Neuron 49, 157–166, January 5, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.neuron.2005.11.014

Predictive Neural Coding of Reward Preference Involves Dissociable Responses in Human Ventral Midbrain and Ventral Striatum

John P. O'Doherty,[1,2,*] Tony W. Buchanan,[3] Ben Seymour,[1] and Raymond J. Dolan[1]
[1] Wellcome Department of Imaging Neuroscience
Institute of Neurology
University College London
12 Queen Square
London WC1N 3BG
United Kingdom
[2] Division of Humanities and Social Science
California Institute of Technology
Pasadena, California 91125
[3] Department of Neurology
University of Iowa
Iowa City, Iowa 52242

Summary

Food preferences are acquired through experience and can exert strong influence on choice behavior. In order to choose which food to consume, it is necessary to maintain a predictive representation of the subjective value of the associated food stimulus. Here, we explore the neural mechanisms by which such predictive representations are learned through classical conditioning. Human subjects were scanned using fMRI while learning associations between arbitrary visual stimuli and subsequent delivery of one of five different food flavors. Using a temporal difference algorithm to model learning, we found predictive responses in the ventral midbrain and a part of ventral striatum (ventral putamen) that were related directly to subjects' actual behavioral preferences. These brain structures demonstrated divergent response profiles, with the ventral midbrain showing a linear response profile with preference, and the ventral striatum a bivalent response. These results provide insight into the neural mechanisms underlying human preference behavior.

Introduction

Choosing between different available foods reflects an elementary form of decision making likely to be of crucial adaptive significance in natural environments. Such decisions are governed in part by individual preferences that are in turn shaped by prior experience. In order to implement decisions about what foods to consume, it is necessary to be able to associate foods with cues in the environment that predict their likely occurrence. Understanding the mechanisms by which the brain learns and encodes preference predictions is important not only for understanding the neural mechanisms of decision making but also for deriving neural markers that predict subsequent preference behavior. A recent study reported cultural modulation of neural and behavioral responses to presentation of two drink brands, by presenting cues relating to brand logos prior to subsequent delivery of the associated drink stimuli (McClure et al., 2004). These cues strongly influenced subjects' choice behavior, suggesting a powerful influence of prior learned predictions on actual choice behavior.

A putative model for learning appetitive and aversive predictions is temporal difference learning (Schultz et al., 1997; Sutton and Barto, 1990). This method involves a prediction error signal that indicates discrepancies between successive predictions of future reward. According to this model, trial-by-trial learning is reflected by a shift in this prediction error signal from the time at which the reward is delivered back to the time at which the predictive cue is first presented. Neurophysiological studies in nonhuman primates indicate that phasic activity in dopamine neurons is a possible neural substrate for this prediction error signal (Hollerman and Schultz, 1998; Montague et al., 1996; Schultz, 1998). Recent neuroimaging studies of appetitive learning indicate the presence of prediction error signals in prominent target structures of dopamine neurons, such as the ventral striatum (nucleus accumbens and ventral putamen) and orbitofrontal cortex (McClure et al., 2003; O'Doherty et al., 2003). Significant prediction error-related activity has also been observed in ventral striatum during aversive learning (Seymour et al., 2004). However, it is not yet clear whether predictive activity in these brain areas is directly related to subjects' actual preference behavior. A key test of the hypothesis that reward predictions underlie behavioral decisions would be to determine whether there is a direct link between neural activity encoding such reward predictions and actual behavioral preferences.

To address this question, we determined subjects' preferences for five different food "flavors": blackcurrant juice, melon juice, grapefruit juice, carrot juice, and a tasteless and odorless control solution. We determined overall preference ranks for each flavor by repeatedly presenting pairs of foods, incorporating all possible combinations of pairs, and recording their preferences on each occasion. Following this, we scanned subjects using fMRI in a Pavlovian conditioning procedure during which they were presented with five different arbitrary visual cues, each of which was reliably associated with the subsequent presentation (5 s later) of one of the five specific foods (Figure 1). We predicted that the (previously neutral) visual cues would come to acquire predictive values according to the subject-specific preferences of the food they predicted.

Multiple behavioral measures were used to provide evidence for Pavlovian conditioning. We used an online measure of subjects' pupillary dilation during the anticipatory interval, after the cue had been delivered but before the juice was presented, to provide an index of learned anticipatory arousal. On each trial, the visual cues were presented on either the left or the right of a fixation cross, and we asked subjects to respond by using a key press to indicate on which side the cue had been presented (before the juice was delivered), enabling us to determine whether reaction times had been modulated by presentation of the cues associated with the

*Correspondence: jdoherty@hss.caltech.edu

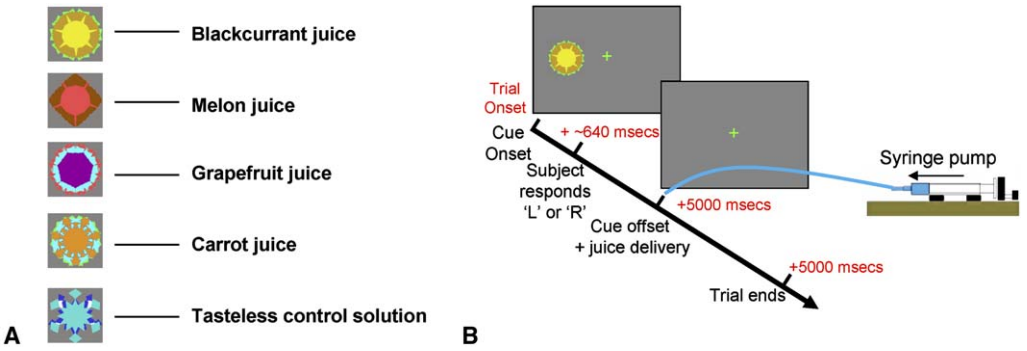

Figure 1. Task Illustration

(A) Illustration of fractal stimuli used in the experiment. Each fractal was paired with a different flavor stimulus. An example pairing is shown here (the actual pairings were counterbalanced across subjects).

(B) Illustration of timeline within a trial. At the beginning of each trial, a cue stimulus was presented on either the left or right side of a fixation cross. A subject's task was to indicate via a button box on which side of the screen the stimulus had been presented. Five seconds later, the cue stimulus presentation was terminated, and at the same time 0.7 ml of the relevant flavor stimulus was delivered intra-orally. A further 5 s later the next trial was triggered.

most or least preferred food stimulus. We also obtained affective ratings for the visual cues before and after the experiment to determine whether subjects changed their affective evaluation of the visual stimuli as a function of the specific juice with which the stimulus had been associated.

