OXFORD

Social Cognitive and Affective Neuroscience, 2023, **18**(1), 1–10

DOI: https://doi.org/10.1093/scan/nsad067
Advance Access Publication Date: 18 November 2023
Original Manuscript

Negative emotion reduces the discriminability of reward outcomes in the ventromedial prefrontal cortex

Lakshman N.C. Chakravarthula and Srikanth Padmala

Centre for Neuroscience, Indian Institute of Science, Bangalore, KA 560012, India
Correspondence should be addressed to Srikanth Padmala, Centre for Neuroscience, Indian Institute of Science (IISc), Bangalore, KA 560012, India.
E-mail: spadmala@iisc.ac.in

Abstract

Reward and emotion are tightly intertwined, so there is a growing interest in mapping their interactions. However, our knowledge of these interactions in the human brain, especially during the consummatory phase of reward is limited. To address this critical gap, we conducted a functional magnetic resonance imaging study to investigate the effects of negative emotion on reward outcome processing. We employed a novel design where emotional valence (negative or neutral) indicated the type of outcome (reward or no-reward) in a choice task. We focused our functional magnetic resonance imaging analysis on the ventro-medial prefrontal cortex (vmPFC), ventral striatum and amygdala, which were frequently implicated in reward outcome processing. In these regions of interest, we performed multi-voxel pattern analysis to specifically probe how negative emotion modulates reward outcome processing. In vmPFC, using decoding analysis, we found evidence consistent with the reduced discriminability of multi-variate activity patterns of reward *vs* no-reward outcomes when signaled by a negative relative to a neutral image, suggesting an emotional modulation of reward processing along the plausible common value/valence dimension. These findings advance our limited understanding of the basic brain mechanisms underlying the influence of negative emotion on consummatory reward processing, with potential implications for mental disorders, particularly anxiety and depression.

Keywords: reward; emotion; valence; ventromedial prefrontal cortex; MVPA

Introduction

The neural substrates underlying reward and emotion have been traditionally investigated in a largely independent manner. However, reward and emotion are tightly intertwined in such a way that the core dimensions of reward (value & salience) and emotion (valence & arousal) have commonalities between them (Cromwell *et al.*, 2020; Sander and Nummenmaa, 2021). Hence, there is a growing interest in mapping the interactions between reward and emotion. Although some previous human neuroimaging work has investigated the brain mechanisms underlying the reward–emotion interactions, one set of those studies focused on the initial choice phase of the decision-making paradigms (Talmi *et al.*, 2009; Park *et al.*, 2011; Aupperle *et al.*, 2015) and the other set of studies have mostly focused on the reward anticipation (Wittmann *et al.*, 2008; Choi *et al.*, 2014; Wei *et al.*, 2016; Padmala *et al.*, 2017; Park *et al.*, 2019). On the other hand, only a limited number of studies have examined the interactions during the consummatory phase of reward (Gorka *et al.*, 2018; Bandyopadhyay *et al.*, 2019; Kim and Anderson, 2020). Hence, very little is known regarding the brain mechanisms underlying the reward outcome–emotion interactions, which is of both basic and clinical relevance (Dillon *et al.*, 2014; Rolls *et al.*, 2020).

Separate lines of human functional magnetic resonance imaging (fMRI) studies focused on reward outcomes and emotion processing has implicated the ventro-medial prefrontal cortex (vmPFC) in the coding of the value and valence dimension, respectively (see Bartra *et al.*, 2013; Lindquist *et al.*, 2016; Oldham *et al.*, 2018, for meta-analytic findings). A few studies that had employed both monetary reward outcomes and emotional stimuli in the same set of participants had reported overlapping activity in the vmPFC (Smith *et al.*, 2010; also see Sescousse *et al.*, 2013, for functional overlap in terms of meta-analytic findings), supporting the notion of common neural currency underlying the encoding of multiple value signals (Levy and Glimcher, 2012). In addition to vmPFC, the ventral striatum is also commonly engaged during receipt of monetary rewards (Bartra *et al.*, 2013; Sescousse *et al.*, 2013; Oldham *et al.*, 2018) and during the processing of emotional stimuli (Sescousse *et al.*, 2010; Smith *et al.*, 2010). On the other hand, the amygdala, a region frequently implicated in encoding the intensity of emotional stimuli in human fMRI studies (Sergerie *et al.*, 2008; Sabatinelli *et al.*, 2011; Lindquist *et al.*, 2016;

Received: 10 June 2023; **Revised:** 11 September 2023; **Accepted:** 17 November 2023

but see Jin *et al.*, 2015) is also engaged during the processing of monetary reward outcomes (Sescousse *et al.*, 2013; Oldham *et al.*, 2018), supposedly reflecting their salient nature (Sescousse *et al.*, 2013). A key question in affective neuroscience beyond the observation of specific brain regions encoding the purported common value/valence or salience/arousal dimensions of reward and emotion is whether and how the neural signatures of reward outcomes and emotion interact along those dimensions. For instance, in regions sensitive to the value/valence dimension, one could expect to find inhibitory-type effects between negative emotion and reward outcomes. In contrast, facilitatory-type effects are expected in regions sensitive to the salience/arousal dimension.

A couple of recent fMRI studies have attempted to address this question by investigating the plausible inhibitory influence of aversive information on reward outcome processing (Gorka *et al.*, 2018; Kim and Anderson, 2020). Gorka *et al.* (2018) employed a modified version of the monetary incentive delay (MID) task (Knutson *et al.*, 2000) to examine how an instructed threat-of-shock context (where aversive shocks were delivered in an unpredictable fashion) compared to a safe context (where shocks were never delivered) influences reward anticipation and outcome processing. Of particular interest to the present study, the authors did not observe that anticipating an aversive shock dampened the reward outcome responses in the vmPFC and ventral striatum. More recently, Kim and Anderson (2020) employed a four-armed bandit task during which participants received one of the four equiprobable outcomes following their choice: monetary reward alone, aversive shock alone, concurrent reward-*plus*-shock and neither of them. Again, of relevance to the present study, attenuation of the reward response in the presence of simultaneous aversive shock was not found in the reward sensitive regions such as the vmPFC and ventral striatum. Overall, the evidence supporting the inhibitory influence of aversive information on reward outcome processing remains elusive.

