Balleine et al. 2008; Dickinson 1985), Pavlovian associations are known to be sensitive to the value of the outcome with which they are associated in that devaluation of outcomes can result in a decrease in cue-elicited Pavlovian conditioned responses (Holland and Straub 1979). Ventral striatum has previously been implicated in learning of stimulus-outcome associations in humans as prediction error signals in this area have been reported during studies of appetitive classical conditioning with a range of different reinforcer types (Bray and O'Doherty 2007; Kim et al. 2006; O'Doherty et al. 2003; McClure et al. 2003). Indeed ventral striatum has been suggested to be a candidate region for implementing the critic.

In the present study, while we found a main effect of reward prediction error in ventral striatum (on average across both juice and money conditions), a direct comparison between money and juice prediction errors revealed significantly stronger correlations with money prediction errors in the nucleus accumbens compared with juice prediction errors. These findings could be taken to indicate that a part of ventral striatum is specifically recruited in response to prediction errors for money and not juice. However, because we did not observe significant prediction error signals in ventral striatum in the juice condition alone, the preceding findings should be interpreted with caution. Many previous studies have reported significant prediction related signals in ventral striatum during learning with juice reward (McClure et al. 2003; O'Doherty et al. 2003, 2004). Therefore it is certainly not the case when taking into account these previous findings that ventral striatum is exclusively engaged during learning with money reward. However, prediction errors for juice reward have in other studies been found to correlate with activity in more lateral regions of ventral striatum, particularly in the ventral parts of putamen (O'Doherty et al. 2003, 2006), whereas money prediction errors often tend to be located more medially within the nucleus accumbens proper (Abler et al. 2006; Hare et al. 2008; Haruno et al. 2006). Our finding of enhanced activity within nucleus accumbens for money prediction errors could be consistent with this trend. It also been reported that prediction errors during learning with a rewarding visual stimulus (attractive faces) elicited activity in medial ventral striatum (Bray and O'Doherty 2007). Taken together these results suggest the possibility that medial ventral striatum may be preferentially engaged by prediction errors for either abstract or visual reinforcers compared with juice. Relatively greater involvement of medial aspects of the ventral striatum in responding to money or visual reinforcers could occur due to the pattern of Pavlovian conditioned responses being elicited through learned association with such stimuli, which may be preparatory (such as eliciting approach) rather than consummatory as would likely be the case during conditioning with juice reward (Konorski 1948). Future studies could address this possibility directly by measuring different types of Pavlovian-conditioned responses during associative learning with money, visual, and juice rewards and comparing these to the pattern of activation in the striatum.

An alternative explanation for the more robust engagement of nucleus accumbens during the money reward condition is that money could be deemed more valuable to subjects than juice reward and that stronger prediction errors in ventral striatum in the money condition relates to a difference in overall salience between the reward conditions. However, several pieces of evidence argue against this interpretation in the case of the present results. First of all, robust prediction error signals were found in dorsal striatum in response to both food and money reward, which indicates that prediction error signals are not generally stronger in the money compared with the food conditions. Moreover, the extent to which subjects favored the high compared with low probability actions within each condition is a behavioral measure of preference for the rewards. Notably, subjects learned to favor the high probability action for the money and juice reward conditions and did not differ between these conditions in the proportion of choice allocations to the high probability actions. On the other hand, subjects were indifferent in their choices between the high and low probability actions for both neutral conditions. These results therefore suggest that subjects value both the food and money rewards equally highly. Finally, subjective pleasantness evaluations for the stimuli associated with the high probability actions were significantly elevated in both juice and money conditions, and no significant difference was found between the stimulus evaluations in the food and money conditions, suggesting that subjectively, both reward contexts were rated as similarly favorable by the subjects.

It is also important to note that the present study can only provide indirect evidence about the role of dopamine in prediction error signaling for different reinforcers. The prediction error signals we are measuring in the target areas of dopamine neurons such as striatum likely depends at least partially on dopaminergic input into these areas (Pessiglione et al. 2006), but it is also likely that other neural signals contribute to the observed prediction-error-related activity in these regions, including intrinsic activity within striatum as well as other nondopaminergic inputs. Nevertheless, the finding of only partially overlapping prediction error BOLD correlates for the different reinforcers raises the question as to whether dopamine neurons do indeed at least partially differentiate between reinforcer types. The results of the present study suggest the importance of following up this question using techniques that allow selective measurement of dopamine signals including single-unit neurophysiology recordings from dopamine centers in the midbrain and/or voltammetric recordings from some of the dopaminoceptive target areas identified in the present study.