To establish whether reward-predicting responses in key brain structures relate to subjects' behavioral preferences, we used a temporal difference learning model to derive a reward-prediction signal that captures the transfer of activity back from the time of reward in the early trials to the time of presentation of the cue in the late trials (Montague et al., 1996; O'Doherty et al., 2003; Sutton, 1988). We correlated this modeled signal with trial-by-trial fMRI data, separately for each cue-food association, and tested for brain regions in which predictive responses to the cues were modulated as a function of subjects' individual preferences. We tested for two types of response profile: activity in which the predictive response scaled linearly with preference (where the fMRI signal increases linearly with increasing preference), and activity which shows a bivalent response profile, with a maximal response to the cues associated with the most and least preferred foods compared to a cue associated with a middle ranked preferred food. We looked for significant effects in a number of key brain structures that have been implicated in reward and reward-related learning: the ventral striatum (incorporating the ventral putamen as well as the nucleus accumbens proper), the midbrain (in the vicinity of the dopaminergic nuclei), the amygdala, and orbitofrontal cortex (O'Doherty, 2004).

Results

Behavioral Results

Pupillary Dilation

To test for significant differential effects in anticipatory pupil dilation following presentation of the cue stimuli as a function of preference, we performed a repeated-measures ANOVA with one factor preference (from most to least preferred), another factor experimental session (session 1 versus session 2), and another factor time within a trial (mean pupilometry responses were binned into five 1 s long epochs from the time of presentation of the cue up until immediately before delivery of the juice). We observed a significant preference × session × time interaction ($F[8,80] = 21.6$, $p < 0.001$), indicating a significant effect of cue preference on anticipatory pupil dilation, evident by the second block of trials. Post hoc analyses revealed that the most and least preferred trials were associated with a significant increase in anticipatory pupil dilation relative to the middle preferred trials (at $p < 0.05$). This provides evidence of increased arousal due to anticipation of the subsequent presentation of the most and least preferred stimuli (Figure 2A).

Reaction Times

We tested for differences in reaction times in responses made to the most and least preferred cues. A two-way repeated-measures ANOVA of median reaction times with one factor preference (Most versus Least Preferred) and the other factor experimental session (session 1 versus session 2) revealed a significant session × preference interaction ($F[1,11] = 5.9$, $p < 0.05$), indicating that responses to the cue associated with the most preferred stimulus were significantly faster than the cue associated with the least preferred stimulus by the second block of trials (Figure 2B).

Affective Evaluation of the Cue Stimuli

We next compared subjects' affective evaluations of the cue stimuli (using a pleasantness scale ranging from −10 [very unpleasant] to +10 [very pleasant]), before and after the experiment. We used a two-way repeated-measures ANOVA with one factor preference (most and least preferred cues) and the other factor session (before and after). The interaction between preference and session approached significance at $p < 0.05$ ($F[1,11] = 4.7$, $p = 0.052$). A post hoc t test revealed that the ratings to the cue associated with the most preferred stimulus were significantly greater than that made to the least preferred stimulus by session 2 ($t[11] = 2.4$, $p < 0.05$). These results suggest that subjects moderated their affective evaluation of the least and most preferred cue stimuli as a function of conditioning. Ratings for the stimuli associated with the least preferred food went

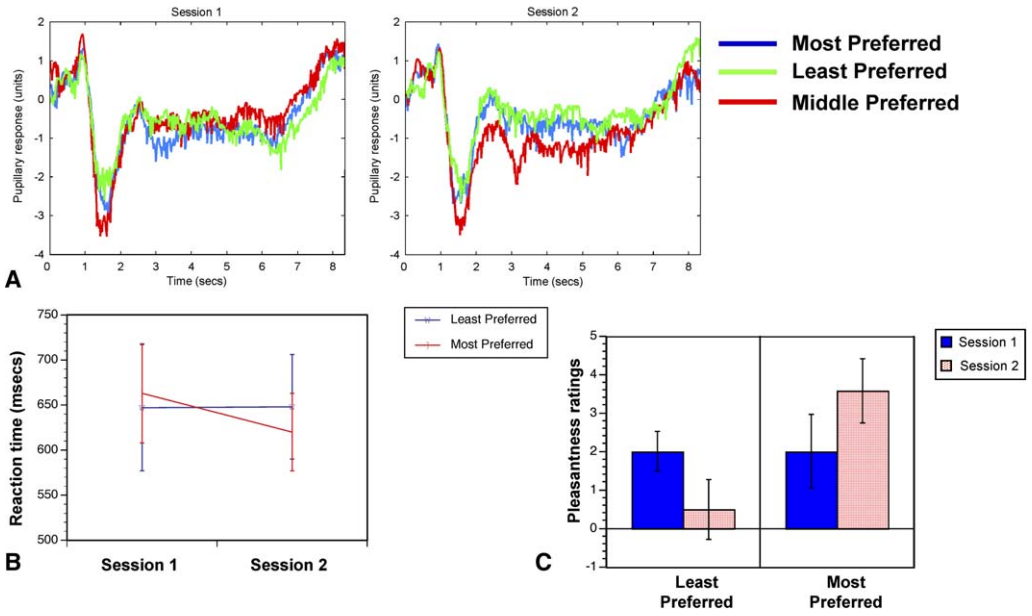

Figure 2. Multiple Behavioral Measures of Learning

(A) Trial averaged pupil dilation for the most, least, and middle preferred stimuli shown separately for both experimental sessions. Anticipatory pupil dilation responses were significantly greater to the cues associated with the most and least preferred stimuli than that to cue associated with the middle preferred stimuli by the second session.

(B) Reaction times for the cues associated with the most and least preferred stimulus plotted for both experimental sessions. The data plotted are the average across subjects of the median reaction time for that cue in each individual subject. Error bars depict standard error of the mean. A significant difference in reaction times to these cues emerged by the second session.

(C) Pleasantness ratings for the cues associated with the most and least preferred stimuli. By the second session, there was a significant difference in pleasantness ratings ascribed to the cues associated with the most and least preferred stimuli.

down across sessions, while ratings for the stimuli associated with most preferred went up, demonstrating the development of learning over the sessions (see Figure 2C).