In the present fMRI study, we sought to further investigate the effects of negative emotion on reward outcome processing by employing a novel experimental design. In addition to ensuring that both the reward and emotion manipulations were concurrent and behaviorally relevant, we made the processing of the former contingent on the latter to achieve their effective integration within the paradigm (see Park *et al.*, 2019 for a similar design in the context of a reward motivation task). Participants performed a simple two-choice task and either received a reward or a no-reward outcome following their response (Figure 1). Specifically, the type of outcome was indicated by the presentation of a negative or neutral emotional scene stimulus in such a way that one of the valence categories indicated reward and the other indicated no-reward. The instructed valence-outcome mapping was reversed during the second half of the experiment, resulting in an orthogonal manipulation with four unique valence-outcome combinations over the entire experiment. We presumed that since our design mandates processing of the emotional valence to determine the type of outcome (reward or no-reward), it would foster the integrated processing of reward and emotional information better than during the concurrent presentation of *separate* reward and aversive stimulus employed previously (Kim and Anderson, 2020).

Using a region of interest (ROI) approach (Poldrack, 2007), we focused our analysis on the vmPFC, ventral striatum and amygdala which were commonly implicated in reward outcome processing (Liu *et al.*, 2011; Bartra *et al.*, 2013; Oldham *et al.*, 2018). In these targeted ROIs, in addition to the conventional univariate analysis, we performed multi-voxel pattern analysis (MVPA; Weaverdyck *et al.*, 2020) to specifically probe how negative emotion modulates reward outcome processing. In several past fMRI studies of reward processing, MVPA techniques have been successful in revealing effects that were undetected in the standard univariate analysis (Kahnt *et al.*, 2010; Vickery *et al.*, 2011; Yan *et al.*, 2016) demonstrating their higher sensitivity

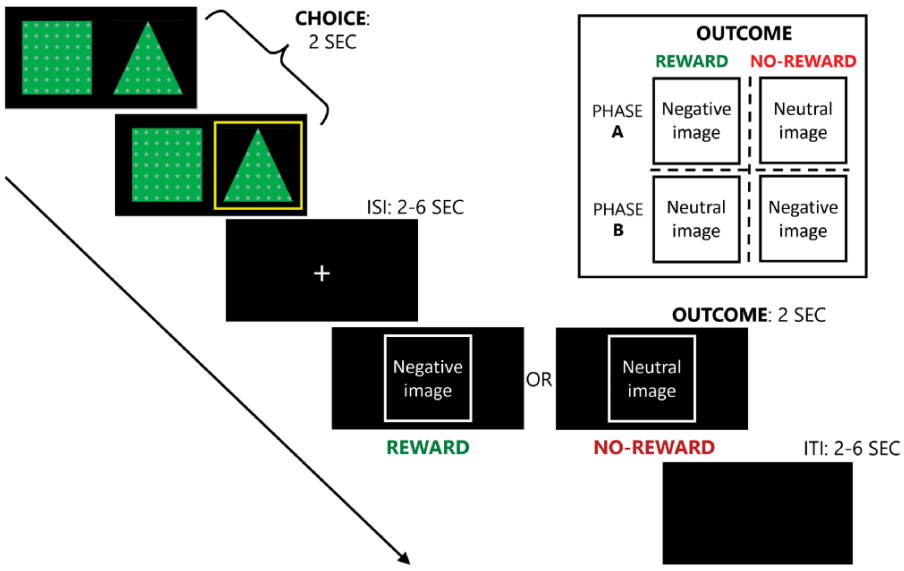

Fig. 1. Experimental design. In each trial, during the initial *Choice* phase, participants were asked to choose between a square or a triangle shape stimulus to maximize their bonus reward. The chosen option was highlighted in yellow. During the subsequent *Outcome* phase, a negative or neutral emotional scene stimulus was presented. The valence of the scene stimulus indicated whether the participant won a reward ('REWARD' outcome) or did not win any reward ('NO-REWARD' outcome). In Phase A, negative stimulus indicated a reward outcome and neutral stimulus indicated a no-reward outcome, and this valence-outcome mapping was reversed in Phase B. The order of Phase A and Phase B was counterbalanced across participants. In each trial, the Choice and Outcome phases were separated by a variable inter-stimulus interval (ISI), where a central white fixation cross was shown. Finally, each trial ended with a variable inter-trial interval (ITI) where the screen remained blank.

(see Kahnt, 2018 for review). We hypothesized that in regions associated with the encoding of common value/valence dimension such as the vmPFC, the multi-variate activity patterns of reward (*vs* no-reward) outcomes would be *less discriminable* when signaled by a negative relative to a neutral scene. In contrast, we expected a opposite pattern in regions that were reported to be sensitive to the arousal/salience dimension such as the amygdala, where multi-variate activity patterns of reward (*vs* no-reward) outcomes were expected to be *more discriminable* when signaled by a negative (*vs* neutral) scene.

Methods

For full information on Methods, please see *Supplementary Material*.

Participants

Thirty-six right-handed healthy adult volunteers (mean age: 24.12 years, SD: 3.39, range: 19–32 years, 21 males) provided informed consent to participate in this fMRI study, approved by the Institutional Human Ethics Committee of the Indian Institute of Science and the Central Ethics Committee of the HCG (HealthCare Global Enterprises Ltd) Hospital, Bangalore.

Paradigm

Task and experimental design

We employed a simple two-choice task for this study. Each trial (Figure 1) started with a *Choice* phase (2 s), followed by a fixation cross that appeared for a jittered inter-stimulus-interval (ISI) of 2–6 s which was followed by the *Outcome* phase (2 s). Finally, a variable inter-trial interval (ITI) of 2–6 s was employed, during which a blank screen was presented. The ISI and ITI values were sampled from an exponential distribution favoring shorter intervals which allowed us to separately estimate the fMRI response during the Choice and Outcome phases (Serences, 2004).

During the Choice phase, the participants chose one of the two shape stimuli (a square and a triangle) presented side-by-side on the screen, and the chosen shape was highlighted following the choice (Figure 1). Participants were instructed that their choice between the two shape stimuli will determine the subsequent monetary bonus reward outcome and were asked to actively make choices to maximize their bonus reward. During the Outcome phase (Figure 1), a neutral or negative scene stimulus was presented on trials where participants chose one of the two shape stimuli during the initial Choice phase. Participants were instructed that the valence of the scene would indicate the type of outcome (reward outcome of Rs. 2 or no-reward outcome of Rs. 0). During one phase of the experiment (phase A), a negative scene indicated that their choice led to a reward outcome and a neutral scene indicated that it led to a no-reward outcome. This instructed mapping between the scene valence and the type of outcome was reversed during the other phase of the experiment (phase B). The order of phase A and phase B was counterbalanced across participants. Overall, our experiment resulted in four unique outcome conditions as part of a 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) within-subjects factorial design. Of note, in our factorial design, when we refer to reward and no-reward outcomes in the 'Negative' condition, it means reward outcomes signaled by a negative image in one phase and no-reward outcomes signaled by a negative image in another phase, respectively. Likewise, when we refer to reward and no-reward outcomes in the 'Neutral' condition, it means reward outcomes signaled by a neutral image in one phase and no-reward outcomes signaled by a neutral image in another phase, respectively.

