Reward activates stimulus-specific and task-dependent representations in visual association cortices.

Anne-Marike Schiffer [1], Timothy Muller [1], Nick Yeung[1] and Florian Waszak [2,3]

[1]Department of Experimental Psychology, University of Oxford, OX13UD, UK
[2]Université Paris Descartes, Sorbonne Paris Cité, Paris, France
[3]Laboratoire Psychologie de la Perception, UMR 8158, Paris, France

Number of pages: 28 incl Title & References
Number of figures: 4
Number of words:
- Abstract: 250
- Introduction: 520
- Discussion: 1498

Short Title:
Stimulus-specific Task-dependent Reward Activation

Acknowledgements:
This work is supported by the Biotechnology and Biological Sciences Research Council (BBSRC) grant number "BB/I019847/1", awarded to NY and FW.

Conflict of Interest
The authors declare no competing financial interests.

Abstract

Humans reliably learn which actions lead to rewards and implement these actions instrumentally. One prominent question is how credit is assigned to the environmental stimuli that were acted upon. Recent functional magnetic resonance imaging (fMRI) studies have provided preliminary evidence that representations of rewarded stimuli are activated at the time of reward delivery, providing possible eligibility traces for credit assignment. The present study sought new evidence of post-reward activation in sensory cortex that satisfied two crucial conditions of learning: that post-reward activity should reflect the category of the stimulus that preceded reward (stimulus specificity), and should occur only if the stimulus was acted on to obtain reward (task dependency). The novel design implemented two different tasks in the fMRI scanner. The first task was a perceptual decision making task on degraded face and house stimuli. Stimulus-specificity was evident as rewards activated the visual association cortices associated with face vs. house perception more strongly after face vs. house decisions. The second task required participants to make an instructed response. The criterion of task-dependency was fulfilled as rewards after face vs. house responses activated the respective association cortices to a larger degree when faces and houses were not only present, but also relevant to the performed task. Because all major analyses concerned trials that unbeknownst to the participants used pure noise images, our study is the first to show these criteria of eligibility traces in credit assignment, and reveal their independence from bottom-up activation of sensory cortices.

1 Introduction

Humans and other animals learn how to act on environmental stimuli to gain reward. Substantial research effort has been devoted to understanding the neural and computational mechanisms by which reward delivery fosters associative learning (Rescorla and Wagner, 1972; Schultz, 2007). This research has revealed that reward-driven learning depends crucially on midbrain dopamine neurons, which display a firing pattern that bears striking resemblance to reward prediction error signals in formal models of reinforcement learning (Schultz, Dayan, and Montague, 1997; Waelti, Dickinson, and Schultz, 2001).

Despite this progress, fundamental questions on reward-based learning remain unanswered. Whereas formal approaches provide computational solutions to the critical problem of *credit assignment* - determining which features are predictive of positive outcomes – little is known about how such eligibility traces (Sutton and Barto, 1990), are represented in the brain. In this study, we aimed to identify neural signatures of eligibility traces with two crucial properties: Eligibility traces in reward-driven learning should be *stimulus specific* and *task dependent*, to guarantee precision of ensuing reward predictions. Stimulus specificity does so by ensuring that the specific environmental conditions that preceded reward will trigger its prediction. Task dependency warrants that environmental conditions are only associated with reward if they were used to perform the rewarded action.

A handful of studies have recently investigated related questions, focussing on the hypothesis that learning should depend on activation of stimulus representations at the time of reward delivery (Arsenault et al., 2013; FitzGerald, Friston, and Dolan, 2012; Pleger et al., 2008; Pleger et al., 2009; Weil et al., 2010). Whereas fMRI in animals has revealed evidence of reward-related ac-

tivity that is stimulus-specific (Arsenault et al., 2013), corresponding evidence in human studies has not been consistently observed (FitzGerald, Friston, and Dolan, 2012; Weil et al., 2010) and whether or not reward-based activity in sensory cortex is stimulus specific remains unclear.

We conceived a new paradigm to investigate this question, in which subjects performed a perceptual discrimination task, deciding whether degraded stimuli contained images of faces or houses. Importantly, the experiment included trials on which, unbeknownst to participants, the stimulus was pure noise. This design renders activation by reward independent of the initial bottom-up activation, independent of potential category specific reward expectations and minimizes the possible contribution of neural adaptation effects (FitzGerald, Friston, and Dolan, 2012; Grill-Spector, Henson, and Martin, 2006).

Of critical interest was whether we could find evidence of stimulus-specific cortical activity (activity in the stimulus-specific ROIs: FFA and PPA) at the time of reward delivery, our first criterion for a neural signature of an eligibility trace. Our second criterion for this neural signature is task dependency. Post reward activation should be stronger for stimuli that were used to gain reward. We compared post-reward activation in the perceptual decision task and a second, instructed response task, hypothesizing that stimulus activity at reward outcomes would be restricted to trials in which the outcomes were experienced as a consequence of an earlier perceptual decision.

In summary, we predicted that activity in the ROIs would show a positive correlation with reward size for associated decisions (demonstrating stimulus-specificity), and would be more influenced by reward following a perceptual decision than following an instructed response (demonstrating task dependency).

[Figure 1 to be inserted here]

2 Materials and Methods

18 right-handed, healthy participants (10 women, age 20 –32 years; mean age 24 years) took part in the study. The participants reported no psychiatric or neurological past or present condition. All procedures were approved by the local ethics committee of the University of Oxford and all participants gave written informed consent.

2.1 Stimulus material

Grayscale photographs of faces and front views of houses (Figure 1), served as stimulus material. In a first step, all images were adjusted in luminance and spatial frequency to the mean of the stimulus pool using the SHINE (Willenbockel et al., 2010) Matlabtool. This measure was taken to prevent categorisation based on surface similarities (Rajimehr et al., 2011; Schyns and Oliva, 1994). A variable percentage of all phases in each Fourier transformed image was scrambled. Images were then back transformed into native space. Three degrees of phase scrambling were applied to yield stimuli to produce the easy, medium or hard recognition trial. Face images were phase scrambled to 70 %, 75 %, and 85% percent. House images were phase scrambled to 50 %, 65 %, or 75 %. These degrees of scrambling were chosen based on pilot testing to produce comparable performance for house and face stimuli across the three levels of degradation. In addition to the three difficulty levels, half of the images were pure noise images with 100 % of all phases scrambled.

