NeuroImage 236 (2021) 118028

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA

Lena M. Schliephake[a,*], Ima Trempler[a,b], Marlen A. Roehe[a,b], Nina Heins[a,b], Ricarda I. Schubotz[a,b]

[a] Department of Psychology, University of Muenster, Muenster, Germany
[b] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

ARTICLE INFO

Keywords:
Prediction error
Predictive coding
Psychophysiological interaction
Visual processing hierarchy
fMRI

ABSTRACT

Surprising scenarios can have different behavioural and neuronal consequences depending on the violation of the expectation. On the one hand, previous research has shown that the omission of a visual stimulus results in a robust cortical response representing that missing stimulus, a so-called *negative prediction error*. On the other hand, a large amount of studies revealed *positive prediction error* signals, entailing an increased neural response that can be attributed to the experience of a surprising, unexpected stimulus. However, there has been no evidence, so far, regarding how and when these prediction error signals co-occur. Here, we argue that the omission of an expected stimulus can and often does coincide with the appearance of an unexpected one. Therefore, we investigated whether positive and negative prediction error signals evoked by unpredicted cross-category stimulus transitions would temporally coincide during a speeded forced-choice fMRI paradigm. Foremost, our findings provide evidence of a behavioural effect regarding the facilitation of responses linked to expected stimuli. In addition, we obtained evidence for negative prediction error signals as seen in differential activation of FFA and PPA during unexpected place and face trials, respectively. Lastly, a psychophysiological interaction analysis revealed evidence for positive prediction error signals represented by context-dependent functional coupling between the right IFG and FFA or PPA, respectively, implicating a network that updates the internal representation after the appearance of an unexpected stimulus through involvement of this frontal area. The current results are consistent with a predictive coding account of cognition and underline the importance of considering the potential dual nature of expectation violations. Furthermore, our results put forward that positive and negative prediction error signalling can be directly linked to regions associated with the processing of different stimulus categories.

1. Introduction

In our everyday life, surprise takes different forms. Consider the situation in which the doorbell rings and you go open the door. Different surprising scenarios are conceivable: (a) you face a stranger who took the wrong door; (b) you find nobody waiting outside (being victim of a knock-a-door-run, presumably); (c) you experience a double surprise because it's not the friend you expected but a package deliverer who took the wrong door. The latter example illustrates that in everyday life, the *omission* of an *expected* stimulus can – and often does – temporally coincide with the *appearance* of an *unexpected* one.

Recent neurocognitive models describe perception as a process of inference where top-down predictions are compared with bottom-up sensory evidence along the visual cortical hierarchy to constantly update our internal model of the outside world (Clark, 2013; Friston, 2005). It has been suggested that this process is carried out by two distinct classes of neurons. Internal representation neurons forward predictions regarding the conditional probability of a stimulus to a lower level whereas prediction error neurons encode the mismatch between the predictions and actual bottom-up information and propagate this prediction error to the next higher level (Egner et al., 2010). To facilitate cognition, only violations of our expectations ("prediction errors") are propagated upwards the cortical hierarchy to update internal representations at the next higher level.

Referencing back to the surprising scenarios mentioned previously, these so-called prediction errors can either be positive or negative depending on the nature of the violation (Keller and Mrsic-Flogel, 2018). Positive prediction error neurons are activated when bottom-up input increases unexpectedly, for example after the presentation of an unexpected visual stimulus. As opposed to that, the activation of negative prediction error neurons represents an unexpected decrease of sensory input, for example when a stimulus has been unexpectedly omitted.

* Corresponding author.
E-mail address: lschliep@uni-muenster.de (L.M. Schliephake).

https://doi.org/10.1016/j.neuroimage.2021.118028
Received 16 October 2020; Received in revised form 23 February 2021; Accepted 30 March 2021
Available online 28 April 2021
1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

While the distinction between positive and negative prediction errors is firmly established in the field of reward prediction (errors), researchers put forward that prediction error signals occur independently of rewards (Gardner et al., 2018; Schiffer et al., 2015) . Consequently, a more generalised concept has been introduced to account for findings showing the presence of prediction errors for unexpected and novel stimuli in the absence of rewards (Horvitz, 2000; Menegas et al., 2017).