Imaging Results

Responses Scaling Linearly with Preference

Our imaging data indicated that in one of our a priori regions of interest there was a response profile that showed a significant correlation with the predictive signal arising from the temporal difference model. Responses in this area scaled linearly with preference. The region we identified corresponds to the ventral midbrain, in the vicinity of one of the main sites of origin of dopaminergic ascending projection systems: the ventral tegmental area (Figures 3A–3C). The subject averaged parameter estimates (from the peak voxel at the group level) are shown in Figure 3D, illustrating the linear trend. To provide further validation of the linearity of the response, we extracted the fitted parameter estimates from the peak voxels in this region from each individual subject and performed a forward stepwise linear regression procedure to test for a significant linear trend in the data as a function of preference. In the stepwise procedure, we also included a more elaborate model incorporating a quadratic response profile (symmetrical around the middle-preferred food) in addition to the linear profile. This procedure found a significant linear trend in the data ($F[1,63] = 20.4349$; $p < 0.001$; $r^2 = 0.242$), and the more elaborate model incorporating the quadratic response did not significantly account for any additional variance (at $p < 0.05$). These data are plotted in Figure 3E. Two outliers are present in the data, on account of the parameter estimates from one particular single subject (for the regressors corresponding to the most and second most preferred stimuli). When these outliers are removed, the linear fit is even more significant ($F[1,62] = 22.726$; $p < 0.001$; $r^2 = 0.271$). Subject averaged percent signal change time-course plots from the ventral midbrain are shown in Figure 4C (alongside model-predicted time courses in Figures 4A and 4B), for each category of preferred stimulus (ranked from most to least preferred), separately for early (first six) and late (last six) trials of each trial type ranked according to preference (from most to least preferred).

No other regions of interest showed significant activity at $p < 0.001$, but linear correlations with preference were found in left amygdala ($-18, -3, -24$; $z = 2.96$; $p < 0.002$) and in medial orbitofrontal cortex ($0, 30, -18$; $z = 2.79$; $p < 0.005$) just below the threshold for significance. We report these results for completeness but do not discuss them further, as they did not reach our significance criterion. Even though we used imaging techniques designed to recover signal in areas with dropout such as orbitofrontal cortex and medial temporal lobes, we cannot rule out the possibility that signal dropout in regions such as orbitofrontal cortex or amygdala contributed to the weak effects in these regions.

In addition to the analysis performed with preference rankings, we also conducted an analysis in which we tested for brain regions in which predictive signals from the temporal difference model scaled according to the

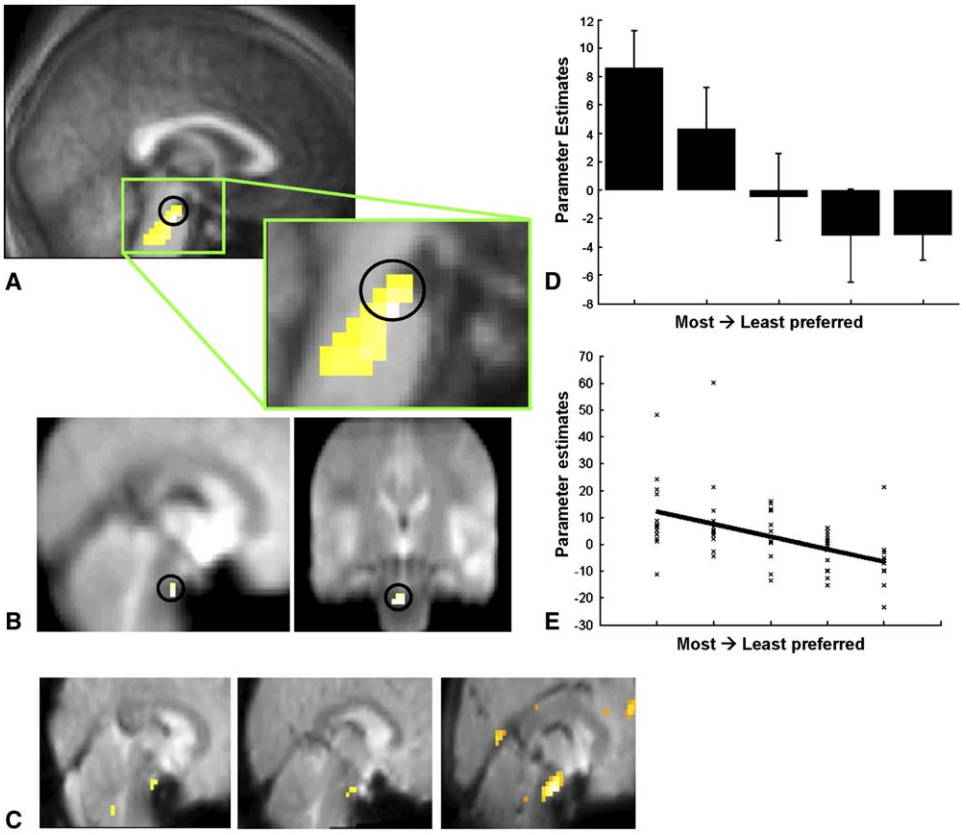

Figure 3. Preference Responses in Ventral Midbrain

(A) Predictive responses in ventral midbrain in the vicinity of the VTA scaling as a linear function of behavioral preference. Predictive responses in this region emerged over the course of learning, responding initially to the food stimulus itself and then transferring back to the cue stimuli by the end of learning. The statistical threshold is set at $p < 0.001$, and results are shown superimposed on the average structural image across subjects. The activation peak is localized to the ventral border of the tegmentum, as illustrated by the black circle, but also extends further down the brainstem into the pons. The coordinates of the peak voxel are [0, −21, −30] ([x, y, z] in MNI space) with the peak $z = 4.21$. The peak voxel remains significant at the $p < 0.001$ level, even after adjusting p values to account for multiple analyses run with different learning rates (adjusted p value of peak voxel = 6.6481×10^{-5}).

(B) To better illustrate the precise localization of the activation peak, a plot of the same activation is shown using a more stringent threshold (set at $p < 0.0001$) overlaid on the mean normalized EPI image (averaged across subjects).

(C) Illustration of VTA activity from three individual subjects overlayed on each subjects' individual mean EPI image. The threshold is set at $p < 0.01$, uncorrected.

(D) Fitted parameter estimates for the temporal difference learning signal are shown from the peak voxel at the group random effects level for each trial type in the order of most to least preferred. Error bars depict standard error of the mean.

(D) Plots of parameter estimates for the temporal difference learning signal from each individual subject in ventral midbrain shown for each of the trial types (ranked in order of preference from most to least preferred). The solid black line depicts a fitted linear regression slope indicating a significant linear trend in the data as a function of preference ($r^2 = 0.242$; $p < 0.001$).

averaged pleasantness ratings for each of the juice stimuli. This analysis yielded a similar pattern of activity in the midbrain to that obtained with preference rankings but at a much lower significance level (which did not meet the criteria for significance at $p < 0.001$). No other areas of interest showed significant effects in this analysis.