2.2 Task

In each trial, participants were first presented with one of the stimuli for 2 seconds. Stimulus presentation was followed by a task image, displayed on the screen for 1.5 seconds; the task image was either a question mark, or an exclamation mark. Question marks instructed participants to press the left or right button to indicate whether they had seen a face or a house (perceptual decision task). Participants were unaware of the fact that half of the images were noise images and instructed to always decide and respond. Images of exclamation marks contained a darkened box on the left or right side underneath the exclamation mark. In these trials, participants had to press the button on the side that corresponded to the box (instructed response task). Importantly, at the time of stimulus presentation, participants did not know what trial type they were in and had to make a perceptual decision regardless. Participants had 1.5 seconds to respond, after which the task image stayed on the screen for the remaining RSI. A cross in a box to the left or right of the task image was displayed during this interval, indicating their previous response. The length of this RSI was randomly drawn from a Poisson distribution with lambda 4, minimum 2, maximum 6, and jittered in steps of 500 ms. The ITI was followed by feedback; which could be rewarding, neutral, or penalizing.

Rewards consisted of images of either one or two moneybags, resulting in the gain of 10 or 20 points, respectively. Penalties were shown as one or two bombs, indicating the loss of either 10 or 20 points. Participants were told that rewards and penalties were contingent on the correctness of their previous response, but were in fact randomly assigned in noise trials. Participants received feedback on their accumulated score each 50 trials. Their final score was converted into a monetary bonus of up to £5 after the scan. The experimental sequence in the

scanner consisted of 275 trials, 138 of which were noise trials; 35 noise trials and 33 signal trials were indication trials; 24-26 noise question trials were followed by a neutral outcome, the remaining noise trials were followed in equal numbers by large rewards, small rewards, small penalties and large penalties. Outcomes in the signal trials were performance contingent, but outcome size was randomly determined. Just prior to the scanning session, participants performed 16 practice trials of the experimental paradigm.

The experimental task was followed by a functional localiser task to determine the ROIs for all planned contrasts. Participants performed an 1-back task while they were presented with two instances of six blocks of 18 images that appeared on the screen for 150 ms, followed by an ISI of 400 ms. Participants had a short break between the first six and second six blocks. They had to switch from making responses with one hand to the other after the break. Each block contained images of only one category; these categories were: unscrambled face images, unscrambled house images, easy medium, and hard face images, easy, medium, and hard house images, pure noise images and unscrambled object images.

2.3 Behavioural analysis

Behaviour in the task was recorded to establish that participants showed performance modulation by the degree of phase scrambling of the signal stimuli and to assess learning and win stay lose shift optimised as markers of learning. It was also established whether participants made face as well as house judgments in the noise trials.

2.4 FMRI procedure

The functional imaging session took place in a 3T Siemens Magnetom Trio scanner (Siemens, Erlangen, Germany). During the scan, participants lay supine on the scanner bed with their left and right index finger resting on two buttons of a centrally placed response box. Participants wore sound attenuating headphones that allowed communication with the experimenter. They viewed the stimuli on the screen via a mirror built into the head coil. Stimuli were displayed at 5 degrees of visual angle to prevent head and eye movements. The functional session engaged a single-shot gradient echo-planar imaging (EPI) sequence sensitive to blood oxygen level dependent contrast (32 slices, parallel to the bicommisural plane, echo time 30ms, flip angle 90°; repetition time 2000ms; interleaved recording). After the functional session was completed, high-resolution 3D T–1 weighted whole-brain MDEFT sequences were recorded for every participant (128 slices, field of view 256mm, 256 by 256 pixel matrix, thickness 1mm, spacing 0.25 mm).

2.5 FMRI data analysis

FMRI data analysis was conducted with the LIPSIA processing tool (Lohmann et al., 2001). For spatial registration, EPI data and 3D MDEFT data were first oriented along the ac-pc axis. The matching parameters (6 degrees of freedom, 3 rotational, 3 translational) of the functional data onto the individual 3D MDEFT reference set were used to calculate the transformation matrices for linear registration. These matrices were subsequently normalized to Talairach

brain size (x=135mm, y=175mm, z=120mm; (Talairach and Tournoux, 1988)) by linear scaling. The normalized transformation matrices were then applied to the functional slices, to transform them using trilinear interpolation and align them with the 3D reference set in the stereotactic coordinate system. The generated output had a spatial resolution of 3 x 3 x 3 mm. Cubic-spline interpolation was used to correct for the temporal offset between the slices acquired in one scan. To remove low-frequency signal changes and baseline drifts, a high pass filter of 1/75 Hz was applied for event related analysis and a high pass filter of 1/125 Hz was applied to the analysis of the localiser blocks. Statistical evaluation was based on a least-square estimation using the general linear model (GLM) for serially auto-correlated observations (Worsley and Friston, 1995). Temporal Gaussian smoothing (4 seconds FWHM) was applied to deal with temporal autocorrelation and determine the degrees of freedom (Worsley and Friston, 1995). A spatial Gaussian filter of FWHM 5.65 mm was applied. Unless otherwise stated, the design matrix was generated by hemodynamic modelling using a γ -function in all contrasts. The onset vectors in the design matrices were modelled in a time-locked event-related fashion. No first derivatives were encompassed in the model, except for the functional localiser.

2.5.1 ROI definition

Functional regions of interest (ROIs) were determined in a two-step approach. As a first step, all blocks from the functional localiser that contained house images, all blocks that contained face images and the object image blocks were entered separately as regressors into a GLM (1). Events were modelled with a box-car function and event length set to block length. House blocks were contrasted with face blocks (HouselocaliserBlock > FacelocaliserBlock) and vice

versa (FacelocaliserBlock > HouselocaliserBlock) on the single subject level and averaged into t-map contrast images. In a separate analysis, the face and house signal trials from the main experiment were entered separately into one GLM (2). Event amplitude was determined by signal strength, with amplitude increasing from 1 to 3 for hard to easy trials. The regressor accounting for house trials was then contrasted with the face trial regressor. (HouseSignalStrength > FaceSignalStrength) and vice versa (FaceSignalStrength > HouseSignalStrength). The resulting contrast images were masked with the contrast images generated based on the functional localiser. The masked images then entered second-level random effects analysis. One-sample t-tests were employed for the group analyses across the contrast images of all subjects that indicate whether observed differences between conditions were significantly different from zero. The bilateral peak voxels of activity in the parahippocampal gyry were used as centres for the parahippocampal place area (PPA) ROI. ROIs were established as 2 x 2 x 2 voxel cubes centred on the bilateral peak coordinates (coordinates). Peak voxels for the fusiform face area (FFA) ROI were generated in a parallel approach, locating peak voxels in the fusiform gyrus; the bilateral ROI was set as a cube of 2 x 2 x 2 voxels around these centres.