Up until today, most previous studies in this field focused on only positive (e.g., Amado et al., 2016; Egner et al., 2010; Meyer and Olson, 2011) or only negative prediction errors (Fiser et al., 2016; den Ouden et al., 2012; Eliades and Wang, 2008; Stanley and Miall, 2007). One reason for this might be the difficulty of designing experiments that can simultaneously differentiate between the neuronal processes associated with both types of prediction errors. However, we think that it is important to consider the potential interplay of both positive and negative prediction error signals as this parallel processing might imply, that in fact, error computations within the brain are carried out by two separate prediction-error circuits: One processing the unexpected omission of a visual stimulus and one processing the unexpected appearance of another stimulus (Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999). Moreover, it is conceivable that depending on the context the processing of positive prediction errors is favoured over the processing of negative prediction errors when for example the appearance of the unexpected stimulus is more relevant for our behaviour than the omission of a competing stimulus.

Therefore, the aim of the current fMRI study was to investigate the effects of positive and negative prediction errors during implicit expectation violations of cross-category stimulus associations. More specifically, we examined fMRI blood oxygenation level-dependent (BOLD) responses during a speeded forced choice task while participants implicitly predicted face or place stimuli generating positive and negative prediction errors.

Several previous studies have shown that face and place stimuli are processed in two distinct brain areas. Faces are preferentially processed by the fusiform face area (FFA) (Haxby et al., 2000; Liu et al., 2002), whereas place and house stimuli elicit higher activations in the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998; Ishai et al., 2000). Moreover, both imagery and expectation of faces and places lead to activity increases in FFA and PPA, respectively (Esterman and Yantis, 2010; O'Craven and Kanwisher, 2000). Introducing these two distinct stimulus categories enabled us to investigate differential brain responses resulting from an omitted stimulus category (e.g., a face) and, at the same time, looking at brain activation resulting from an unexpectedly presented stimulus from the other category (e.g., a place).

In addition to stimulus-specific prediction error signals in FFA and PPA, we expected frontal areas to respond to expectation violation, especially the right IFG. Activity in this area has been found to generally increase for prediction errors in different modalities (Chao et al., 2018; Trempler et al., 2020) presumably scaling with the amount of modification needed to adapt the current predictive model accordingly (Alexander and Brown, 2018; El-Sourani et al., 2019; Keller and Mrsic-Flogel, 2018).

Before fMRI scanning, participants were trained to implicitly learn the probability of the two category transitions within a sequence of face and place stimuli, which occurred at the same base rate. Crucially, within-category transitions (face-face and place-place) occurred with a much lower probability than between-category transitions (face-place and place-face). For these more expected stimuli, faster reaction times and higher accuracy levels were hypothesised. With regard to BOLD fMRI, we hypothesised to find both evidence of positive and negative prediction errors during expectation violation. By using two distinct stimulus categories, we were able to test positive and negative prediction errors within different category-specific areas, namely the FFA and the PPA: for unexpectedly as compared to expectedly appearing stimuli, we hypothesised a positive prediction error as reflected in a BOLD increase in the brain region preferentially processing this stimulus category and at the same time, a negative prediction error as reflected in a BOLD increase in the brain area that was non-preferentially processing the current stimulus category representing the omitted stimulus category. Here, activity evoked by a face stimulus represents "preferred" beta activity in FFA and "non-preferred" beta activity in PPA. The same applies to place stimuli where "preferred" activity can be observed in PPA and "non-preferred" activity in FFA. Furthermore, we were interested in changes in functional connectivity between stimulus-specific prediction error signals in PPA and FFA and the right IFG that has previously been implicated to play an important role in model updating and error processing (e.g., El-Sourani et al., 2019). To this end, we conducted a psychophysiological interaction analysis (PPI) and examined whether expectation violation modulates the functional connectivity between the FFA and PPA on the one hand, and the right IFG on the other hand.

2. Materials and methods

2.1. Participants

Thirty-four volunteers (25 females) participated in the current study. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). After initial inspection of the behavioural data, one participant was excluded from all further analyses, because of comparably low performance levels (accuracy more than 2 *SDs* below the mean accuracy level of all other participants). The remaining 33 participants (24 females) were between 18 and 35 years of age ($M = 24$; $SD = 3.49$). All participants had normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Before participation, all volunteers gave written informed consent. Participants were debriefed and reimbursed after taking part in the fMRI experiment. The study was performed following the Declaration of Helsinki and had been approved by the ethics committee of the University of Muenster.

2.2. Stimuli

In total, 32 photographs were used in the current paradigm. We employed 16 unique, colour photographs of faces (8 females, 8 males) with neutral facial expression and 16 unique place images (8 indoor, 8 outdoor). Since our participants usually identify as female/women or male/men, we wanted to make sure that pictures of both women and men were included in the stimuli used for the paradigm. Subsequently, we also included two subcategories for the place stimuli, indoor and outdoor, respectively.