Bivalent Responses as a Function of Preference

We also tested for regions that significantly correlated with a bivalent response profile (responding to the cues associated with the most and least preferred compared to the middle preferred stimulus). One region of interest showed a strong bivalent response: the ventral striatum, bilaterally (Figures 5A and 5B). Subject averaged responses in the ventral striatum are shown in Figure 5C. In order to determine whether the striatum was demonstrating a quadratic response as a function of preference or else an absolute valued linear response (i.e., scaling linearly both with increasing preference and also with decreasing preference relative to the middle preferred stimulus), we extracted the parameter estimates from the peak voxel in each individual subject and performed a forward stepwise linear regression procedure, adding linear, quadratic, and absolute valued linear terms to the model in a stepwise fashion. Addition of the quadratic term provided a significantly better fit to the data than the linear term alone ($F[1,62] = 24.3$; $p < 0.001$; $r^2 = 0.278$), indicating that a quadratic response profile was a good description of the data. Further inclusion of the orthogonalized component of the absolute valued regressor (to account for the additional variance explained by the absolute valued response over and above the quadratic response) produced a significantly better fit over the quadratic model alone ($F[1,62] = 11.7$; $p < 0.005$). This result indicates that according to our

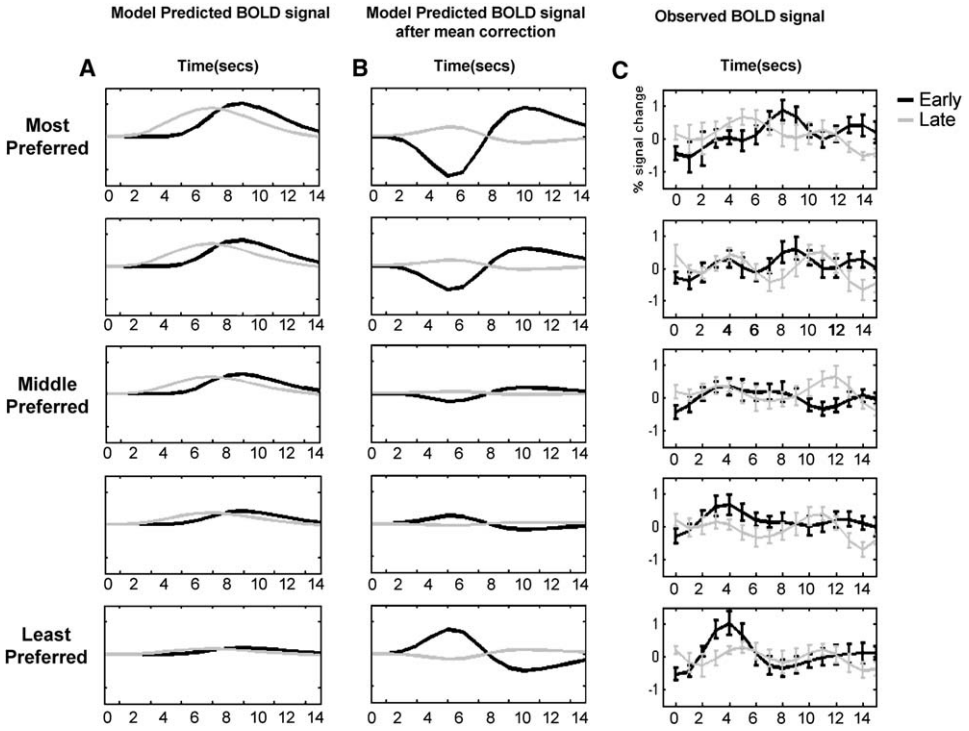

Figure 4. Time Course Plot from Ventral Midbrain

(A) Model-predicted time course plots of the average evoked hemodynamic response during early (trials 1 to 6) and late (trials 24 to 30) trials for each cue-juice pairing (ranked in order from most to least preferred). The model-predicted response of each trial is convolved with a canonical hemoydnamic response function and then averaged across trials. The response for each trial type is shown scaled according to preference (assuming a perfect linear response as a function of preference).

(B) Model-predicted time course plot shown after mean-correction to simulate the fact that due to only a small number of baseline events in the present experiment (1/6 of the total), the evoked hemodynamic response does not return to baseline but instead oscillates around a mean level of activity over all the trials.

(C) Percent signal change subject averaged time course plots shown for each trial type, presented in order of preference of the associated flavor stimulus (from most to least preferred). The time course is shown separately for early (in black; trials 1 to 6) and late trials (in red; trials 24 to 30). The time course is extracted from the peak voxel in ventral midbrain (of the linear preference contrast) for each individual subject and then averaged across subjects. Error bars reflect the standard error across subjects. This observed time course shows some correspondence to the model-predicted time courses for the early and late trials (after mean correction), indicating that the observed fMRI signal reflects a linear change as a function of preference around a mean activity level in this region.

analysis the best description of the response profile in the ventral striatum is that it is demonstrating an absolute linear valued response as a function of preference anchored around the middle preferred stimulus (see Figure 5D). Subject averaged evoked BOLD signals in striatum are plotted in Figure 6C (alongside model-predicted time courses in Figures 6A and 6B), separately for early (first six) and late (last six) trials of each trial type ranked according to preference (from most to least preferred).

Gustatory Responses in Insular Cortex

We also conducted an additional analysis in which we modeled responses at the time of presentation of the flavor stimuli. A comparison of the flavor stimuli compared to the baseline trial (in which no flavor stimulus was presented) revealed activity in primary gustatory cortex (in mid-anterior insula and adjoining frontal operculum; see Figure 7). Responses in this area did not show a significant effect of preference, as tested by a repeated-measures analysis of variance on the parameter estimates across subjects from the peak voxel in this region ($F[4,48] = 0.541$; $p = 0.706$). These results provide support to the suggestion that preference effects observed elsewhere are unrelated to the sensory properties of the stimuli.

Discussion

We show that neural responses to a predictive cue in two key human brain regions, ventral striatum and midbrain, reflect the subjective value of the associated food reward as indexed by behavioral preference. Moreover, such representations develop with learning, responding initially to the food stimulus itself, and then over the course of learning transferring back to the time of presentation of the predictive cue. We suggest that such value-weighted representations may play an important role in guiding action selection when subjects choose between actions that lead to different available awards.

Activity in the ventral mid-brain scaled linearly with behavioral preference. Thus, the greater the activity in this area to a predictive cue, the more the associated food stimulus was preferred. While we also obtained multiple behavioral and physiological measures that discriminated in a relatively crude manner between either