2.5.2 Decision-specific activation at noise stimulus presentation

To show that noise stimuli were treated as if they contained signal, the first contrast tested whether the ROIs would show significant activation in line with the perceptual decision on noise trials. Two regressors were entered into the GLM, one accounting for the presentation of noise stimuli that would be followed by a house decision and the corresponding regressor for noise stimuli that

were followed by a face decision. Events were time-locked to noise stimulus presentation, modelled with an event length of 1 and event amplitude of 1. We estimated the main effect of each regressor separately and contrasted face noise trials with house noise trials (FaceDecision > HouseDecision). The mean beta scores extracted from the FFA and PPA ROIs entered a repeated measures ANOVA to estimate main effects and interactions of decision (face or house) and ROI (FFA or PPA).

2.5.3 Reward network response

The second contrast aimed to show that reward after noise trials would result in the network response associated with learning from rewards. We parametrically modelled BOLD increase from neutral to large reward trials after both face and house decisions to noise stimuli. Events were modelled time-locked to reward presentation, modelled with an event length of 1 and event amplitude ranged from 1 (neutral) to 3 (large reward).

2.5.4 Stimulus-specific activation at reward outcome

The most critical features for post-reward activation to be classified as an eligibility trace were stimulus specificity and task dependency. These effects were tested in a GLM which contained the following 8 regressors: 4 regressors separately accounting for parametric modulation of reward size after house responses in noise decision trials, face responses in noise decision trials, instructed 'house'

responses, and instructed 'face' responses. Large rewards were modelled with an amplitude vector of 2, small rewards with an amplitude vector of 1. Events were modelled time-locked to reward presentation. Further, this GLM contained separate regressors for penalty outcomes modelled by size after face decisions and house decisions and separate regressors for neutral outcomes after each type of decision. The main parametric contrasts for stimulus specificity were (FaceDecisionReward) and (HouseDecisionReward). The stimated beta values for post reward activity scaling with reward size in the FFA and PPA for face vs. house decisions respectively were entered into a 2 x 2 ANOVA to test for stimulus-specific post reward activation.

2.5.5 Task-dependent activation at reward outcome

The second main contrast tested for the assumption that reactivation should be task dependent, i.e. depend on a perceptual decision, as opposed to an instructed response. To test this hypothesis, we performed a repeated measures ANOVA on the beta values from the regressors FaceDecisionReward, FaceInstructionReward, HouseDecisionReward, and HouseInstructionReward. Further, we directly contrasted [(FaceDecisionReward) > (FaceInstructionReward)] in the FFA ROI. Events were modelled time-locked to reward presentation and reward size was modelled parametrically with the same amplitude vector as in the stimulus-specificity analysis.

2.5.6 Trial-type sensitivity

For the last contrast, a separate GLM was defined to compare post reward activation for different trial types. To test whether noise stimuli would deliver a more sensitive context for post reward activation effects than signal stimuli, we included the following four main regressors to assess stimulus specificity and task dependency: reward size parameter after house responses in noise decision trials, face responses in noise decision trials, instructed 'house' responses in noise trials, and instructed 'face' responses in noise trials. To allow comparison, the GLM further included the corresponding regressors for signal trials. Mean betas from all parametric contrasts were entered into a 2 x 2 x 2 x 2 repeated measures ANOVA for further analysis.

T-test were performed on beta values from the contrast of regressors in the respective GLMs. For whole-brain analysis, acquired t-values were transformed to z-scores. To correct for false-positive results, an initial z-threshold was set to 2.56 (p <0.05, one-tailed, uncorrected for multiple comparisons). The results were corrected for multiple comparisons at the cluster level, using cluster size and cluster value thresholds that were obtained by Monte-Carlo simulations. The employed significance level was p= 0.05. The reported activations are significantly activated at $p <= 0.05$, corrected for comparison at cluster level.

3 Results

3.1 Behavioural analysis

Analysis of participants' performance indicated that they engaged with the task, and confirmed that the paradigm effectively created three different levels of difficulty, with performance in the hardest level of difficulty being close to the implemented chance performance in noise trials. Participants made on average 75.8 % correct responses (SD= 12.9 %) on signal trials. One dataset was excluded from the analysis because performance was below 2 standard deviations from the mean. The remaining 17 participants achieved on average: 88.9 %, 79.9 %, and 63.3 % on easy, medium, and hard face signal trials and 92.1 %, 79.4 %, and 58.4 % on the corresponding house signal trials. In instructed response trials, participants reached on average 89.27 % correct responses (SD = 9.67 %). To test that while signal trials created a plausible context for noise trials, they were not clearly distinguishable from noise trials, we assessed how many levels of degradation participants thought they had encountered. Of 17 participants, 9 indicated that the experiment implemented 3 levels of difficulties, while 4 participants believed that there had been "3–4" levels of difficulty. Only 2 participants correctly estimated that there had been 4 levels of difficulty, while the remaining participants indicated 5, and 50 levels of degradation, respectively. It thus appears as if for most subjects, noise trials were not clearly distinguishable from signal trials. Incidentally, only one participant reported he had noticed that a few trials did not contain signal. This participant did not realise that half of the trials were noise trials.

In a next step, we assessed whether participants made use of feedback to adapt their behaviour. We therefore assessed participants' performance changes

over the course of the experiment. As expected, performance improved, as revealed in a 2 x 3 repeated measures ANOVA with the factors TIME (level: first half of experiment /second half of experiment) and DEGRADATION (level: easy/medium/hard). This analysis revealed a marginally significant main effect of time ($F(1,15) = 4.41$, $p = 0.051$), a significant main effect of degradation ($F(2,30) = 42.85$, $p = 0.000$) and a significant interaction between the two main factors ($F(2,30) = 4.23$, $p = 0.032$). Descriptively, participants' performance improved particularly on hard trials (Figure 2a). Participants thus seem to learn from feedback integrating this information to modify their behaviour.

Because feedback was assigned randomly in noise trials (and learning was impossible), performance modification by feedback was assessed on a trial-by-trial basis instead. To this end, we assessed for successive noise trials how likely participants were to repeat a rewarded response or switch away from a penalized one. Such a win-stay/lose-shift behaviour would indicate pseudo-learning from positive feedback. A one sample t-test revealed significant difference in stay probabilities for rewarded compared to penalized noise trials ($t(16) = 17.46$, $p = 0.000$), with participants being more likely to repeat a rewarded response. We thus find evidence, both in signal, as well as in noise trials that participants pay attention to feedback and adapt their performance accordingly.