Face stimuli were drawn from the Radboud Face Database (Langner et al., 2010). Place images were collected from the internet and, like the face images, cropped and resized using GIMP. All stimuli were individually displayed in the centre of a uniform grey background.

2.3. Trial sequence and task design

The stimulus randomisation was programmed using Matlab (Version R2018a; The MathWorks Inc., Natick, MA, USA). Stimuli were presented and participants' responses recorded using the Presentation Software (Version 18.1, Neurobehavioral Systems, Inc., Berkeley, CA). The 32 photographs were randomly assigned to two different stimulus sets. In both sets, 16 images including four photographs of each category (female, male, indoor, outdoor) were repeated eight times resulting in 128 trials per block. To introduce statistical regularities, we manipulated the transitional probabilities of the stimulus sequences. Transitions were either within- or between-category so that a face followed a face (within-category) or a place stimulus (between-category) and a place followed either a place (within-category) or a face stimulus (between-category). The probability of within-category transitions was 35% whereas the between-category transition probability was 65% (Fig. 1). With this manipulation, we ensured that participants would be able to predict the

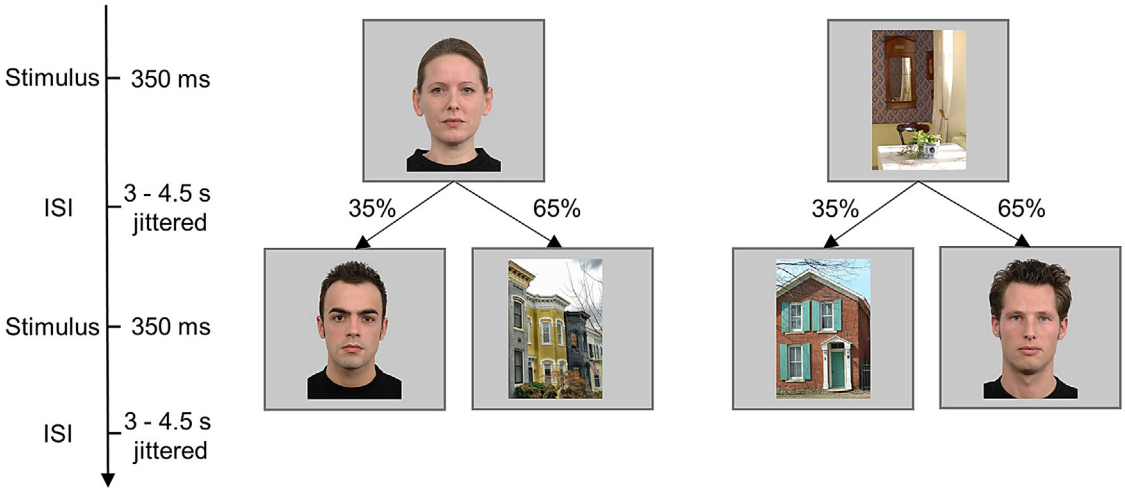

Fig. 1. Example stimuli including stimulus transitions with their corresponding transitional probabilities and underlying time course. The "expected" condition included between-category transitions (i.e., Face–Place; Place–Face) whereas the "unexpected" condition included within-category transitions (i.e., Face–Face; Place–Place.). The transition probabilities were independent of the subcategories (i.e., female/male; indoor/outdoor). For each trial, participants had to indicate the stimulus category (face or place) via a button press (left or right).

most likely succeeding stimulus category when presented with the current stimulus. Hereafter, we will therefore call the 65% transition probability condition "expected" and the 35% transition probability condition "unexpected". The paradigm contained additional manipulations, including reoccurring stimulus pairs of different probabilities that we are pursuing separately. The pairs were imbedded in the more likely stimulus category transition meaning that they were formed by using unequal stimulus categories. In total, there were eight stimulus pairs consisting of a particular face that was preceded by a particular place or a particular place that was preceded by a particular face. In half of the blocks, the second stimulus followed the first stimulus in 100% of the cases and in the other half the probability was 75%. This additional manipulation was included to investigate whether identity learning (that would follow the learning of the particular stimulus pairs) would be distinguishable from category learning (that would follow the learning of the probability of the category transitions). The inclusion of stimulus pairs was pursued in a different analysis, which was not part of this article. Moreover, because of the specific nature of the pairs, we did not expect this manipulation to influence the present analyses in any way .