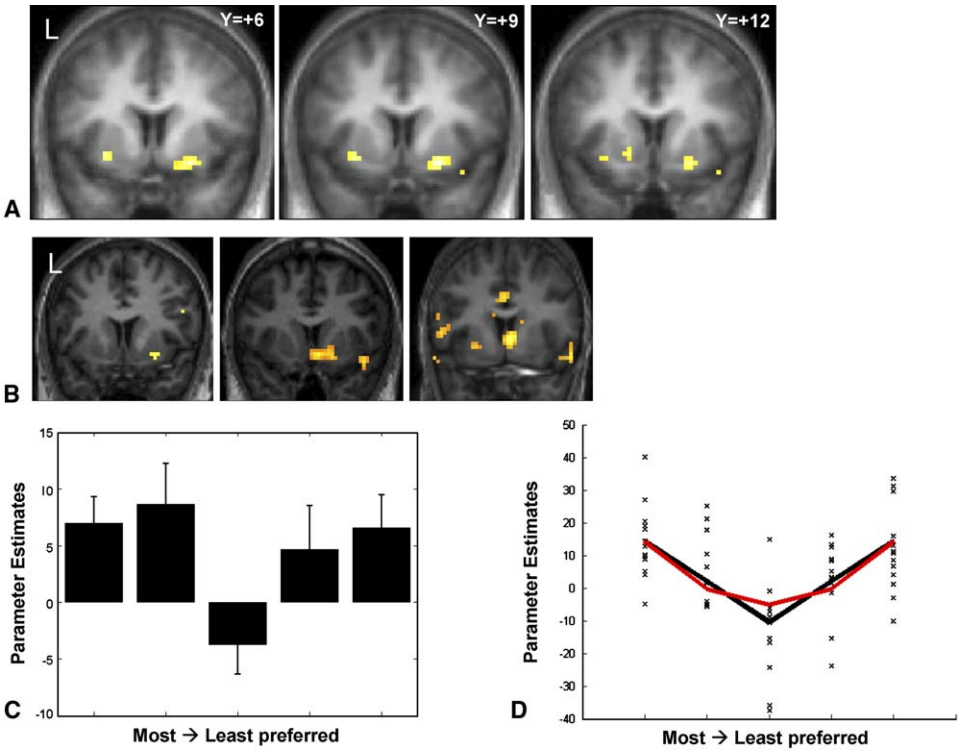

Figure 5. Preference Responses in Ventral Striatum

(A) Predictive responses in ventral striatum as a quadratic function of behavioral preference. The threshold is set at $p < 0.001$, and results are shown superimposed on the average structural image across subjects. Coronal slices are shown through the ventral striatum at the given y coordinates (top left of each slice). The coordinates for the peak voxel in the right striatum are [24, 9, −15] with a peak z value = 4.33; and in the left striatum are [−30, 9, −12], peak z = 3.92. These areas survived small volume correction within two 10 mm spheres defined around coordinates derived from a previous study of temporal difference learning at $p < 0.001$, corrected, in right striatum and at $p < 0.01$, corrected, in left striatum (O'Doherty et al., 2004). The peak voxel remains significant at the $p < 0.001$ level even after adjusting p values to account for multiple analyses run with different learning rates (adjusted p value of peak voxel = 3.6718×10^{-5}).

(B) Illustration of activity in ventral striatum from three individual subjects overlayed on each subject's individual structural image. The threshold is set at $p < 0.01$, uncorrected.

(C) Fitted parameter estimates for the temporal difference learning signal from the group level peak voxel in the right ventral striatum indicating a bivalent response as a function of preference. Error bars depict standard error of the mean.

(D) Parameter estimates for the temporal difference learning signal shown separately from the peak voxels in each individual subject (for the bivalent preference contrast), plotted as a function of preference (from most to least preferred). Stepwise linear regression revealed that a quadratic response profile showed a significant fit to the data (at $p < 0.001$; fitted quadratic response is shown as solid red line). However, significantly more variance was explained by an absolute valued linear response (significant at $p < 0.005$). This response profile is linear for both decreasing and increasing preference anchored around the middle preferred stimulus, and the fit of this response profile to the data is depicted in the figure as a solid black line.

most, least, or middle preferred items, such measures were not as sensitive an index of behavioral preference as our fMRI data. This provides some support to the suggestion that brain imaging could be a more sensitive predictor of subsequent preference behavior than traditional psychophysiological or behavioral assays, at least at the group level (Wilkinson and Halligan, 2004).

In the present study, we do not discriminate between a value signal and a prediction error signal, as the only means to tell these signals apart would be to induce an error in prediction by the omission of expected reward. Thus, ventral midbrain responses may either reflect value or its derivative (prediction error). It should be noted that in previous imaging studies in which errors in reward prediction were induced, prediction error activity was reported in the ventral striatum but not the ventral midbrain, favoring the possibility that ventral midbrain responses relate to value and not prediction error (McClure et al., 2003; O'Doherty et al., 2003, 2004). The finding of a univalent response in the ventral midbrain is consistent with reports that dopamine neurons show enhanced firing for rewards and predictors of reward (Schultz, 1998). However, activity in ventral midbrain may not directly reflect the activity of intrinsic dopamine neurons. Although the link between the blood oxygenation level dependent (BOLD) signal and the underlying neural activity in dopaminergic midbrain remains unexplored, evidence from visual cortex suggests that the BOLD signal is more likely to reflect afferent input into a brain region as well as intrinsic processing within the area (Logothetis et al., 2001). Thus, one possibility is that what is indexed in enhanced BOLD signal in this area is effectively activity within inputs to these dopamine neurons.

Whereas ventral midbrain responses demonstrate a univalent predictive signal with increasing activity to increasing levels of preference, the ventral striatum showed a bivalent signal with a maximal response to

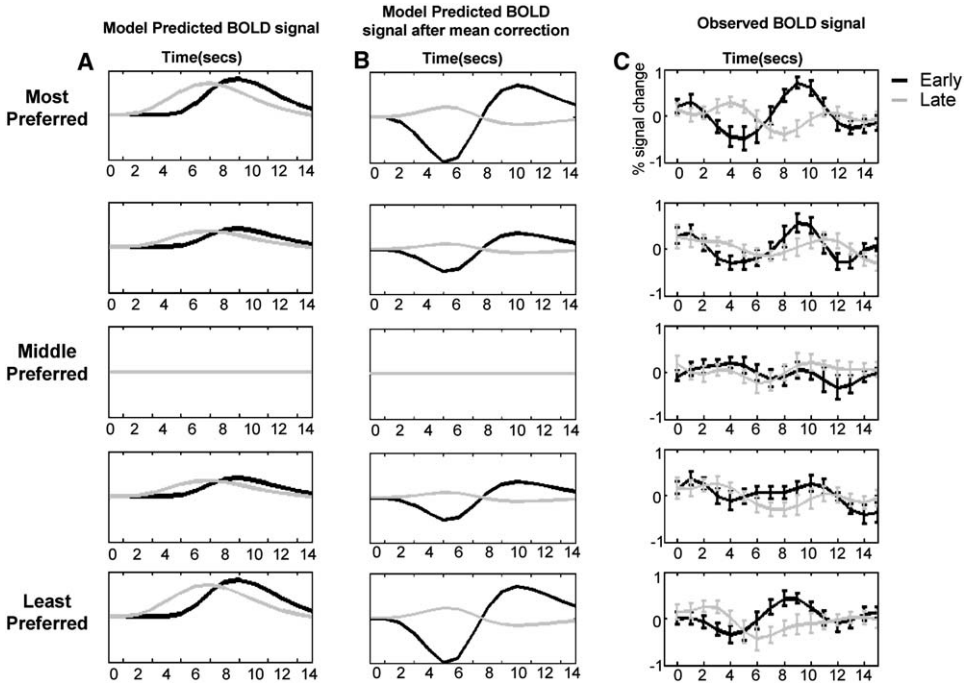

Figure 6. Model-Predicted and Actual Time Course Plots from Ventral Striatum (Ventral Putamen)

(A) Model-predicted time course plots of the average evoked hemodynamic response during early (trials 1 to 6) and late (trials 24 to 30) trials for each cue-juice pairing (ranked in order from most to least preferred). The model-predicted response of each trial is convolved with a canonical hemoydnamic response function and then averaged across trials. The response for each trial type is shown scaled according to preference (assuming a perfect V-shaped response as a function of preference).