As a last measure of the credibility of the manipulation, but also the comparability of BOLD effects, we compared the distribution of perceptual judgments on noise trials: Participants showed balanced judgements, with no strong preferences on the group level. Thus, face decisions were on average made on 49% of all trials ($SD = 7.8$ standardised%, range = 31 - 61 %).

3.2 FMRI analysis

The FFA ROI for this and all further analyses was derived from masking the (FaceSignalStrength > HouseSignalStrength) contrast with the (FacelocaliserBlock > HouselocaliserBlock) contrast and was centred on the peak coordinates x = –38, y = –51, z = –15 and x = 34, y = –60, z = –15. The PPA ROI for this and all further analyses was derived from masking of the (HouseSignalStrength > FaceSignalStrength) contrast with the (HouselocaliserBlock > FacelocaliserBlock) contrast was centred on the peak coordinates in x = –26 , y = –41 , z = –6 and x = 25, y = –44 , z = –6 (Figure 3).

To determine whether participants treated noise stimuli as if they contained signal, we estimated the BOLD activity in the region of interests (ROIs), at the time when noise stimuli were presented, in relation to the subsequent perceptual judgment. Activity in these stimulus-specific ROIs provided clear evidence that noise stimuli were treated as if they contained some (albeit weak) signal. Participants' individual beta values for the two conditions: viewing noise stimuli that were then judged to be faces (FaceDecision) and viewing noise stimuli that were judged to be houses (HouseDecision) were estimated in the two ROIs (Figure 3). These individual beta values were then entered into a repeated measures ANOVA with the factors DECISION (face/house) and ROI (FFA/PPA). This yielded no significant effect of decision, but a significant main effect of ROI (F(1,16) = 5.93, p = 0.027) and a significant interaction between DECISION and ROI (F(1,16) = 38.33, p = 0.000). The significant interaction is further illustrated by the direct contrasts of conditions within the ROIs. These contrasts showed significantly more activity in the FFA for pending face vs. house judgments (t(16) = 2.25, p = 0.019) and significantly more activity in the PPA than FFA preceding house vs. face judgments (t(16) = 3.39, p = 0.002). (Figure

3b).

We further established that positive outcomes to noise trials would activate the network of brain regions classically associated with learning from reward (O'Doherty et al., 2003; O'Doherty, 2004). Corrected for multiple comparisons at the whole brain level (cluster threshold $z = 2.56$), the according parametric contrast established the hypothesized positive correlation of BOLD signal with reward size in the network classically signifying reward, that is the right nucleus accumbens and right subgenual anterior cingulate gyrus/vmPFC. The network further included bilateral hippocampal activation which has also been repeatedly been associated with learning from rewarding outcomes (Figure 3 c).

[Figure 3 to be placed here]

3.2.1 Stimulus-specific activation at reward outcome

The primary aim of the present study was to identify proposed neural correlates of eligibility traces supporting reinforcement learning, which we hypothesize to be reflected in post-reward activation in ROIs that represent stimulus categories. This stimulus specificity was defined as the first criterion to make post reward activation a plausible correlate of credit assignment. The stimulus-specificity effect was assessed in two separate contrasts that modelled the parametric effect of reward size. The two parameters modelled BOLD activity increase in the ROIs separately for reward after face and house decisions, respectively. A repeated measures ANOVA on the mean beta values from the parametric contrasts with the factors ROI (PPA /FFA) and RESPONSE (house /face) yielded

a significant main effect of ROI (F = 5.104, p = 0.038), no significant main effect of RESPONSE ($F(1,16) = 1.59$, $p = 0.22$), and a statistically significant interaction ($F(1,16) = 8.98$, $p = 0.009$), in line with the hypothesis of stimulus-specific post-reward ROI activity (Figure 4). To investigate the degree to which both areas contributed to this overall effect, we conducted one sample t-tests on group averaged beta values from the parametric contrast to test for deviations from zero (FaceDecisionReward), which yielded a significant result in the FFA ROI ($t(16) = 2.45$, $p = 0.013$), but no significant result in the PPA ROI. The corresponding analysis of post-reward activity after house decisions parameter did not reveal significant activation in either ROI. Thus, visually identical reward images (pictures of money bags) activated stimulus representations differentially in a decision-contingent manner, an effect mostly carried by an increase in FFA activity following reward stimuli to face decisions compared to reward stimuli following house decisions (Figure 4).

3.2.2 Task-dependent activation at reward outcome

Our second criterion for post-reward activation to be a marker of credit assignment was task dependency. Task dependency requires post-reward activation to be specific to perceptual decision tasks. We modelled four separate parametric contrasts: reward size for face decisions, house decisions, instructed face responses and instructed house responses. Entering the beta values from the RESPONSE (face/house), ROI (FFA/PPA), and TASK (decision/instructed) conditions from this parametric analysis into a repeated measures 2 x 2 x 2 ANOVA revealed no significant main effects, but a marginally significant interaction for RESPONSE x TASK (F = 3.125, p = 0.096) and the hypothesized

significant interaction of ROI x TASK x RESPONSE ($F(1,16) = 10.84$, $p = 0.005$). This significant three-way interaction is indicative of a stronger positive relationship between reward size and response-specific ROI activity in the perceptual decision task than in the instructed response task (Figure 4), satisfying the task-dependency criterion. Because activity in the FFA ROI was modulated to a higher degree by stimulus specificity, we assessed within the FFA whether this stimulus-specific activity was also task-dependent. Therefore, face decisions were contrasted with trials in which participants made an instructed response with the same key. This contrast [(FaceDecisionReward) > (FaceInstruction-Reward)] yielded a significant result in the FFA ROI ($t(16) = 2.02$ $p = 0.03$), showing that the significant effect of stimulus specificity in the FFA was task-dependent.

3.2.3 Trial-type sensitivity

The present paradigm focussed on establishing stimulus specificity and task dependency by focussing on trials with noise stimuli, using stimuli with true (house or face) signal primarily to create a credible context for those critical noise trials within our perceptual judgment task. It is nevertheless instructive to analyse reward-induced activity following the signal trials for comparison with other recent studies of reward-related activation in sensory cortex. Signal trials differ from noise trials in several notable respects, for example because these perceptual decisions would involve more reliance on bottom-up features and because reward probability and neural adaptation effects are confounded with signal strength.