Unbeknownst to the participants, the reward outcome on every trial was predetermined using a pseudo-randomized order (same for all participants) with the constraint that the same outcome was not repeated for more than three trials in a sequence. This was done to ensure an equal number of trials for the four unique outcome conditions in our factorial design. Although outcomes were predetermined, we included an initial choice phase in our paradigm (see Delgado *et al.*, 2000; Tricomi *et al.*, 2004) for a similar strategy) because previous work has reported that passive delivery of reward outcomes (i.e. without involving any instrumental choice or action context) did not activate regions of our interest specifically the sub-cortical amygdala and ventral striatum (Elliott *et al.*, 2004; Zink *et al.*, 2004).

For each of the four outcome conditions, 36 trials were employed resulting in a total of 144 trials. These trials were divided into four runs (with 36 trials in each run) with two runs in each phase. Each run contained an equal number of predetermined reward and no-reward outcome trials (18 trials of each). At the end of each run, the amount of bonus reward won in that particular run along with the cumulative bonus reward won until then was displayed to the participants. The instruction about the valence-outcome mapping changed between the two phases. Each phase began only after verbal confirmation by the experimenter that the participants clearly understood the instructed valence-outcome mapping. Finally, at the end of each phase, subjective ratings of pleasantness of the experienced reward/no-reward outcomes were collected on a scale of 1–9 units. Example question: '*Please rate how you felt when you received REWARD outcome signaled by a NEGATIVE image*' with labels '*UNPLEASANT*' and '*PLEASANT*' at the extremities of the 1–9 scale.

ROI definitions

To enhance statistical power, we primarily focused on the ROI analysis in the vmPFC, bilateral ventral striatum, and bilateral amygdala that were commonly engaged during the processing of reward outcomes (Bartra *et al.*, 2013; Sescousse *et al.*, 2013; Oldham *et al.*, 2018). The ROIs were selected from a recent meta-analysis of fMRI studies that examined reward (*vs* no-reward) outcome-related processing separately from reward anticipation (Oldham *et al.*, 2018). See Supplementary Material for details.

Univariate analysis

For the individual-level analysis, the representative time-series data in each ROI was modeled using multiple linear regression (using *3dDeconvolve* in AFNI) with the following task-related regressors. During the *Choice* phase, two regressors were included for square and triangle choices separately, and during the *Outcome* phase, four regressors were included corresponding to each outcome type (Neutral-Reward, Neutral-NoReward, Negative-Reward, Negative-NoReward). These six regressors were modeled for 2 s from the corresponding stimulus onset and convolved with a canonical gamma (GAM) variate hemodynamic response function model (Cohen, 1997). The Choice and Outcome phase regressors of no response trials [pooled over all conditions; 0.78 ± 0.9 trials (mean $\pm$ SD)], whenever applicable, were also included in the model as regressors of no interest. Additional regressors included in the GLM model were six estimated motion parameters, their derivatives, and polynomial regressors (of degrees 0–3) separately

for each run to account for baseline and drifts of the MR signal. Finally, we excluded volumes [0.6 ± 1.36% (mean ± SD)] with a frame-to-frame displacement of >1.5 mm Euclidean distance (half of the original voxel size; Siegel *et al.*, 2014) from the GLM analysis (using the *censor* option in *3dDeconvolve*).

For the group-level analysis, our primary focus in this study was on the Outcome phase of the task. Hence, for each ROI, estimated beta coefficients of the four outcome regressors were submitted to a 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA using JASP software (Love *et al.*, 2019). Accounting for multiple comparisons using the Bonferroni correction for the number of ROIs tested, the *P*-value was set at 0.01 for these analyses for an overall α-value of 0.05.

Multi-voxel pattern analysis

Our primary motivation for employing MVPA in this study was to capitalize on its sensitivity to across voxels variability (which is discarded in the univariate analysis) and its lower sensitivity to between-subject variability (which is the primary source of variability in univariate analysis) that may lead to better detection of effects of interest when compared to the standard univariate analysis (Davis *et al.*, 2014). Specifically, we employed decoding analysis using a linear support vector machine classifier in our targeted ROIs (Weaverdyck *et al.*, 2020). We were primarily interested in probing if the ability to classify multi-variate activity patterns of reward *vs* no-reward outcomes differs based on whether negative or neutral scenes signaled those outcomes. A significant difference in classification accuracies between negative and neutral conditions would be indicative of emotional modulation during the reward outcome processing.

Estimation of single-trial activity

Decoding analysis was performed using single-trial activity (i.e. beta) estimates of the four outcome conditions. For each ROI, the single-trial beta estimation was performed on the representative timeseries data in a GLM framework using 3dLSS command in AFNI (See Supplementary Material for details).

Decoding analysis

As noted above, in each of our targeted ROIs, we were primarily interested in probing whether the ability to classify reward vs. no-reward outcomes was modulated by the valence of the emotional scenes that signaled them. At first, decoding analysis was performed in each participant to classify reward *vs* no-reward outcomes *separately* in each valence condition that indicated the outcome (i.e. negative and neutral). The following classification was implemented in the five targeted ROIs (Table 2), using custom MATLAB scripts with functions from The Decoding Toolbox (Hebart *et al.*, 2015). In each ROI, the reward *vs* no-reward outcomes were classified using a linear support vector machine algorithm (default cost parameter C = 1) *separately* in each valence condition. For cross-validation, leave-one-trial-out approach was used where the classifier was trained with multi-variate activity patterns of reward and no-reward outcome trials and classification accuracy was calculated on a single left-out pair of reward and no-reward trials in each fold. In each valence condition, the average classification accuracy across folds was computed and was considered as cross-validation accuracy. Then, in each valence condition, the cross-validation accuracies were averaged across participants to get group-level average classification accuracy. Secondly, concerning our primary research question of interest, cross-validation accuracies were compared between valence conditions. To do so, for each ROI, difference in cross-validation accuracies between negative and neutral conditions was calculated in each participant, and then averaged across participants to get group-level average differential classification accuracy.

In each ROI, for assessing statistical significance of the cross-validation accuracies, we followed the non-parametric permutation-based testing advised in Stelzer *et al.* (2013), which was employed in several recent studies (e.g. Palenciano *et al.*, 2019). See Supplementary Material for details.

Pleasantness ratings analysis

To probe for potential modulatory effect of negative (*vs* neutral) emotion on the subjective ratings of pleasantness, the ratings collected for each of the four outcome conditions were submitted to a 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA using JASP software. The *P*-value was set at 0.05 for this analysis.