(B) Model-predicted time course plot shown after mean-correction to simulate the fact that due to only a small number of baseline events in the present experiment (1/6 of the total), the evoked hemodynamic response does not return to baseline but instead oscillates around a mean level of activity over all the trials.

(C) Percent signal change subject averaged time course plots shown for each trial type (right panel), presented in order of preference of the associated flavor stimulus (from most to least preferred). The time course is shown separately for early (in black; trials 1 to 6) and late trials (in red; trials 24 to 30). The time course is extracted from the peak voxel in ventral striatum (for the bivalent preference contrast) of each individual subject and then averaged across subjects. Error bars reflect the standard error across subjects. A shift in the time to peak of the response is evident as a function of learning (by comparing early trials to late trials) for the most and least preferred trials compared to the middle preferred trials, as reflected in the similarity of the observed time course to the predicted time course after mean correction shown in (B).

least and most preferred stimuli and lowest response to the middle preferred. Previous studies have reported predictive signals in ventral striatum during appetitive learning (to juice and money rewards), as well as during learning with aversive stimuli such as pain (Becerra et al., 2001; Jensen et al., 2003; Knutson et al., 2001; Seymour et al., 2004). Here, we explored responses related to preference rankings for everyday food stimuli, none of which on their own would be considered to be strongly aversive. Nevertheless, we found that ventral striatum responded equally strongly to the predictor of the least preferred food as to the predictor of the most preferred food. One interesting possibility that arises from these findings is that ventral striatum responses may encode the relative value of the available stimuli, rather than coding for their objective value independently of the context in which they are presented. This possibility will need to be tested in a future experiment in which the same reward is presented in different contexts (i.e., alongside different combinations of rewards

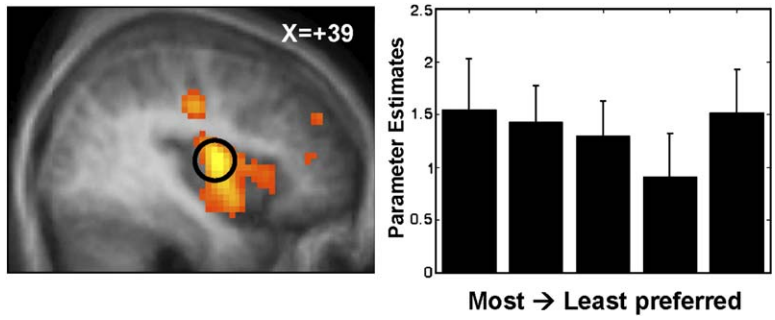

Figure 7. Responses in Primary Gustatory Cortex to Presentation of Flavor Stimuli

(A) Sagittal slice through right insular cortex (at X = +36), demonstrating significant flavor-related activation in the vicinity of primary gustatory cortex, located in the middle/anterior insula and adjacent frontal operculum. The threshold is set at $p < 0.001$, uncorrected.

(B) Parameter estimates extracted from a peak voxel in dorsal mid-insula (at coordinates [36, −3, 6]; peak $z = 6.15$) ranked as a function of preference (from most to least preferred). Error bars depict standard error of the mean. A repeated-measures analysis of variance revealed no significant effect of preference in this region (even at $p < 0.05$).

with variable preference values), in order to establish whether the striatal responses to a reward-predicting cue scales according to relative preference, as is known to be the case in the orbitofrontal cortex (Tremblay and Schultz, 1999). Furthermore, it will also be useful in future experiments to explore whether predictive responses in striatum also scale in a similar manner to cues associated with stimuli that are found to be truly aversive by subjects.

The divergence in response profile between the ventral midbrain and striatum observed here has important implications for understanding the manner in which these structures interact during learning. The reward-prediction error theory of reward learning stipulates that value responses in the striatum and elsewhere are learned via phasic prediction error activity of afferent dopamine neurons. However, the results presented here suggest that there is not a simple linear relationship between the activity in midrain dopaminergic loci and responses in target structures such as the striatum. One possibility is that an opponent signal scaling positively with decreasing preference is also providing input to striatum, leading to the bivalent predictive responses seen here. A candidate opponent signal could be the phasic activity of serotonin neurons, as suggested in a recent theory proposing opponent interactions between serotonin and dopamine (Daw et al., 2002). Such a proposal remains speculative in the absence of direct neurophysiological evidence. Yet, divergent response profiles in these two structures suggests that learning of value representations in striatum may not be mediated exclusively via an afferent dopaminergic signal.

While the imaging results described here are interpreted in the context of Pavlovian conditioning, it is also possible that some of the effects we see pertain to the influence of the Pavlovian cue on the button press (used to indicate whether the stimulus is presented on the left or the right of the screen), an effect known as Pavlovian to instrumental transfer. In this study, the button press was not made contingent on obtaining a reward; nevertheless, it is possible that at least some subjects perceived such a contingency, and in such cases instrumental conditioning mechanisms may have been invoked. However, Pavlovian to instrumental transfer effects are unlikely to account in large part for the significant preference signal reported in the ventral striatum. This is because in a previous paradigm we used a purely passive Pavlovian conditioning task in which juice was delivered following a cue presentation without any requirement of subjects to perform a behavioral response (O'Doherty et al., 2003). Even in this case, we observed significant prediction error effects to reward in the same part of ventral putamen. Thus, it is reasonable to assume that the effects we observe in the present study (at least in the ventral striatum) are mostly due to Pavlovian and not instrumental conditioning effects.

It has long been known that associating brand items with other rewarding or appetitive stimuli, through the process of classical conditioning, makes it possible to modulate subjects' preferences (Gorn, 1982). This process may account in large part for the efficacy and power of advertising. At a broader level, we suggest that our findings provide insight into the neural mechanisms by which such preference signals can be acquired through experience. An obvious extension of our approach and that of McClure et al. (2004) would be to pair arbitrary cue stimuli associated with a given food with other rewarding stimuli (such as attractive faces or pleasant music) and then evaluate the degree to which behavioral preference, and its neuronal correlates, can be experimentally modulated as a function of such associative learning (Cox et al., 2005). The principal implication of the present study is that it provides an account of how predictive representations, learned through classical conditioning, come to elicit activity in the human brain that relate directly to subsequent behavioral preference. We suggest that such representations play an important role in the guidance of action based upon future reward, a form of elementary behavioral decision making.