2.4. Task

All participants completed a 20-min training one day before the fMRI session. During the training, participants were exposed to the same stimulus sets that would later be presented in the fMRI session. The training consisted of four experimental blocks including two block repetitions of the two different stimulus sets. The stimuli were presented on a computer screen. The aim of this training was twofold: first, to familiarise participants with the task and second, to induce learning of the statistical regularities of the stimulus sequence. Photographs were presented for 350 ms with inter-trial-intervals (ITI) of 2000 ms. After each block, participants had a short break of 7 s.

Before training, participants were told that we were investigating how the brain reacts while being exposed to different types of images. By pressing one of two buttons, participants were asked to respond as quickly and accurately as possible as to whether the shown photograph was a face or a place on every trial. The button allocation determining which button represented which image category was counterbalanced across all participants. During the main task in the MR scanner, participants had to perform the same task. Here, photographs were shown for 350 ms followed by a jittered ITI of 3000, 3500, or 4000 ms. This time, participants completed six blocks resulting in 768 trials in total and a test session of about 50 min in the scanner.

2.5. FMRI data acquisition

Whole-brain imaging data were recorded with a 3-T Siemens Magnetom Prisma scanner (Siemens, Erlangen, Germany) using a 20-channel head coil. Functional blood oxygenation level-dependent (BOLD) images were acquired parallel to the anterior commissure/posterior commissure line with a T2*-weighted gradient echo planar imaging (EPI) sequence (64 × 64 data acquisition matrix; 192 mm field of view (FOV); 90° flip angle; time of repetition (TR) = 2000 ms; echo time (TE) = 30 ms). Each volume consisted of 33 adjacent axial slices with a slice thickness of 3 mm and a gap of 1 mm, resulting in a voxel size of 3 × 3 × 4 mm. Structural images were acquired for each participant using a standard Siemens 3D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxel size (1 × 1 × 1 mm) in a 256-mm FOV (256 × 256 matrix, 196 slices; RT = 2130 ms; TE = 2.28 ms). Stimuli were projected on a screen that was positioned behind the scanner bore. They were presented in the centre of the field of vision by a video projector, and participants viewed the screen by a 45° mirror, which was fixated on the top of the head coil and adjusted for each participant to provide a good view of the entire screen. Participants' right index and middle finger were placed on two response buttons, matching the response contingencies from the training session. Participants' arms were stabilized on form-fitting cushions. Additionally, foam padding around the head was applied to prevent motion artefacts. Earplugs and noise-cancelling headphones were provided to reduce scanner noise.

2.6. Data analysis

2.6.1. Behavioural data analysis

The following behavioural analysis steps were applied to the training data as well as the behavioural data from the main experiment. Reaction times and accuracy (percentage of correct responses) analyses were performed with Matlab (Version R2019a; The MathWorks Inc., Natick, MA, USA). Trials with no response and incorrect responses were categorised as error trials. Additionally, we defined a maximum response window of 1500 ms starting from trial onset during which participants' responses were classified as valid as for example suggested by McKendrick et al. (2014). Responses recorded after these 1500 ms were also categorised

as error trials. Task performance reflected the mean percentage of correctly answered trials of all correct and error trials, which was calculated for expected and unexpected trials. Due to technical issues during the recording of the training session of the first participant, we had to exclude this participants' behavioural data from the analysis of the training. Additionally, as already indicated earlier, one participant was excluded from all analyses because of comparably poor performance (75.46% correct responses). Therefore, 32 participants remained for the analysis of the training session and 33 participants remained for the main analyses. Since we hypothesised that responses for the expected stimulus categories would display an increase in accuracy aligned with a decrease in reaction times, these two aspects of performance were subjected to Bonferroni corrected individual paired-samples *t*-tests.

2.6.2. FMRI data analysis

*Preprocessing:*All preprocessing and statistical analyses were performed with SPM12 (www.fil.ion.ucl.ac.uk/spm) and custom Matlab scripts (Version R2019a; The MathWorks Inc., Natick, MA, USA). For each participant, functional images were slice time corrected and spatially realigned to the first volume. Structural images were coregistered to the mean functional image and then used to calculate transformation parameters for normalising the functional images to the Montreal Neurological Institute (MNI) template brain. The normalised functional images (resampled at 3 mm^3) were spatially smoothed with a Gaussian kernel of full width at half-maximum of 6 mm^3.