To compare the two types of trials, we modelled reward parameters for noise

as well as signal trials in each task, implementing eight separate regressors for the factorial combination of TRIAL-TYPE (noise /signal), RESPONSE (house /face), and TASK (decision /instructed). As a first pass, we established whether the documented stimulus specificity and task dependency effects were again reliable in noise trials in a 2 x 2 x 2 (ROI x RESPONSE x TASK) repeated measures ANOVA. This yielded a marginally significant interaction of ROI and TASK ($F = 3.09$, $p = 0.098$) and a significant 3-way interaction between ROI, RESPONSE and TASK ($F(1,16) = 10.19$, $p = 0.006$), with ROI activity being higher for the associated response in decision tasks. Thus we find support for the result from the original stimulus specificity and task dependency analyses in this alternative GLM.

Given this confirmation, we next compared post-reward BOLD responses for the two different trial-types in a 2 x 2 x 2 x 2 repeated measures ANOVA including the factor ROI (FFA/PPA) crossed with RESPONSE, TASK, and TRIAL-TYPE. This analysis replicated the stimulus-specificity effect in a marginally significant interaction of ROI x RESPONSE ($F(1,16) = 3.47$ $p = 0.081$). Further, it showed a significant 4-way interaction ($F(1,16) = 18.11$, $p = 0.001$). The significant 4-way interaction indicates trial-type sensitivity of the established effects, expressed as a difference between noise and signal trials with regard to task dependency. Because the differential activation of the ROIs was repeatedly shown to be decision dependent, and not pronounced for indication trials, we focussed our further comparison of noise and signal trials on decision trials only. We therefore investigated the effect of trial-type sensitivity for decision trials in a 2 x 2 x 2 repeated measures ANOVA with the factors: ROI (FFA/ PPA), RESPONSE (face/house), and TRIAL-TYPE (noise/signal) and found a significant interaction of ROI and RESPONSE ($F(1,16) = 7.32$, $p = 0.016$) indicating stimulus specificity in decision tasks, a significant interaction of ROI

and TRIAL-TYPE ($F(1,16) = 10.13$, $p = 0.006$), and a significant 3-way interaction ($F(1,16) = 7.45$, $p = 0.015$). Thus, we find confirmation of the stimulus specificity and task dependency effect, but a difference between noise and signal trials.

[Figure 4 to be inserted here]

4 Discussion

To interact successfully with their environment, humans must learn how to acquire rewarding outcomes. The neural basis of reward processing has been studied extensively. However, we still know very little about how reward-yielding tasks are represented in the brain. One possibility would be that eligibility traces for credit assignment become apparent as activation of sensory cortices representing components of the rewarded task. The present fMRI study investigated the neural correlates of post-reward task representations in visual association cortices in a perceptual decision making task. We defined two criteria for BOLD signal increase to be a potential correlate of eligibility traces: activation in a sensory association cortex should be stimulus-specific, i.e. reflect the stimulus category of the rewarded response, and task-dependent, i.e. should only occur if the stimulus was relevant to the task. In line with our hypotheses, we found the representation of a stimulus to be activated post reward, especially if it was relevant for the correct response in the rewarded task. This effect was established in a significant interaction of response category (face or house decision) and BOLD activity in the FFA or PPA. Moreover, the effect was specific to trials in which the stimulus category was task relevant, fulfilling the second

criterion of task dependency. These effects were particularly pronounced in the FFA.

4.1 Credit assignment

Reinforcement learning (RL) theory provides a solution to the credit assignment problem, as it explains how events that predict reward are assigned a higher value and become targets of behaviour (Sutton, 1988; Sutton and Barto, 1990; Dayan and Niv, 2008; Daw and Doya, 2006). However, the neural underpinnings of this mechanism are still unclear. In particular, it is known that reward prediction and prediction errors elicit neural activity in the basal ganglia and vmPFC and result in dopamine release in the midbrain, but it is yet to be established how this reward response fosters the representations of rewarded tasks. One proposal is that reward signal increases synaptic plasticity in sensory areas (Jay, 2003; Pennartz et al., 2011; Lisman, Grace, and Duzel, 2011). Support evidence comes from studies that have shown modulation of neural activity according to anticipated reward (value) (Brosch, Selezneva, and Scheich, 2011; Serences, 2008). These findings indicate that pairing with reward changes the neuronal representation of an event's value in sensory cortices; this effect may explain why associations between stimuli and reward can prime behaviour (Wimmer and Shohamy, 2012; Hickey, Chelazzi, and Theeuwes, 2010; Hickey and van Zoest, 2012). However, while the data show that learning results in value coding in sensory areas, they do not explain how the association between the reward and the representation is shaped during learning. That is, these studies show that credit assignment takes place and may be linked to dopamine, but they do not target the question how neural activity representing relevant stimuli is linked to neural correlates of reward during learning.

The present study sheds light on this question as it demonstrates the stimulus-specificity and task-dependency of post reward activation.

4.2 Stimulus specificity

If post-reward activation of the task underlies learning, an exact representation of the previous stimulus should be traceable after reward delivery. This notion has been tested in a number of studies (FitzGerald, Friston, and Dolan, 2012; Pleger et al., 2008; Pleger et al., 2009; Weil et al., 2010). However, results of paradigms studying reward-related activation of visual association areas have been ambiguous. Using high-resolution fMRI in monkeys, Arsenault and colleagues (2013) successfully showed post-reward activation in sensory cortices, but only in trials which did not entail the visual stimulus itself. Conversely, several human fMRI studies have failed to find evidence of stimulus-specific activity in sensory cortex following reward delivery (FitzGerald, Friston, and Dolan, 2012; Weil et al., 2010).

In the present study, we measured effects following decisions on noise stimuli that contained no objective signal. This may have rendered the design especially sensitive to stimulus specific activation for a number of reasons. First, adaptation to individual stimuli has been suggested to explain the observation of non-specific activation effects (FitzGerald, Friston, and Dolan, 2012). The present study analysed the post-reward activation following perceptual decisions on noise. Activity in the respective ROIs was thus dependent on the judged category representations, not on lower level features of the individual stimuli. Category representation might be less prone to sensory adaptation than lower level features. Second, decision under noise and top-down driven post-reward activation rely on feedback projections which differ from feedforward projections

conveying sensory input. They might therefore activate the same level in the cortical hierarchy of stimulus representation, as well as specifically the same cortical layer (Markov et al., 2013) within a given area. This may increase the overlap of the locus of BOLD response measured for decision under noise and top-down driven activation, increasing positive correlation between decision-specific activity and post reward stimulus-specific activity. Third, in studies with true 'signal' stimuli, strong anticipation of reward may modulate activity in the sensory cortices prior to reward delivery (Brosch, Selezneva, and Scheich, 2011; Serences, 2008). Thus, reward delivery may have had little effect on activity given that it was delivered in a performance-dependent manner in tasks where participants performed above chance (FitzGerald, Friston, and Dolan, 2012; Pleger et al., 2008; Weil et al., 2010). Here, however, rewards in noise trials could not be anticipated, as feedback was assigned randomly, limiting a positive correlation between signal strength, reward anticipation and reward. Collectively, these crucial features of our noise stimuli may have made the present study more sensitive to stimulus-specific post-reward effects compared to previous studies.