Behavioral choice measures

Since participants were instructed that their choice of the shape stimulus determines the reward outcome, the proportion of square and triangle choices was calculated separately during each phase of the main experiment. As noted previously, in one phase, negative and neutral stimuli signaled reward and no-reward outcomes respectively, and in the other phase this valence-outcome mapping was reversed. Additionally, in each of these phases, the proportion of rewarded trials following square and triangle choices was computed to confirm that the choices did not differ in their expected reward values. In case of each metric, the average value across participants was statistically compared against the chance value (50%) using a one-sample *t*-test. The *P*-value was set at 0.05 for this analysis.

Results

fMRI univariate analyses

The central aim of this study was to assess the influence of negative emotion on reward outcome processing. To probe this, in each targeted ROI, we performed a 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA on the beta estimates of the four outcome conditions. As expected, a strong main effect of *Reward Outcome* was detected in all the ROIs, with activity during the reward outcomes being higher than that of the no-reward outcomes (Figure 2, Table 1). A main effect of *Emotion* was observed in the bilateral amygdala such that the activity during the negative scenes was higher than that of the neutral scenes (Figure 2, Table 1). Crucially, the effect of primary interest, the interaction between *Reward Outcome* and *Emotion*, was not detected in any of the ROIs.

For completeness, we also report the 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA results from whole-brain voxel-wise univariate analysis in the Supplementary Material (Figures S2 and S3; Tables S3 and S4). We did not detect any clusters exhibiting *Reward Outcome* × *Emotion* interactions.

fMRI multi-voxel pattern analyses

First, we performed MVPA-decoding analysis to test whether reward *vs* no-reward outcomes can be classified separately under

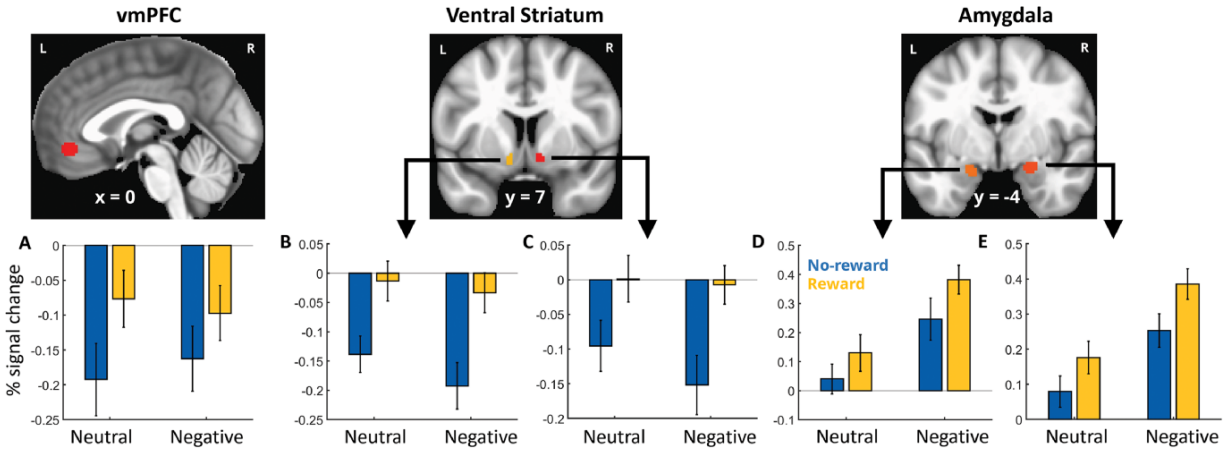

Fig. 2. Univariate fMRI analysis results from the targeted ROIs in (A) ventro-medial prefrontal cortex (vmPFC), (B) left ventral striatum, (C) right ventral striatum, (D) left amygdala and (E) right amygdala. For each ROI, the bar plot shows the average group-level estimates of the four outcome conditions of interest. The error bars represent within-subject error bars based on O'Brien and Cousineau (2014)

Table 1. Univariate ROI analysis results from the 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA

	Main effect: *Reward outcome*		Main effect: *Emotion*		Interaction effect: *Reward outcome* * *Emotion*	
ROI name	*F*(1,35)	*P*-value	*F*(1,35)	*P*-value	*F*(1,35)	*P*-value
Ventromedial PFC	15.362	<0.001	0.064	0.801	1.713	0.199
Left ventral striatum	39.719	<0.001	3.732	0.061	0.074	0.787
Right ventral striatum	30.967	<0.001	3.073	0.088	1.718	0.199
Left amygdala	9.125	0.005	45.629	<0.001	1.367	0.250
Right amygdala	18.744	<0.001	61.12	<0.001	0.458	0.503

Table 2. MVPA ROI-decoding analysis results

		Negative condition		Neutral condition		Negative *minus* Neutral	
S. No.	ROI name	Classification accuracy (Mean ± SE)	*P*-value	Classification accuracy (Mean ± SE)	*P*-value	Classification accuracy difference (Mean ± SE)	*P*-value
1	Ventromedial PFC	53.67 ± 1.28	1.00E-04	58.46 ± 1.46	<0.00001	−4.79 ± 1.96	1.00E-04
2	Left ventral striatum	58.76 ± 1.61	<0.00001	56.17 ± 1.78	<0.00001	2.59 ± 2.09	0.122
3	Right ventral striatum	60.28 ± 1.48	<0.00001	56.04 ± 1.61	<0.00001	4.24 ± 2.06	0.015
4	Left amygdala	56.08 ± 1.50	<0.00001	53.63 ± 1.36	1.00E-04	2.45 ± 2.04	0.086
5	Right amygdala	58.35 ± 1.38	<0.00001	55.51 ± 1.40	<0.00001	2.84 ± 1.59	0.062

the neutral and negative valence conditions. We observed that the classification accuracy was above the chance performance in all five ROIs under negative and neutral conditions (Table 2). Then, we tested whether reward *vs* no-reward outcome classification accuracy differed based on the emotional valence of the image that signaled the outcomes, which is of primary interest to our hypothesis. We observed the following (Figure 3, Table 2): (i) in vmPFC, the reward *vs* no-reward classification accuracy *decreased* under negative compared to the neutral condition; (ii) in the right ventral striatum, the classification accuracy *increased* under negative compared to the neutral condition (a similar pattern was also observed in the left ventral striatum but was not statistically significant) and (iii) in bilateral amygdala, the classification accuracy numerically *increased* under negative compared to the neutral condition but was not statistically significant.

Pleasantness ratings analysis

The 2 *Reward Outcome* (reward, no-reward) × 2 *Emotion* (negative, neutral) repeated-measures ANOVA of subjective ratings revealed a main effect of *Reward Outcome* [$F(1,35) = 35.487$; $P < 0.001$], where higher pleasantness ratings were reported for reward compared to no-reward outcomes (Table 3). Neither the main effect of *Emotion* [$F(1,35) = 0.278$; $P = 0.602$] nor the interaction between *Reward Outcome* and *Emotion* were detected [$F(1,35) = 0.226$; $P = 0.638$].