Region of interest (ROI) analysis: To analyse our main task we used a summary statistic random effects approach. At the first level (within-subjects), we estimated parameters encoding condition-specific activations. This involved specifying stimulus functions for each trial type. These functions were then convolved with a canonical hemodynamic response function to form our condition regressors. Separate regressors were entered for faces and places as well as expected and unexpected trials resulting in ten different regressors: face_expected (FE; face preceded by a place), face_unexpected (FU; face preceded by a face), place_expected (PE; place preceded by a Face), place_unexpected (Place preceded by a Place) as well as six regressors for the motion parameters (three translations and three rotations). In order to analyse BOLD responses with regard to our two ROIs (FFA and PPA), we used probabilistic masks representing FFA and PPA, respectively, from the SPM Anatomy toolbox (Lorenz et al., 2017). Therefore, all ROIs were statistically independent from our data. We then used these ROIs to extract each subject's beta values of the activity associated with each trial type during the main task. Since a profound amount of research has shown that face images are dominantly processed by the right hemisphere (Rossion, 2014; Bukowski et al., 2013; Bentin et al., 1996; Rangarajan et al., 2014), we decided to include only the right FFA and correspondingly also only the right PPA as ROIs into our design. Therefore, beta values were subjected to a $2 \times 2 \times 2$ repeated measures ANOVA with the factors: ROI (right FFA vs. right PPA), expectation (expected vs. unexpected), and preference (preferred vs. non-preferred). This coding was adapted from previous studies looking at stimulus-specific activity with regard to expectation and surprise (Yon et al., 2018). The three-way ANOVA was calculated to make sure that the ROIs did not have a differential influence on the two factors Expectation and Preference, which were of main interest to us. To test whether unexpected stimuli modulated responses in the preferred or non-preferred condition, we implemented post-hoc contrasts of the conditions expected_preferred vs. unexpected_preferred and expected_non-preferred vs. unexpected_non-preferred as Bonferroni corrected paired two-sided *t*-tests. Greenhouse–Geisser correction for sphericity violations were used for reported degrees of freedom and *p*-values where appropriate.

Psychophysiological interaction analysis: We conducted a condition-specific psychophysiological interaction (PPI) analysis to investigate the context-dependent functional coupling between the right IFG and stimulus-specific prediction error processing areas of our task. Specifically, we were interested in the functional coupling of the right IFG and positive and negative prediction error related signals resulting from unexpected face or place stimuli FU > FE and PU > PE. We performed two independent PPIs to be able to identify whether potential FFA and PPA activity was reflecting positive or negative prediction error effects. To define our seed ROI, we constrained our search to a volume of interest that was based on the group fMRI analysis contrasting unexpected with expected trials ($U > E$) and functionally localised the right IFG using the Anatomy Toolbox (Eickhoff et al., 2005). This definition of our seed ROI followed a standard procedure described in O'Reilly et al. (2012). To this end, we extracted the time course of right IFG activity with a 6 mm radial sphere using voxels that showed peak activation for the contrast ($U > E$); $x = 51$, $y = 17$, $z = 17$) with marsbar (Poldrack, 2007). Furthermore, we extracted the first eigenvariate of our seed sphere and allowed actual VOIs to vary in size between participants but restricted them to the first level masks. We then *z*-transformed the time course values and generated the PPI regressor by multiplying the physiological regressor (time course of right IFG) with the convolved psychological regressors (FU > FE and PU > PE). For each participant, both the physiological and the PPI regressors were added to the original design matrix. The resulting matrices were then entered into a random-effects group analysis in which the PPI regressor was tested. As we had strong *a priori* hypotheses about the areas of the brain regarding their involvement in right IFG coupling during unexpected face and place trials, we used the same independently defined ROIs (right FFA and right PPA, respectively) as in the univariate analysis to correct for multiple testing (small volume correction). According to our hypotheses, positive coupling between the right IFG and the right FFA and PPA during unexpected trials would imply the conjecture that the right IFG is updating the current internal model based on surprise.

3. Results

3.1. Behavioural results

3.1.1. Training

Implicit learning of the transition probabilities between stimulus categories was already hypothesised for the training session. Therefore, this learning should result in shortened response times as well as elevated accuracy levels for expected as compared to unexpected stimuli. This was indeed the case, as response times for expected when compared to unexpected trials were significantly faster (419.79 ms ($SD = 55.08$) vs. 429.55 ms ($SD = 53.23$), $t_{(31)} = -2.82$, $p < 0.005$). Additionally, participants showed higher accuracy levels (percentage of correct responses, CR) for expected vs. unexpected trials (93.26% CR ($SD = 15.47$) vs. 91.20% CR ($SD = 14.64$); $t_{(31)} = 4.17$, $p < 0.001$).