Experimental Procedures

Subjects

Thirteen healthy right-handed normal subjects were included in the experiment, of which eight were female (mean age, 27.5; range, 21–40). The subjects were preassessed to exclude those with a prior history of neurological or psychiatric illness. All subjects gave informed consent, and the study was approved by the local research ethics committee.

Experimental Protocol

Before scanning, subjects took part in a preference-ranking procedure, in which on each trial the subject was presented with a choice between two of the five juice stimuli and was asked to choose which one they preferred. Each possible combination of stimulus pairs was presented to the subject, and preference rankings were derived. Once preference rankings had been derived, subjects were placed in the scanner and were given 0.7 ml aliquots of each of the five juices in random order and asked to evaluate each juice for its pleasantness, using a scale ranging from −10 to +10, where −10 = very nonpleasant, +10 = very pleasant, and 0 = neutral.

The first of two ~15 min scanning sessions was then initiated. Each session consisted of 90 trials each of 10 s duration. There were six main trial types, each presented 15 times and in random order throughout the session. On each trial, one out of six arbitrary fractal cue stimuli was presented on a gray background to either the left or the right of a central fixation cross. The subjects' task was to respond with a button press as soon as possible after the beginning of the trial to indicate on which side of the fixation cross the stimulus had appeared. After a further 5 s, the next trial was scheduled. For five of the trial types, presentation of a specific cue stimulus was consistently followed 5 s later by intra-oral delivery of 0.7 ml of one of the four different juices or else the tasteless control stimulus. The sixth or baseline trial type involved a cue stimulus that was followed 5 s later by nothing. For all trials, after a further 5 s, the next trial was scheduled. The specific allocation of cue stimuli to a given trial type was counter-balanced across subjects. Once the first session was completed, following a brief break, the second session was initiated, which involved a further 15 repetitions of the same six trial types.

On completion of the experiment, subjects were again asked to provide pleasantness ratings in the scanner for each of the juices. They were then removed from the scanner, and preference rankings were again tested. There were no significant differences in pleasantness ratings from before and after the experiment (interaction term of ANOVA with one factor pleasantness ratings and the other factor session [before an after]: $F[4,9] = 1.5$, $p = 0.27$). Furthermore, Kronbach's α for test-retest reliability of the preference ratings (from before to after) the experiment was 0.90. This indicates highly stable pleasantness and preference ratings for the food stimuli over the course of the experiment.

Flavor Stimulus Presentation

The flavor stimuli were contained in five 50 ml syringes that were attached to an SP220I electronic syringe pump (World Precision

Instruments Ltd, Stevenage, UK), positioned in the scanner control room and delivered to the subjects via five separate 6 meter long 3 mm wide polythene tubes, which were placed into the subject's mouth via a specifically designed five-way disposable mouthpiece (which kept each tube separate but enabled each juice to be delivered centrally in the oral cavity). The syringes were also attached to a computer-controlled valve system that enabled the different tastes to be delivered independently along the tubing. The apparatus was controlled by the stimulus-presentation computer positioned in the control room, which also received volume trigger pulses from the scanner. The visual stimuli were viewed on a projector screen positioned to the rear of the scanner and viewed through a mirror attached to the head coil ~4 cm from the subject's head.

Preference Rankings

Subjects had relatively diverse preference rankings. The blackcurrant juice was the most popular stimulus, ranked as the most preferred by 6 out of 13 subjects, though other subjects ranked melon (four subjects), grapefruit juice (one subject), or carrot juice (one subject) as their most preferred. The grapefruit juice was perhaps the least popular stimulus, ranked by seven subjects as their least preferred stimulus, though other subjects found the tasteless control solution (three subjects), carrot juice (three subjects), or melon juice (one subject) to be their least preferred. Middle ranked stimuli were even more heterogenous, with four subjects rating the tasteless control solution, three subjects rating the blackcurrant juice, three subjects rating the grapefruit juice, and two subjects rating the carrot juice as their middle preferred stimulus. Given the considerable variance in preference rankings between subjects, it is unlikely that neural effects related to preference at the group level can be easily attributed to systematic differences in the sensory (gustatory, texture, or olfactory) properties of the stimuli.

Imaging Procedure

The functional imaging was conducted by using a 3 Tesla Siemens Allegra head-only MRI scanner to acquire gradient echo T2*-weighted echo-planar images (EPI) with BOLD contrast. We employed a special sequence designed to optimize functional sensitivity in OFC and medial temporal lobes (Deichmann et al., 2003). This consisted of tilted acquistion in an oblique orientation at 30* to the AC-PC line, as well as application of a preparation pulse with a duration of 1 ms and amplitude of −2 mT/m in the slice selection direction. The sequence enabled 36 axial slices of 3 mm thickness and 3 mm in-plane resolution to be acquired with a TR of 2.34 s. Coverage was obtained from the base of the orbitofrontal cortex and medial temporal lobes to the superior border of the dorsal anterior cingulate cortex. Subjects were placed in a light head restraint within the scanner to limit head movement during acquisition. A T1-weighted structural image was also acquired for each subject. Functional imaging data were acquired in two separate 390 volume runs. To detect transient head movements due to swallowing, we attached a 1.5 cm long copper coil with a radius of 0.5 cm to the neck of each subject. Small movements of the coil induced a current in the magnetic field that could be detected when amplified using one channel of an EEG system positioned in the scanner room (National Hospital for Neurology and Neurosurgery, London, UK). This produced a time series over the whole experiment reflecting transient head movement.

Temporal Difference Model

The temporal difference (TD) learning model used in this study is that described by Schultz et al. (Schultz et al., 1997). On each trial, the predicted value (V) at any time t within a trial is calculated as a linear product of the weights w_i and the presence or absence of a CS stimulus at time t, coded in the stimulus representation vector $x_i(t)$:

$$\hat{V}(t) = \sum_i w_i x_i(t).$$

Learning occurs by updating the predicted value of each time point t in the trial by comparing the value at time $t + 1$ to that at time t, leading to a prediction error or $\delta(t)$:

$$\delta(t) = r(t) + \gamma \hat{V}(t+1) - \hat{V}(t)$$

where $r(t)$ = reward at time t.

The parameter γ is a discount factor, which determines the extent to which rewards that arrive earlier are more important than rewards that arrive later on. In the present study, we set $\gamma = 0.99$. The weights w_i are then updated on a trial-by-trial basis according to the correlation between prediction error and the stimulus representation:

$$\Delta w_i = \alpha \sum_t x_i(t)\delta(t)$$

where α = learning rate.