Behavioral choice measures

Across both phases of the main experiment, participants exhibited a high response rate (mean: 99.22%; SD: 0.9%) indicating that they were attentive to the choice task. In both Phase A and Phase B, the choice proportions of the *square* and the *triangle* stimuli did not

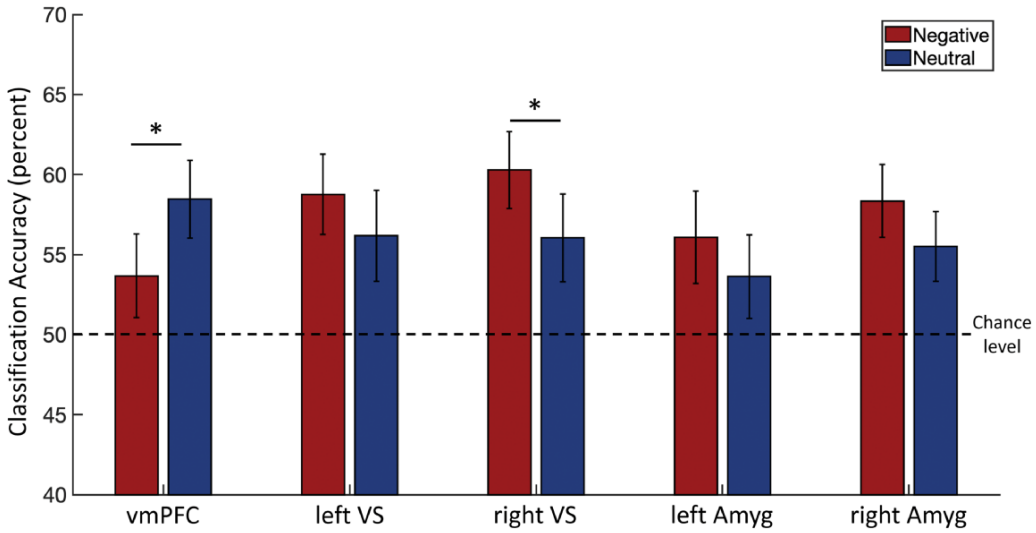

Fig. 3. Reward *vs* no-reward outcome classification accuracies from MVPA-decoding analysis in the targeted ROI. The error bars represent within-subject error bars based on O'Brien and Cousineau (2014). Asterisk indicates a significant difference between the reward *vs* no-reward outcome classification accuracies of negative and neutral emotion conditions assessed using the non-parametric permutation-based testing. vmPFC: ventro-medial prefrontal cortex; VS: ventral striatum; Amyg: amygdala.

Table 3. Pleasantness ratings of the four outcome conditions on a scale of 1–9 with 1 being Unpleasant and 9 being Pleasant

Condition	Mean ± SD
No-Reward Outcome signaled by a *Neutral image*	4.00 ± 1.87
Reward Outcome signaled by *a Neutral image*	6.17 ± 2.09
No-Reward Outcome signaled by a *Negative image*	4.14 ± 1.91
Reward Outcome signaled by a *Negative image*	6.47 ± 1.99

differ from 50% across the participants (Table 4), indicating that the choice pattern was not skewed towards any single stimulus. Additionally, we also computed the proportions of reward outcomes following square and triangle choices in both phases and found them not to significantly differ from the chance level (Table 4). This result confirmed that the reward outcome expectancies for the two choices were not skewed in favor of one over the other.

Discussion

In the present fMRI study, using a novel experimental design where emotional valence indicated the type of outcome, we investigated the influence of negative emotion during reward receipt. We focused our analysis in a set of ROIs involving vmPFC, bilateral ventral striatum, and bilateral amygdala that were frequently implicated in reward outcome processing. Based on the proposals that the usually referred core dimensions of reward and emotion have commonalities between them (Cromwell *et al.*, 2020; Sander and Nummenmaa, 2021), we expected negative emotion would modulate reward outcome signatures along those common dimensions. In vmPFC, using MVPA decoding analysis, we found evidence consistent with the reduced discriminability of reward (*vs* no-reward) outcome signatures when signaled by a negative relative to a neutral image, suggesting an emotional modulation of reward outcome processing along the common value/valence dimension.

Extensive previous work conducted separately in reward and emotion domains has implicated vmPFC in the coding of value and valence dimensions, respectively (Bartra *et al.*, 2013; Lindquist *et al.*, 2016). More importantly, by including both the reward outcome and emotion manipulations in the same set of participants, Smith *et al.* (2010) has reported that fMRI activity in vmPFC represents a common neural currency for the coding of the value/valence dimension (also see Sescousse *et al.*, 2013). Based on the notion of this common value/valence dimension, we hypothesized that the signatures of reward (*vs* no-reward) outcomes would become less discriminable when signaled by negative relative to neutral emotional scenes in vmPFC. In line with our hypothesis, we found that the classification accuracy of decoding reward vs. no-reward outcomes was lower in the negative relative to the neutral condition. Multiple past fMRI studies employing MVPA have reported successful decoding of reward outcomes in the vmPFC (Kahnt *et al.*, 2010; Vickery *et al.*, 2011; Yan *et al.*,

Table 4. Behavioral choice measures during the main experiment. In Phase A, negative stimulus signaled a reward outcome and neutral stimulus signaled a no-reward outcome. In Phase B, neutral stimulus signaled a reward outcome and negative stimulus signaled a no-reward outcome. The *t*(35) and corresponding *P*-values are based on a one sample *t*-test against the chance value (50%)

	Phase A			Phase B		
Condition	Mean ± SD (%)	*t*(35)	*P*-value	Mean ± SD (%)	*t*(35)	*P*-value
Square choices	51.43 ± 9.08	0.94	0.35	50.46 ± 9.73	0.28	0.78
Triangle choices	47.92 ± 8.82	−1.42	0.16	48.61 ± 9.57	−0.87	0.39
Square choices that led to reward outcome	49.01 ± 5.64	−1.06	0.30	50.54 ± 5.65	0.57	0.57
Triangle choices that led to reward outcome	51.00 ± 6.14	0.98	0.33	48.74 ± 6.35	−1.19	0.24

2016). Our findings indicate that such an ability to classify multi-voxel fMRI activity patterns of reward and no-reward outcomes in vmPFC *decreased* when the outcomes were signaled by negative compared to neutral scenes revealing emotional modulation of the reward outcome processing.