3.1.2. Main experiment

As for the training session, if implicit learning of the transitional probabilities between stimulus categories would still be evident during the main experiment, performance with regard to predictable trials should be facilitated. This pattern was observed in response times for expected vs. unexpected trials (576.46 ms ($SD = 89.85$) vs. 589.59 ms ($SD = 86.53$), respectively; $t_{(32)} = -4.00$, $p < 0.001$) (Fig. 2A). A facilitation effect could also be observed with regard to accuracy (CR). CR was higher for expected vs. unexpected trials (97.05% CR ($SD = 2.77$) vs. 95.16% CR ($SD = 4.19$); $t_{(32)} = 4.44$, $p < 0.001$) (Fig. 2B).

3.2. fMRI results

The appearance of an unexpected stimulus should, on the one hand, result in a positive prediction error with regard to the regions' preferred stimulus category (i.e., FFA activation for unexpected faces and PPA activation for unexpected places) as well as a negative prediction error signal regarding the regions' non-preferred stimulus category (i.e., PPA activation of unexpected faces and FFA activation for unexpected faces).

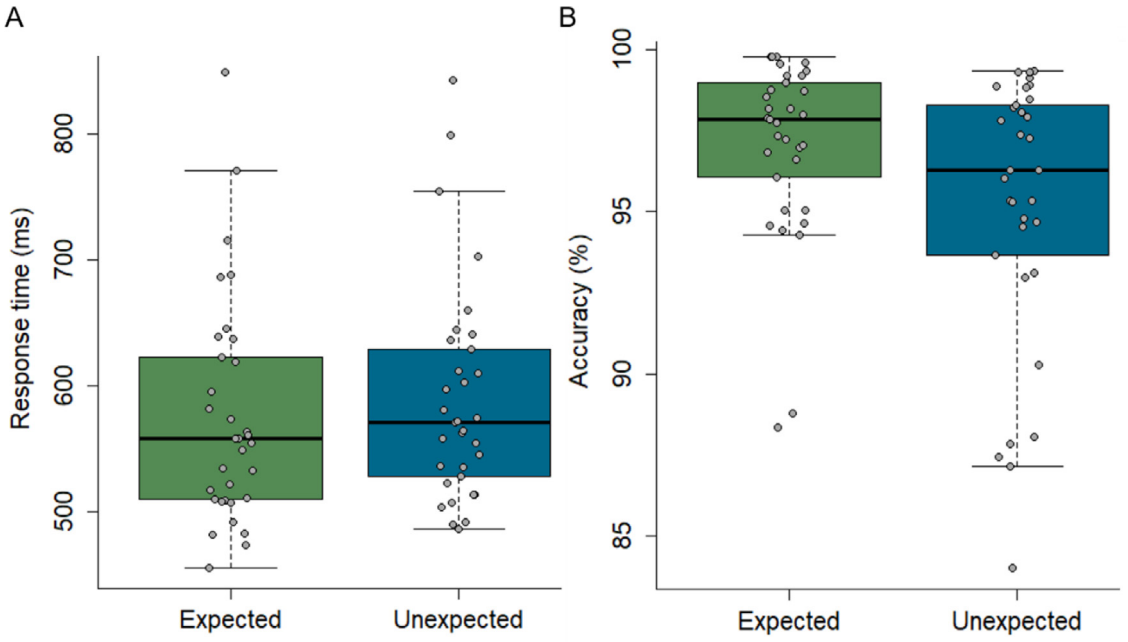

Fig. 2. Behavioural measures. **(A)** Mean response times for expected and unexpected trials and **(B)** mean percentage of correct responses. Black horizontal lines represent the mean values, boxes represent the standard error of the mean (*SEM*), and grey vertical lines represent the standard deviation (*SD*). The grey circles represent individual data points of the 33 participants.