One aspect of the current findings is that effects are reliable when investigated as interactions of decisions and ROIs, but only within the FFA when the ROIs are analysed separately. Previous evidence of decision-dependent effects in the FFA but not the PPA (Summerfield et al., 2006) has been attributed to several possible factors. First participants may rule out faces before making house decisions (Summerfield et al., 2006). If house responses were results of guessing when no decision in favour of faces can be made, lower confidence in this judgment could result in less credit assignment. Alternatively, PPA activity may rely on more specific bottom-up features and be less modulated by top-down input than the FFA. While the PPA was modulated at the time point

of decision making in the present study, in contrast to the results reported by Summerfield and colleagues (2006), house decision activity in the PPA led to slightly lower parameter estimates than face decision activation in the FFA. This may suggests that top-down activation is weaker, or target different subregions of the parahippocampal gyrus, than bottom-up driven activation.

4.3 Task dependency

Most objects require a specific manipulation to yield the desired outcomes. However, at any given moment a large number of objects is present in our environment, and a given object may afford different actions depending on the task context. As a consequence, reward needs to activate the specific representations of only those objects that were involved in the current task. Global post reward activation including irrelevant stimulus representations would yield a new credit assignment problem (Roelfsema and van Ooyen, 2005).

This implies that if post-reward activation observed in the current study was a marker of a credit assignment, we would expect it to reflect whether the stimulus that preceded the reward was relevant to the rewarded task. The established post-reward activation in previous studies was tested in setups where stimuli that preceded reward were always task relevant, rendering it difficult to interpret the established effects as correlates of either credit assignment or less specific reward-driven activation. In the present study, of two different tasks, only one required a response that was stimulus specific. We showed that stimulus-specific post-reward activation was dependent on the relevance of the stimulus for the reward yielding task. Thus, our study shows that reward selectively increases activity in sensory areas representing objects that have been used to perform a task, and not globally in sensory areas representing any object

currently present within the context of the task.

This novel finding relates to the recently developed attention-gated reinforcement learning model (Roelfsema, van Ooyen, and Watanabe, 2010; Roelfsema and van Ooyen, 2005). This model suggests that sensory cortices become fine-tuned towards relevant (reward-predicting) features of a stimulus, while distracter features are suppressed. It assumes that a global RL signal increases synaptic plasticity, while an attentional gating mechanism ensures that this increased synaptic plasticity is limited to task-relevant dimensions (Roelfsema, van Ooyen, and Watanabe, 2010; Roelfsema and van Ooyen, 2005). Extending the framework of this model, our finding suggests that this attentional gating is dynamically modulated by the current task.

4.4 Conclusion

In sum, the present study has established stimulus-specific and task-dependent activation following reward delivery in a perceptual decision task. The established features of stimulus specificity and task dependency are important evidence that post-reward activation may be a cortical signature of eligibility traces for credit assignment. This finding is a substantial step towards closing the gap between well-defined computational concepts in reward-based learning and their neural implementation. Important questions for future research will concern the mechanisms of maintaining relevant representations until reward delivery and which dopaminergic circuits, e.g. involving the hippocampus or the ventral striatum, mediate this form of learning.

References

Arsenault, John T, Koen Nelissen, Bechir Jarraya, and Wim Vanduffel (2013). Dopaminergic reward signals selectively decrease fmri activity in primate visual cortex. *Neuron* 77(6): 1174–1186.

Brosch, Michael, Elena Selezneva, and Henning Scheich (2011). Representation of reward feedback in primate auditory cortex. *Frontiers in systems neuroscience* 5: 5.

Daw, N.D. and K. Doya (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology* 16(2): 199–204.

Dayan, Peter and Yael Niv (2008). Reinforcement learning: The good, the bad and the ugly. *Current Opinion in Neurobiology* 18(2): 185–196.

FitzGerald, Thomas H. B., Karl J. Friston, and Raymond J. Dolan (2012). Action-specific value signals in reward-related regions of the human brain. *The Journal of Neuroscience* 32(46): 16417–16423.

Grill-Spector, K, R Henson, and A Martin (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences* 10(1): 14–23.

Hickey, Clayton, Leonardo Chelazzi, and Jan Theeuwes (2010). Reward changes salience in human vision via the anterior cingulate. *The Journal of Neuroscience* 30(33): 11096–11103.

Hickey, Clayton and Wieske Zoest (2012). Reward creates oculomotor salience. *Current Biology* 22(7): R219–R220.

Jay, Therese (2003). Dopamine: a potential substrate for synaptic plasticity and memory mechanisms. *Progress in Neurobiology* 69(6): 375–390.

Lisman, John, Anthony A Grace, and Emrah Duzel (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in neurosciences* 34(10): 536–47.

Lohmann, G, K Mueller, V Bosch, H Mentzel, S Hessler, Li Chen, S Zysset, and D Y Cramon (2001). Lipsia - a new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society* 25(6): 449–457.

Markov, Nikola T, Mária Ercsey-Ravasz, David C Van Essen, Kenneth Knoblauch, Zoltán Toroczkai, and Henry Kennedy (2013). Cortical high-density counterstream architectures. *Science (New York, N.Y.)* 342(6158): 1238406.

O'Doherty, J P (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology* 14(6): 769–776.

O'Doherty, J P, P Dayan, K J Friston, H Critchley, and R J Dolan (2003). Temporal Difference Models and Reward-Related Learning in the Human Brain. *Neuron* 38(2): 329–337.

Pennartz, C M A, R Ito, P F M J Verschure, F P Battaglia, and T W Robbins (2011). The hippocampal-striatal axis in learning, prediction and goal-directed behavior. *Trends in neurosciences* 34(10): 548–59.

Pleger, Burkhard, Felix Blankenburg, Christian C. Ruff, Jon Driver, and Raymond J. Dolan (2008). Reward facilitates tactile judgments and modulates hemodynamic responses in human primary somatosensory cortex. *The Journal of Neuroscience* 28(33): 8161–8168.