In addition to vmPFC, we also investigated the influence of negative emotion on reward outcome processing in the amygdala. Based on the separate set of fMRI findings from emotion and reward domains that support the notion that responses in the amygdala are sensitive to the arousal/salience dimension (Sergerie *et al.*, 2008; Sabatinelli *et al.*, 2011; Sescousse *et al.*, 2013; Lindquist *et al.*, 2016), we hypothesized that signatures of reward (*vs* no-reward) outcome processing would be more discriminable when signaled by negative (*vs* neutral) emotional scenes. In the MVPA analysis, in line with our prediction, classification accuracy of reward *vs* no-reward outcomes was numerically higher in the negative (relative to neutral) condition in the bilateral amygdala, but not statistically significant. Clearly, additional studies are required to further assess the effects of negative emotion on reward outcome processing in the amygdala.

As noted in the 'Introduction' section, some recent attempts at probing for inhibitory influence of aversive stimuli on reward outcome processing yielded null findings (Gorka *et al.*, 2018; Kim and Anderson, 2020). A few observations could be made regarding the design elements of the above studies that might have reduced their ability to detect the hypothesized effects. First, in Gorka *et al.* (2018), although reward outcomes were delivered based on task performance, aversive shock stimulation was delivered *independent* of performance. Hence, the salient nature of performance-contingent reward outcomes (Tricomi *et al.*, 2004; Zink *et al.*, 2004) might have been resistant to any adverse influence of anticipating a behaviorally irrelevant aversive shock. Secondly, in the study by Kim and Anderson (2020), both the reward outcome and aversive shock were delivered concurrently and contingent on the participant's choice making them behaviorally relevant. However, as the authors noted, maybe their concurrent mode of presentation, where they were delivered via two different sensory modalities (visual text information for revealing monetary reward outcome and tactile stimulation for shock), might not have been potent enough to result in inhibitory interactions.

Specifically, compared to Kim and Anderson (2020), we made two major changes in the design and analysis strategy of the present study. First, in terms of experimental design, we ensured tighter coupling between reward and emotion manipulations by making the valence of the emotional scenes indicate the reward/no-reward outcomes rather than the concurrent presentation of the *separate* reward and aversive stimulus. The robust main effect of reward outcome in the univariate analysis (see Table 1) confirms that participants effectively employed the instructed valence-outcome mapping resulting in greater responses during reward compared to no-reward outcomes in all the ROIs that were frequently reported to be engaged during reward receipt (Bartra *et al.*, 2013; Sescousse *et al.*, 2013; Oldham *et al.*, 2018). However, univariate analysis did not reveal interactions between emotion and reward outcome processing in any of the ROIs (Table 1), indicating that the major change made in our experimental design alone was insufficient. Secondly, we employed MVPA, which has higher sensitivity compared to the standard univariate analysis (Davis *et al.*, 2014; Kahnt, 2018). Specifically, in the context of reward outcome processing, several fMRI studies which employed MVPA reported higher sensitivity in detecting the effects of interest (Kahnt *et al.*, 2010; Vickery *et al.*, 2011; Yan *et al.*, 2016). In the present study, unlike the standard univariate analysis, the MVPA-decoding analysis detected the hypothesized modulatory effects of negative emotion on reward outcome processing in the vmPFC. Simulations conducted by Davis and colleagues (Davis *et al.*, 2014) have shown that MVPA analysis is sensitive to the magnitude of voxel-level variability in the effect of interest (i.e. reward *vs* no-reward outcomes effect in the present study) within subjects, even when the sign of the effect remains the same in each voxel. Moreover, MVPA is insensitive to subject-level variability in mean activation across an ROI, which is the primary source of variance in the standard univariate analysis. Thus, it is possible for a true but weak effect to remain undetected in a univariate analysis but could be picked up in an MVPA analysis specifically in cases of large across-subject variability (Davis *et al.*, 2014). Overall, our novel experimental design combined with MVPA uncovered the influence of negative emotion on reward outcomes.

Although the observed emotional modulation of reward outcome processing in the vmPFC was attributed to interaction along a common value/valence dimension, vmPFC was also implicated in processing pleasure that might have elicited during reward receipt (Kühn and Gallinat, 2012; Berridge and Kringelbach, 2015). In the subjective ratings of pleasantness collected in the present study, we found the main effect of reward outcome where participants reported higher pleasantness during the experience of reward (*vs* no-reward) outcomes. However, of primary interest, interaction effects were not detected, suggesting a lack of emotional modulation of pleasure associated with processing reward outcomes. Of note, because of time constraints in the MRI setup, we had collected only one pleasantness rating for each of the four valence-outcome conditions. Hence, the null interaction effects observed with these single rating scores should be interpreted with caution. Future studies may collect pleasantness ratings of experienced outcomes on each trial (see Sescousse *et al.*, 2010; Buchel *et al.*, 2018), and correlate those ratings with vmPFC responses to further characterize the nature of the observed interaction.

One limitation of the present study that merits discussion is the inability of our experimental design to tease apart the processing of experienced reward value from prediction errors that signal the discrepancy between received and expected reward outcomes. Given that our task did not involve any orthogonal manipulation of reward predictability following the Choice phase, positive and negative reward prediction errors of equal magnitude were expected to be generated during reward and no-reward outcomes, respectively. This would result in a high correlation between the prediction error signals and outcome value signals in our design. Hence, the current study cannot provide unequivocal evidence for the emotional modulation of reward value *vs* prediction error signals in our regions of interest. However, some past fMRI studies that employed ingenious experimental paradigms to tease apart the value from prediction error signals have implicated vmPFC in the coding of experienced reward value, whereas the ventral striatum is believed to be involved in encoding reward prediction errors (Pagnoni *et al.*, 2002; Hare *et al.*, 2008; Rutledge *et al.*, 2010; Rohe *et al.*, 2012). In the MVPA analysis, we found that the classification accuracy of decoding reward *vs* no-reward outcomes was higher under negative compared to neutral valence condition in the right ventral striatum. We speculate that negative emotion might have modulated the coding of reward prediction errors in the (right) ventral striatum. The enhanced decoding accuracy in the negative

condition could also be attributed to the possibility of a non-specific increase in overall arousal and attention in the presence of salient negative (relative to neutral) images, which is supported by some previous work that implicated ventral striatum in saliency processing (Zink *et al.*, 2004; Jensen and Walter, 2014; but see Sabatinelli *et al.*, 2007). Nevertheless, we expect future studies to include explicit manipulations of reward predictability to delineate how emotional information modulates prediction error and other value/salience related signals in the ventral striatum (see Watanabe *et al.*, 2013).