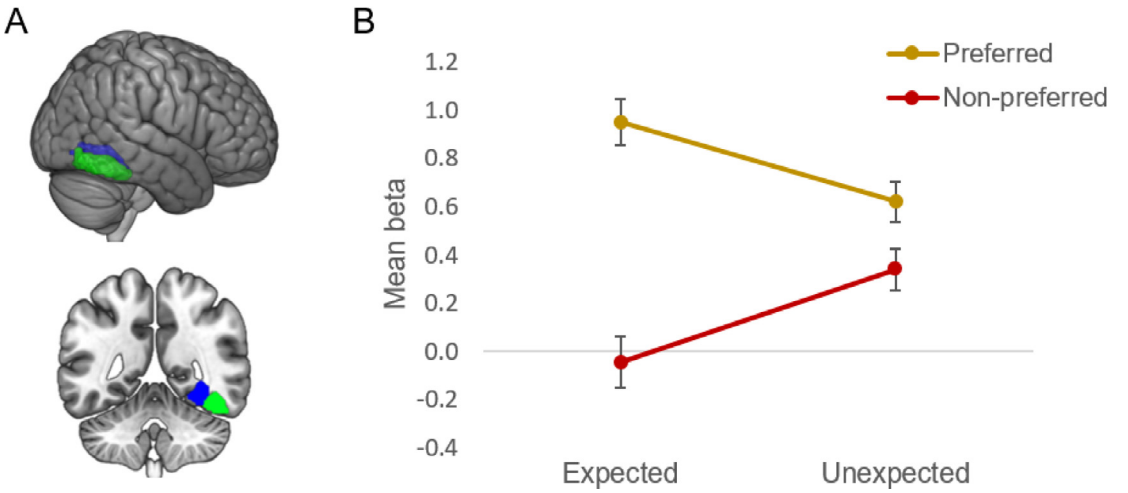

Fig. 3. Region of interest (ROI) fMRI data. **(A)** Right FFA and right PPA ROIs used for beta value extraction. **(B)** Mean group activation estimates (β parameters +/− SEM) for expected_preferred, expected_non-preferred, unexpected_preferred, and unexpected_non-preferred trials.

A three-way repeated measures ANOVA was run on a sample of 33 participants to examine the effect of expectation, preference, and ROI on the extracted beta values. There was no significant main effect of expectation showing that there was no general prediction error effect for both positive and negative surprises ($F_{(1,32)} = 0.183$, $p = 0.672$). However, the effect of expectation was dependent on whether the presented image was a region's preferred or non-preferred stimulus as revealed by a significant two-way interaction between preference and expectation ($F_{(1,32)} = 149.54$, $p < 0.001$) (See Fig. 3). This interaction effect, however, was independent of the two ROIs as the results yielded no significant three-way interaction between the factors ROI, preference, and expectation ($F_{(1,32)} = 0.02$, $p = 0.901$). The results of the post-hoc contrasts revealed a significant difference between the conditions expected_preferred and unexpected_preferred as on average, expected_preferred stimuli evoked larger beta values as unexpected_preferred stimuli ($t_{(32)} = 5.31$, $p < 0.001$). On the other hand, the comparison between expected_non-preferred and unexpected_non-preferred trials yielded a significant difference resulting from higher beta values for unexpected_non-preferred trials ($t_{(32)} = -5.44$, $p < 0.001$). With regard to our hypotheses, the results show no general prediction error effect of both positive and negative surprises as revealed by the non-significant main effect of expectation. Moreover, the prediction error effect that was expected to result from unexpected as compared to expected trials seemed to be dependent on whether the presented stimulus was preferred or non-preferred to our ROIs. The results revealed an increase for unexpected as compared to expected trials only within a stimulus' non-preferred region which represents a negative prediction error effect in our study. As opposed to that, the univariate analysis did not show significant positive prediction error effect, as on average, beta values for expected trials as compared to unexpected trials were higher in the preferred condition. A detailed overview of all main effects and interactions can be found in the Supplementary material.