Pleger, Burkhard, Christian C. Ruff, Felix Blankenburg, Stefan KlÃ¶ppel, Jon Driver, and Raymond J. Dolan (2009). Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biol* 7(7): e1000164.

Rajimehr, Reza, Kathryn J. Devaney, Natalia Y. Bilenko, Jeremy C. Young, and Roger B. H. Tootell (2011). The parahippocampal place areaâ responds preferentially to high spatial frequencies in humans and monkeys. *PLoS Biol* 9(4): e1000608–.

Rescorla, Robert A and Allan R Wagner (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory* 2: 64–99.

Roelfsema, Pieter R and Arjen Ooyen (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural computation* 17(10): 2176–2214.

Roelfsema, Pieter R, Arjen Ooyen, and Takeo Watanabe (2010). Perceptual learning rules based on reinforcers and attention. *Trends in cognitive sciences* 14(2): 64–71.

Schultz, Wolfram (2007). Behavioral dopamine signals. *Trends in neurosciences* 30(5): 203–10.

Schultz, Wolfram, Peter Dayan, and P Read Montague (1997). A Neural Substrate of Prediction and Reward. *Science* 275(June 1994): 1593–1599.

Schyns, Philippe G. and Aude Oliva (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science* 5(4): 195–200.

Serences, John T. (2008). Value-based modulations in human visual cortex. *Neuron* 60(6): 1169–1181.

Summerfield, C, T Egner, J Mangels, and J Hirsch (2006). Mistaking a house for a face: neural correlates of misperception in healthy humans. *Cerebral cortex (New York, N.Y. : 1991)* 16(4): 500–508.

Sutton, R S and A G Barto (1990). Time-Derivative Models of Pavlovian Reinforcement.

Sutton, Richard S (1988). Learning to predict by the methods of temporal differences. *Machine learning* 3(1): 9–44.

Talairach, Jean and Pierre Tournoux (1988). *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. Thieme.

Waelti, Pascale, Anthony Dickinson, and Wolfram Schultz (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412(6842): 43–48.

Weil, Rimona Sharon, Nicholas Furl, Christian C. Ruff, Michael Symmonds, Guillaume Flandin, Raymond J. Dolan, John Driver, and Geraint Rees (2010). Rewarding feedback after correct visual discriminations has both general and specific influences on visual cortex. *Journal of Neurophysiology* 104(3): 1746–1757.

Willenbockel, V, J Sadr, D Fiset, G. O. Horne, F. Gosselin, and J. W. Tanaka (2010). Controlling low-level image properties: The shine toolbox. *Behavior Research Methods* 42: 671–684.

Wimmer, G. Elliott and Daphna Shohamy (2012). Preference by association: How memory mechanisms in the hippocampus bias decisions. *Science* 338(6104): 270–273.

Worsley, Keith J and Karl J Friston (1995). Analysis of fmri time series revisited - again. *NeuroImage* 2(3): 173–181.

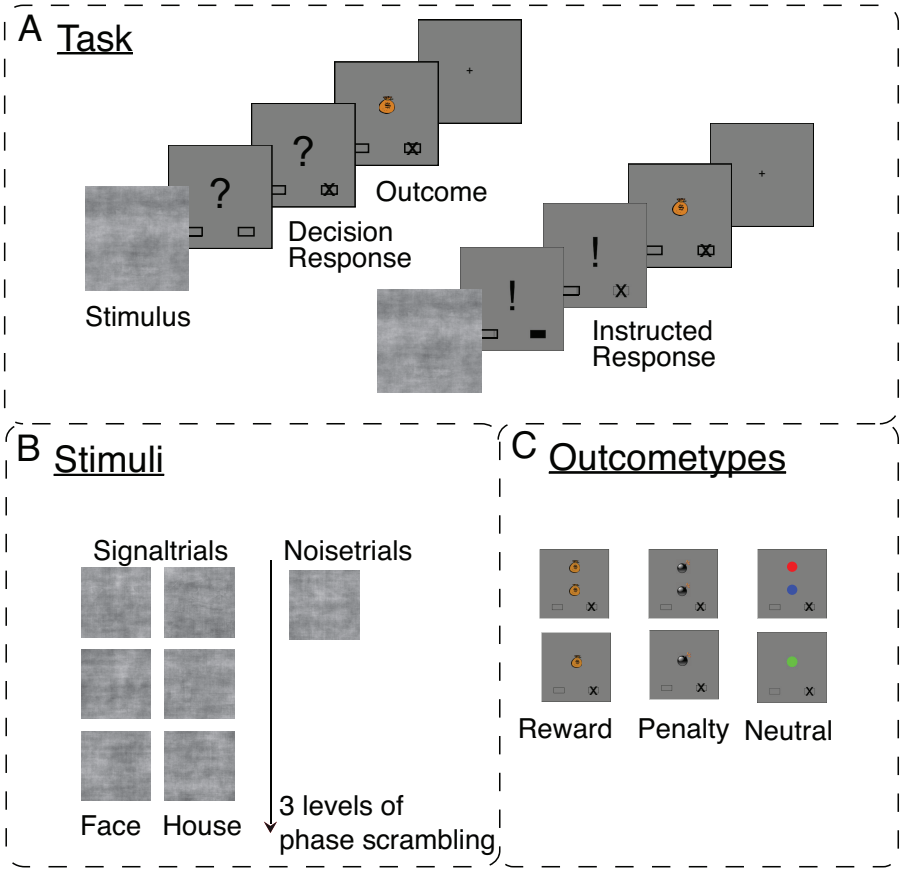

Figure 1: **A** The main task was a perceptual decision task in which participants were first presented with a stimulus, which they had to classify a either face or house. When the question mark appeared, participants had to indicate their decision with a left or right button press. They then received positive (as shown), negative, or neutral feedback. The second task made up 25 % of all trials. Here, the initial stimulus was followed by an instruction (darkened box), which button to press. Feedback was again delivered in the same format as for the perceptual decision task.**B** Three levels of degradation and noise trials were included, yielding graded performance levels. Stimulus degradation was achieved by phase scrambling greyscale images of faces and houses. Participants were unaware of the existence of noise stimuli.**C** Participants experienced 5 levels of outcome valance, 2 levels of reward, 2 levels of penalty and a neutral outcome. Large and small rewards or penalties resulted in the gain or loss of 20 or 10 points, respectively. Outcome valence (reward or penalty) was performance contingent on signal trials and in the instructed response task, but randomly assigned for perceptual decisions in noise trials. A quarter of all trials were followed by neutral outcomes instead independent of performance.