One potential caveat while interpreting the null interaction effects in our standard univariate ROI analysis is worth mentioning. As noted earlier, we observed a robust main effect of reward in all our ROIs (and also in the retrospective subjective ratings data), attesting to the effectiveness of our reward manipulation. However, we did not detect a main effect of negative (*vs* neutral) valence in the univariate ROI analysis and subjective ratings. The lack of evidence for our emotional manipulation suggests that the brief presentation of a negative emotional image might not have been sufficient to transiently induce negative affect to interfere with the reward outcome processing, contributing to the weaker interaction effects. Hence, more potent affect manipulations employing other intense aversive stimuli or negative mood induction could allow one to detect the hypothesized interaction effects even in the standard univariate analysis. Nevertheless, the strong effect of reward (*vs* no-reward) outcomes independent of whether a negative or neutral valenced image signaled them suggests that the neurocircuitry of consummatory reward processing is reasonably robust to the particular visual features and connotations of the stimulus signaling the outcome.

In the present study, we employed negative stimuli to specifically examine the emotional modulation of reward outcome signatures when signaled by valence-incompatible images. Future work could complement these findings by investigating how signaling outcomes by valence-compatible images (i.e. positive stimuli) would influence the reward outcome signatures in vmPFC and other brain regions. Although some previous fMRI studies have included both monetary rewards and erotic images as outcomes in their task paradigms (Sescousse *et al.*, 2010; Buchel *et al.*, 2018), they were presented on separate trials leaving the question of how positive emotion modulates reward outcome signatures unexplored. Of note, one previous ERP study reported that concurrent presentation of an incidental happy (relative to neutral) emotional facial stimulus enhanced the coding of monetary loss (*vs* gain) outcomes as reflected in the feedback-related negativity (FRN) component (Bandyopadhyay *et al.*, 2019).

In conclusion, using a novel experimental design combined with MVPA, our study demonstrated the emotional modulation of reward outcome processing in the human brain. Specifically, in vmPFC, decoding analysis revealed reduced discriminability of reward (*vs* no-reward) outcome signatures when signaled by a negative relative to a neutral emotional stimulus. Extending the notion of the commonality between the core dimensions of reward and emotion (Cromwell *et al.*, 2020; Sander and Nummenmaa, 2021), the MVPA results in vmPFC elucidated how negative emotion modulates reward outcome processing along the common value/valence dimension. These findings advance our limited understanding of the basic brain mechanisms underlying the influence of negative emotion during the consummatory phase of reward processing, with potential implications for mental illnesses such as anxiety and depression (Dillon *et al.*, 2014; Rolls *et al.*, 2020).

Supplementary data

Supplementary data is available at *SCAN* online.

Data Availability Statement

The data and codes supporting the findings of this study will be made available by the authors upon reasonable request.

Author contributions

L.C. and S.P. designed the research, interpreted the results and wrote the manuscript. L.C. performed the data collection and analysis. Both authors approved the final version of the article.

Conflict of interest

The authors declared that they had no conflict of interest with respect to their authorship or the publication of this article.

Acknowledgements

The support for this work was provided in part by a Cognitive Science Research Initiative grant from the Department of Science and Technology, Government of India (DST/CSRI/2021/87 to S.P.). We are grateful to Prof. Luiz Pessoa and Dr Nicola Sambuco for their valuable feedback on the earlier versions of the manuscript. We thank two anonymous reviewers for their constructive feedback on the previous version of the manuscript.

References

Aupperle, R.L., Melrose, A.J., Francisco, A., Paulus, M.P., Stein, M.B. (2015). Neural substrates of approach-avoidance conflict decision-making. *Human Brain Mapping*, **36**(2), 449–62.

Bandyopadhyay, D., Pammi, V.C., Srinivasan, N. (2019). Incidental positive emotion modulates neural response to outcome valence in a monetarily rewarded gambling task. *Progress in Brain Research*, **247**, 219–51.

Bartra, O., McGuire, J.T., Kable, J.W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, **76**, 412–27.

Berridge, K.C., Kringelbach, M.L. (2015). Pleasure systems in the brain. *Neuron*, **86**(3), 646–64.

Buchel, C., Miedl, S., Sprenger, C. (2018). Hedonic processing in humans is mediated by an opioidergic mechanism in a mesocorticolimbic system. *Elife*, **7**, e39648.

Choi, J.M., Padmala, S., Spechler, P., Pessoa, L. (2014). Pervasive competition between threat and reward in the brain. *Social Cognitive and Affective Neuroscience*, **9**(6), 737–50.

Cohen, M.S. (1997). Parametric analysis of fMRI data using linear systems methods. *NeuroImage*, **6**(2), 93–103.

Cromwell, H.C., Abe, N., Barrett, K.C., *et al.* (2020). Mapping the interconnected neural systems underlying motivation and emotion: a key step toward understanding the human affectome. *Neuroscience and Biobehavioral Reviews*, **113**, 204–26.

Davis, T., LaRocque, K.F., Mumford, J.A., Norman, K.A., Wagner, A.D., Poldrack, R.A. (2014). What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. *NeuroImage*, **97**, 271–83.

Delgado, M.R., Nystrom, L.E., Fissell, C., Noll, D., Fiez, J.A. (2000). Tracking the hemodynamic responses to reward and punishment in the striatum. *Journal of Neurophysiology*, **84**(6), 3072–7.

Dillon, D.G., Rosso, I.M., Pechtel, P., Killgore, W.D., Rauch, S.L., Pizzagalli, D.A. (2014). Peril and pleasure: an RDOC-inspired examination of threat responses and reward processing in anxiety and depression. *Depression and Anxiety*, **31**(3), 233–49.

Elliott, R., Newman, J.L., Longe, O.A., Deakin, J.W. (2004). Instrumental responding for rewards is associated with enhanced neuronal response in subcortical reward systems. *NeuroImage*, **21**(3), 984–90.

Gorka, A.X., Fuchs, B., Grillon, C., Ernst, M. (2018). Impact of induced anxiety on neural responses to monetary incentives. *Social Cognitive and Affective Neuroscience*, **13**(11), 1111–9.

Hare, T.A., O'doherty, J., Camerer, C.F., Schultz, W., Rangel, A. (2008). Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *Journal of Neuroscience*, **28**(22), 5623–30.

Hebart, M.N., Görgen, K., Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, **8**, 88.

Jensen, J., Walter, H. (2014). Incentive motivational salience and the human brain. *Restorative Neurology and Neuroscience*, **32**(1), 141–7.

Jin, J., Zelano, C., Gottfried, J.A., Mohanty, A. (2015). Human amygdala represents the complete spectrum of subjective valence. *Journal of Neuroscience*, **35**(45), 15145–56.

Kahnt, T., Heinzle, J., Park, S.Q., Haynes, J.-D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, **107**(13), 6010–5.