The main conclusions of the present study are that a region of dorsal striatum correlates with prediction errors during learning with both juice and monetary rewards, suggesting that this area is involved in mediating learning of stimulus-response associations irrespective of the nature of the reward being used to reinforce such instrumental associations. While prediction error activity was largely overlapping within the dorsal striatum, we also found some evidence for partial specificity within the ventral striatum in that the nucleus accumbens showed stronger responses to prediction errors during learning with money compared with juice reward, which may suggest at least partial differentiation of ventral striatal circuitry involved in learning Pavlovian associations.

ACKNOWLEDGMENTS

We thank I. Kagan, S. Wagner, and M. Tyszka for technical assistance and advice.

GRANTS

This research is supported by a Searle Scholarship and a grant from the National Science Foundation to J. P. O'Doherty.

REFERENCES

Abler B, Walter H, Erk S, Kammerer H, Spitzer M. Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage* 31: 790–795, 2006.

Balleine BW, Daw ND, and O'Doherty J P. Multiple forms of value learning and the function of dopamine. In: *Neuroeconomics: Decision Making and the Brain*, edited by Glimcher PW, Camerer CF, Poldrack RA, Fehr E. New York: Academic Press, 2008, p. 367–388.

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

Barto AG. Reinforcement learning and adaptive critic methods. In: *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, edited by White DA, Sofge DA. New York: Van Norstrand Reinhold, 1992, p. 469–491.

Barto AG. Adaptive critics and the basal ganglia. In: *Models of Information Processing in the Basal Ganglia*, edited by Houk JC, Davis JL, Beiser BG. Cambridge, MA: MIT Press, 1995, p. 215–232.

Bray S, O'Doherty J. Neural coding of reward-prediction error signals during classical conditioning with attractive faces. *J Neurophysiol* 97: 3036–3045, 2007.

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

Dayan P, Balleine BW. Reward, motivation, and reinforcement learning. *Neuron* 36: 285–298, 2002.

Deichmann R, Gottfried JA, Hutton C, Turner R. Optimized EPI for fMRI studies of the orbitofrontal cortex. *Neuroimage* 19: 430–441, 2003.

Dickinson A. Actions and habits: the development of a behavioural autonomy. *Philos Trans R Soc Lond B Biol Sci* 308: 67–78, 1985.

Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R. Analysis of fMRI time-series revisited. *Neuroimage* 2: 45–53, 1995.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A. Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J Neurosci* 28: 5623–5630, 2008.

Haruno M, Kawato M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Neurophysiol* 95: 948–959, 2006.

Holland PC, Straub JJ. Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. *J Exp Psychol Anim Behav Process* 5: 65–78, 1979.

Kim H, Shimojo S, O'Doherty JP. Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biol* 4:e233, 2006.

Konorski, J. *Conditioned Reflexes and Neuron Organization.* Cambridge, UK: Cambridge Univ. Press, 1948.

McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. *Neuron* 38: 339–346, 2003.

Mirenowicz J, Schultz W. Importance of unpredictability for reward responses in primate dopamine neurons. *J Neurophysiol* 72: 1024–1027, 1994.

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

Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936–1947, 1996.

Montague PR, King-Casas B, Cohen JD. Imaging valuation models in human choice. *Annu Rev Neurosci* 29: 417–448, 2006.

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

O'Doherty JP. Lights, camembert, action! The role of human orbitofrontal cortex in encoding stimuli, rewards, and choices. *Ann NY Acad Sci* 1121: 254–272, 2007.

O'Doherty JP, Buchanan TW, Seymour B, Dolan RJ. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron* 49: 157–166, 2006.

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

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

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behavior in humans. *Nature* 442: 1042–1045, 2006.

Roesch MR, Calu DJ, Schoenbaum G. Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci* 10: 1615–1624, 2007.

Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 80: 1–27, 1998.

Tanaka SC, Doya K, Okada G, Ueda K, Okamoto Y, Yamawaki S. Prediction of immediate and future rewards differentially recruits cortico-basal ganglia loops. *Nat Neurosci* 7: 887–893, 2004.