We used this algorithm to derive a theoretical prediction error signal to model learning-related changes over the course of the 30 presentations of each trial type (across both sessions 1 and 2). The signal took the form of a phasic response, which over the course of learning shifted its responses from the time of presentation of the reward (5 s into the trial) back to the time of presentation of the cue stimulus (at the beginning of each trial). In this analysis, we used a six time point TD model, in which the time of presentation of the reward was designated to occur at time point 5, and the time of presentation of the cue stimulus at time point 1. We report results using a learning rate (α) of 0.1, which shows strong activation in both ventral midbrain and ventral striatum (the same learning rate was used for all subjects). Using this learning rate, convergence (i.e., complete learning) occurs by the end of the 30 trials. This learning rate is slightly slower than that used in previous studies (where typically $\alpha = 0.2$ was found to be optimal; O'Doherty et al., 2003, 2004). However, this study differs from previous studies in the increased interstimulus interval (5 s instead of 3 s used previously) and the increased number of trial types. This could account for the slower learning rate observed here. For completeness, we tested other learning rates (from 0.2 through to 0.8), but responses were maximal for the learning rate shown.

Image Analysis

Image analysis was performed using SPM2 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). To correct for subject motion, the images were realigned to the first volume, spatially normalized to a standard T2* template with a resampled vowel size of 3 mm^3, and spatial smoothing was applied 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.

For the statistical analysis, each trial was modeled as having five time points: the time of presentation of the cue, three interim time points, and the time of presentation of the reward. The TD prediction error signal (described above) was entered into the general linear model as a parametric regressor to capture changes in neural activity over time as a function of learning for each trial type and for each time point within a trial. The main feature of this signal in the case of the present study was that it captures a shift in the timing of activity within a trial from the time of presentation of the reward itself back to the time of presentation of the cue, over the course of learning. These regressors were then convolved with the canonical hemodynamic response function and correlated with each subjects' fMRI data in SPM.

In addition, the six scan-to-scan motion parameters produced during realignment were included to account for residual effects of scan-to-scan motion. To take into account transient head motion effects produced by, for example, swallowing, we also included an additional motion regressor that featured the output of the motion-detector coil, band-pass filtered appropriately and subsampled to the number of scans in the experiment. Linear contrasts between regressors were computed at the individual subject level to detect regions showing responses to the TD regressor that scaled in a linear or quadratic fashion as a function of preference. To enable inference at the group level, the contrasts from each individual subject were taken to the second level, and random-effects group statistics were computed. A priori we defined the midbrain (in the vicinity of the ventral tegmental area and substantia nigra), ventral striatum, orbitofrontal cortex, and amygdala as areas of interest. By ventral striatum, we refer to the ventral aspects of the striatum incorporating both the nucleus accumbens proper as well as adjacent ventral parts of the putamen. It is important to note that our definition of ventral striatum is more extensive than sometimes used in the literature. Often this term is used to refer exclusively to the part of the striatum

encompassing the nucleus accumbens and olfactory tubercle. Inclusion of the ventral parts of the putamen in the definition used here is motivated by recent findings that the ventral part of putamen has similar cytoarchitectonic characteristics as the nucleus accumbens proper (Holt et al., 1997; Karachi et al., 2002), as well as by the fact that reward-predictive responses have frequently been reported in this area in previous fMRI studies (O'Doherty et al., 2003, 2004; McClure et al., 2003). Results are reported in areas of interest at $p < 0.001$, uncorrected.

The structural T1 images were coregistered 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 (Duvernoy, 1999).

Acknowledgments

This research was supported by a Programme Grant to R.J.D. from the Wellcome Trust. T.W.B. was supported by a short-term fellowship provided by the Human Frontiers Science Program.

Received: April 22, 2005
Revised: August 11, 2005
Accepted: November 2, 2005
Published: January 4, 2006

References

Becerra, L., Breiter, H.C., Wise, R., Gonzalez, R.G., and Borsook, D. (2001). Reward circuitry activation by noxious thermal stimuli. Neuron *32*, 927–946.

Cox, S.M., Andrade, A., and Johnsrude, I.S. (2005). Learning to like: a role for human orbitofrontal cortex in conditioned reward. J. Neurosci. *25*, 2733–2740.

Daw, N.D., Kakade, S., and Dayan, P. (2002). Opponent interactions between serotonin and dopamine. Neural Netw. *15*, 603–616.

Deichmann, R., Gottfried, J.A., Hutton, C., and Turner, R. (2003). Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage *19*, 430–441.

Duvernoy, H.M. (1999). The Human Brain (Vienna: Springer-Verlag).

Gorn, G.J. (1982). The effects of music in advertising on choice behavior: A classical conditioning approach. J. Mark. *46*, 94–101.

Hollerman, J.R., and Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. Nat. Neurosci. *1*, 304–309.

Holt, D.J., Graybiel, A.M., and Saper, C.B. (1997). Neurochemical architecture of the human striatum. J. Comp. Neurol. *384*, 1–25.

Jensen, J., McIntosh, A.R., Crawley, A.P., Mikulis, D.J., Remington, G., and Kapur, S. (2003). Direct activation of the ventral striatum in anticipation of aversive stimuli. Neuron *40*, 1251–1257.

Karachi, C., Francois, C., Parain, K., Bardinet, E., Tande, D., Hirsch, E., and Yelnik, J. (2002). Three-dimensional cartography of functional territories in the human striatopallidal complex by using calbindin immunoreactivity. J. Comp. Neurol. *450*, 122–134.

Knutson, B., Adams, C.M., Fong, G.W., and Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J. Neurosci. *21*, RC159.

Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. Nature *412*, 150–157.

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38*, 339–346.

McClure, S.M., Li, J., Tomlin, D., Cypert, K.S., Montague, L.M., and Montague, P.R. (2004). Neural correlates of behavioral preference for culturally familiar drinks. Neuron *44*, 379–387.

Montague, P.R., Dayan, P., and Sejnowski, T.J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. *16*, 1936–1947.

O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr. Opin. Neurobiol. *14*, 769–776.

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

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

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

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

Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664–667.

Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. Mach. Learn. *3*, 9–44.

Sutton, R.S., and Barto, A.G. (1990). Time derivative models of Pavlovian Reinforcement. In Learning and Computational Neuroscience: Foundations of Adaptive Networks, M. Gabriel, and J. Moore, eds. (Cambridge, MA: MIT Press), pp. 497–537.

Tremblay, L., and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. Nature *398*, 704–708.

Wilkinson, D., and Halligan, P. (2004). The relevance of behavioural measures for functional-imaging studies of cognition. Nat. Rev. Neurosci. *5*, 67–73.