Kahnt, T. (2018). A decade of decoding reward-related fMRI signals and where we go from here. *NeuroImage*, **180**, 324–33.

Kim, A.J., Anderson, B.A. (2020). The effect of concurrent reward on aversive information processing in the brain. *NeuroImage*, **217**, 116890.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *NeuroImage*, **12**(1), 20–7.

Kühn, S., Gallinat, J. (2012). The neural correlates of subjective pleasantness. *NeuroImage*, **61**(1), 289–94.

Levy, D.J., Glimcher, P.W. (2012). The root of all value: a neural common currency for choice. *Current Opinion in Neurobiology*, **22**(6), 1027–38.

Lindquist, K.A., Satpute, A.B., Wager, T.D., Weber, J., Barrett, L.F. (2016). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex*, **26**(5), 1910–22.

Liu, X., Hairston, J., Schrier, M., Fan, J. (2011). Common and distinct networks underlying reward valence and processing stages: a meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, **35**(5), 1219–36.

Love, J., Selker, R., Marsman, M., *et al.* (2019). JASP: graphical statistical software for common statistical designs. *Journal of Statistical Software*, **88**, 1–17.

O'Brien, F., Cousineau, D. (2014). Representing error bars in within-subject designs in typical software packages. *The Quantitative Methods for Psychology*, **10**(1), 56–67.

Oldham, S., Murawski, C., Fornito, A., Youssef, G., Yücel, M., Lorenzetti, V. (2018). The anticipation and outcome phases of reward and loss processing: a neuroimaging meta-analysis of the monetary incentive delay task. *Human Brain Mapping*, **39**(8), 3398–418.

Padmala, S., Sirbu, M., Pessoa, L. (2017). Potential reward reduces the adverse impact of negative distractor stimuli. *Social Cognitive and Affective Neuroscience*, **12**(9), 1402–13.

Pagnoni, G., Zink, C.F., Montague, P.R., Berns, G.S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, **5**(2), 97–8.

Palenciano, A.F., González-García, C., Arco, J.E., Pessoa, L., Ruz, M. (2019). Representational organization of novel task sets during proactive encoding. *Journal of Neuroscience*, **39**(42), 8386–97.

Park, H.R., Kostandyan, M., Boehler, C.N., Krebs, R.M. (2019). Winning smiles: signalling reward by overlapping and non-overlapping emotional valence differentially affects performance and neural activity. *Neuropsychologia*, **122**, 28–37.

Park, S.Q., Kahnt, T., Rieskamp, J., Heekeren, H.R. (2011). Neurobiology of value integration: when value impacts valuation. *Journal of Neuroscience*, **31**(25), 9307–14.

Poldrack, R.A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, **2**(1), 67–70.

Rohe, T., Weber, B., Fliessbach, K. (2012). Dissociation of BOLD responses to reward prediction errors and reward receipt by a model comparison. *European Journal of Neuroscience*, **36**(3), 2376–82.

Rolls, E.T., Cheng, W., Feng, J. (2020). The orbitofrontal cortex: reward, emotion and depression. *Brain Communications*, **2**(2), fcaa196.

Rutledge, R.B., Dean, M., Caplin, A., Glimcher, P.W. (2010). Testing the reward prediction error hypothesis with an axiomatic model. *Journal of Neuroscience*, **30**(40), 13525–36.

Sabatinelli, D., Bradley, M.M., Lang, P.J., Costa, V.D., Versace, F. (2007). Pleasure rather than salience activates human nucleus accumbens and medial prefrontal cortex. *Journal of Neurophysiology*, **98**(3), 1374–9.

Sabatinelli, D., Fortune, E.E., Li, Q., *et al.* (2011). Emotional perception: meta-analyses of face and natural scene processing. *NeuroImage*, **54**(3), 2524–33.

Sander, D., Nummenmaa, L. (2021). Reward and emotion: an affective neuroscience approach. *Current Opinion in Behavioral Sciences*, **39**, 161–7.

Serences, J.T. (2004). A comparison of methods for characterizing the event-related BOLD timeseries in rapid fMRI. *NeuroImage*, **21**(4), 1690–700.

Sergerie, K., Chochol, C., Armony, J.L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, **32**(4), 811–30.

Sescousse, G., Redouté, J., Dreher, J.-C. (2010). The architecture of reward value coding in the human orbitofrontal cortex. *Journal of Neuroscience*, **30**(39), 13095–104.

Sescousse, G., Caldú, X., Segura, B., Dreher, J.-C. (2013). Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, **37**(4), 681–96.

Siegel, J.S., Power, J.D., Dubis, J.W., *et al.* (2014). Statistical improvements in functional magnetic resonance imaging analyses produced by censoring high-motion data points. *Human Brain Mapping*, **35**(5), 1981–96.

Smith, D.V., Hayden, B.Y., Truong, T.-K., Song, A.W., Platt, M.L., Huettel, S.A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *Journal of Neuroscience*, **30**(7), 2490–5.

Stelzer, J., Chen, Y., Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage*, **65**, 69–82.

Talmi, D., Dayan, P., Kiebel, S.J., Frith, C.D., Dolan, R.J. (2009). How humans integrate the prospects of pain and reward during choice. *Journal of Neuroscience*, **29**(46), 14617–26.

Tricomi, E.M., Delgado, M.R., Fiez, J.A. (2004). Modulation of caudate activity by action contingency. *Neuron*, **41**(2), 281–92.

Vickery, T.J., Chun, M.M., Lee, D. (2011). Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron*, **72**(1), 166–77.

Watanabe, N., Sakagami, M., Haruno, M. (2013). Reward prediction error signal enhanced by striatum–amygdala interaction explains the acceleration of probabilistic reward learning by emotion. *Journal of Neuroscience*, **33**(10), 4487–93.

Weaverdyck, M.E., Lieberman, M.D., Parkinson, C. (2020). Tools of the trade multivoxel pattern analysis in fMRI: a practical introduction for social and affective neuroscientists. *Social Cognitive and Affective Neuroscience*, **15**(4), 487–509.

Wei, P., Wang, D., Ji, L. (2016). Reward expectation regulates brain responses to task-relevant and task-irrelevant emotional words: ERP evidence. *Social Cognitive and Affective Neuroscience*, **11**(2), 191–203.

Wittmann, B.C., Schiltz, K., Boehler, C.N., Düzel, E. (2008). Mesolimbic interaction of emotional valence and reward improves memory formation. *Neuropsychologia*, **46**(4), 1000–8.

Yan, C., Su, L., Wang, Y., *et al.* (2016). Multivariate neural representations of value during reward anticipation and consummation in the human orbitofrontal cortex. *Scientific Reports*, **6**(1), 1–12.

Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., Chappelow, J.C., Berns, G.S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, **42**(3), 509–17.