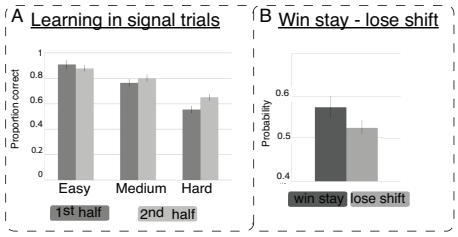

Figure 2: Feedback integration for signal stimuli was apparent in the performance improvement from the 1st half (dark grey bars) of the experiment to the 2nd (light grey bars) on perceptual decision trials. Feedback integration for noise trials was evident in win stay - lose shift behaviour. Participants were significantly more likely to stay than to shift response when it was rewarded and more likely to shift than to stay when a response was penalised.

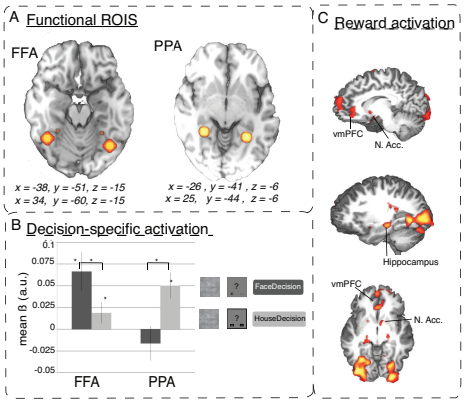

Figure 3: **A** Functional ROIs were created by masking the contrast (FaceLocaliserBlock > HouseLocaliserBlock) and (HouseLocaliserBlock > FacelOcaliserBlock), respectively, with the corresponding parametric contrast for signal increase with signal strength from the main experiment. **B** Contrasting BOLD response for stimuli leading up to face decisions (dark grey) on signal trials with those leading up to house decisions (light grey) revealed significantly stronger activity for face decisions in the FFA (left) and for house decisions in the PPA (right) compared to the respective alternative decision. **C** Rewarding outcomes after decisions on noise trials activated a network classically associated with reward delivery, including the ventromedial prefrontal cortex, and nucleus accumbens, as well as the anterior hippocampus.

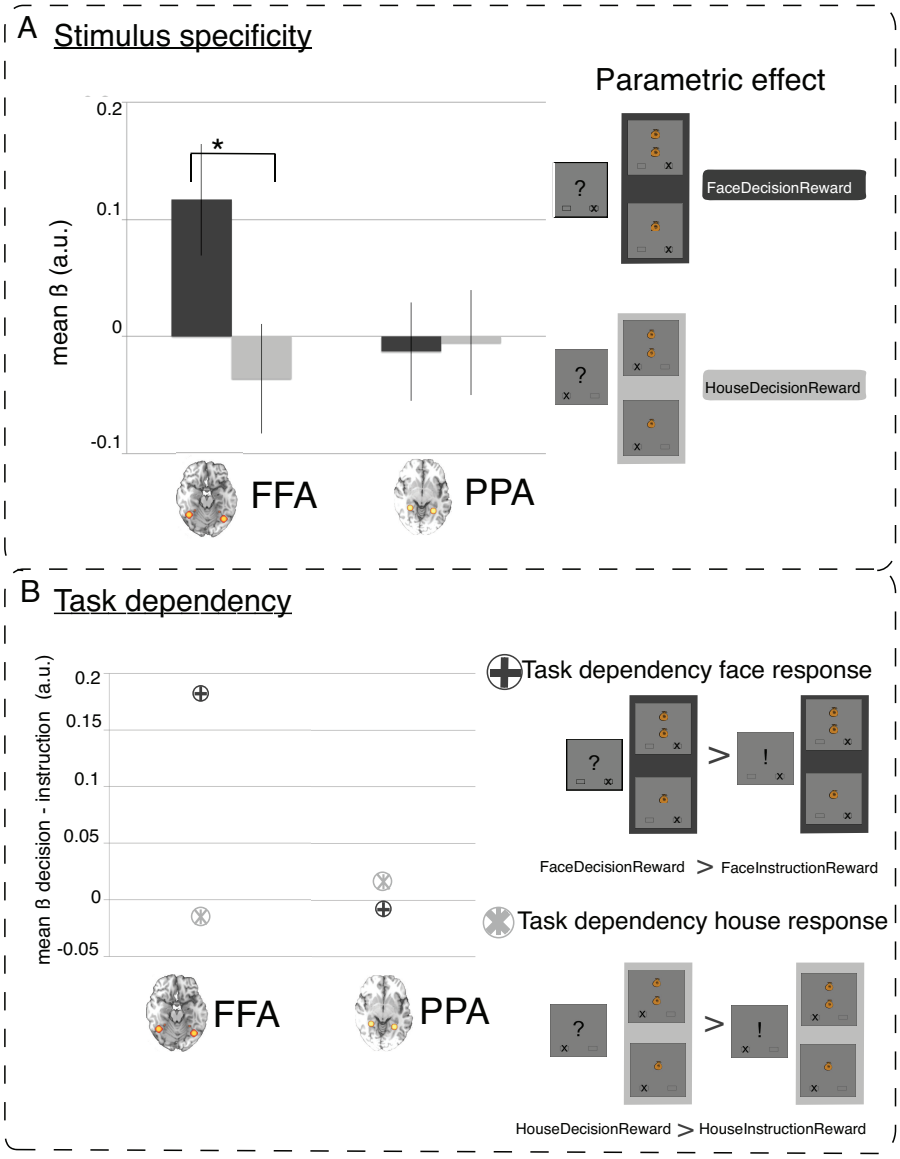

Figure 4: **A** *Stimulus-specific* activity in noise trials was measured as an increase in BOLD activity with reward size, that was significantly stronger for the decision associated with an ROI (face for FFA/ house for PPA) than for the opposite decision. Dark grey bars show mean betas for parametric increase with rewardsize for face decisions, light grey bars show mean betas for parametric increase with rewardsize for house decisions. Left: Activity in the FFA ROI, right: activity in the PPA ROI. **B** *Task dependency* was reflected in a larger parameter estimate (mean beta) for BOLD increase with reward size in the associated ROIs (FFA left, PPA right) in the perceptual decision compared to the instructed response tasks. Markers refer to the difference of mean betas in the perceptual decision and instructed response task; the difference for face responses are shown in dark grey, house responses in light grey.