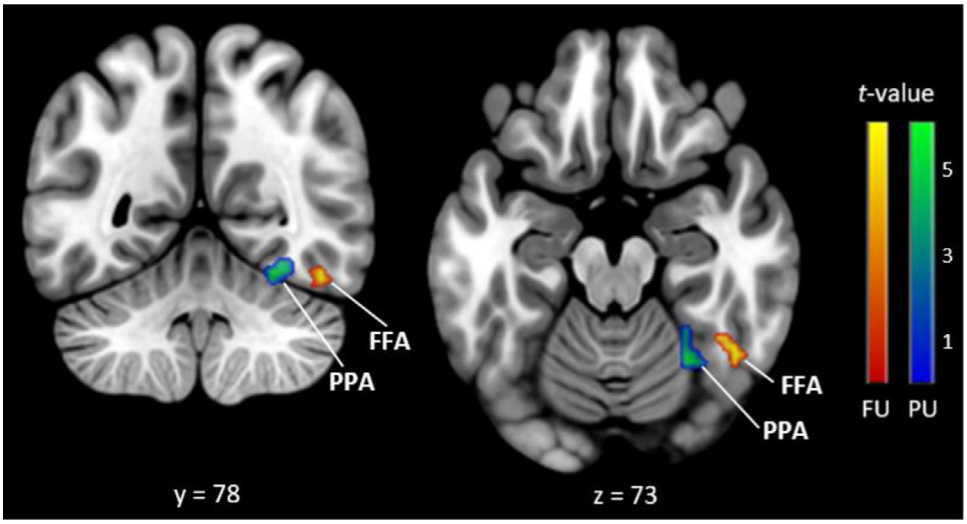

Fig. 4. Results of PPI analyses for the right IFG as seed region on the right FFA for FU > FE (in red) and on the right PPA for PU > PE (in blue). For display purposes, the activation identified by the PPI in the respective ROIs (FFA, PPA) is shown at $p < 0.005$ (*uncorrected*). Significance of the activation in the two ROIs is tested using SVC at $p < 0.05$, FDR [SVC]-corrected (see Table 1).

Table 1
PPI analyses of the right IFG. After small volume correction, the results were FDR corrected at $p < 0.05$ voxel level. Cluster extent (k) is indicated in voxels. MNI, Montreal Neurological Institute.

Region	k	t-value of peak voxel	MNI coordinates		
			x	y	z
FU > FE					
Right fusiform gyrus	17	3.26	45	−52	−19
PU > PE					
Right parahippocampal place area	41	4.33	30	−58	−16

To further investigate the principles of positive and negative prediction error effects, we implemented two PPI analyses, where we examined whether the right IFG was coupled with negative and positive prediction error signals during unexpected faces and places when compared to expected faces and places, respectively. After using small volume correction for our predefined ROIs and a threshold of $p < 0.05$ (FDR) at voxel level, results demonstrated that for unexpected faces when compared to expected faces, the right IFG showed context-dependent connectivity with the right FFA. Correspondingly, for unexpected places as compared to expected places, the right IFG was significantly coupled with activation in the right PPA, corroborating the IFG's stimulus-specific engagement in processing positive prediction error signals for unexpected stimuli. See Fig. 4 and Table 1 for an overview of the PPI results.

4. Discussion

The current fMRI study investigated co-occurring positive and negative prediction error signals resulting from expectation violations of learned cross-category stimulus-stimulus transitions. We examined the BOLD response evoked by the unexpected omission of one stimulus category (e.g., a face) and the simultaneous unexpected presentation of the other stimulus category (e.g., a place). The results of our univariate analysis provide evidence for negative prediction error signals as seen in increased brain activation evoked by unexpected and non-preferred stimuli (e.g., PPA activation during unexpected faces). However, we did not find a positive prediction error effect as the ROI analysis revealed a significant activation increase in expected preferred as compared to unexpected non-preferred stimuli (e.g., FFA activation increase for expected faces when compared to unexpected faces). This finding seems to be more in line with previous literature on prediction enhancement effects. Nevertheless, we found evidence for positive prediction error signalling in our PPI analyses. Here, the results suggest context-dependent functional coupling during unexpected faces as compared to expected faces between the right inferior frontal gyrus (IFG) and right FFA as well as coupling between right IFG and right PPA during unexpected places compared to expected places. This implicates a network that updates the internal model after the appearance of an unexpected stimulus through involvement of the right IFG. In the following, we will discuss these findings in more detail.

Behavioural results confirmed that participants learned the given probability distribution of stimulus transitions. Thus, stimuli matching the expected transition probabilities were categorised quicker and more often correctly than non-matching stimuli; hence, reflecting the hypothesised facilitation effect resulting from predictive processes (Esterman and Yantis, 2010; Turk-Browne et al., 2010). As hypothesised, corresponding effects of predictive processing were also reflected in stimulus- and region-specific neural activation patterns. Specifically, the omission of expected stimuli lead to an increased hemodynamic response in the unexpected stimulus' non-preferred brain region, reflecting a negative prediction error (Egner et al., 2010). Negative prediction error signals have mainly been investigated in so-called omission paradigms where the expected stimulus is withheld but a robust cortical response in the relevant cortical area can still be measured (den Ouden et al., 2012; Fiser et al., 2016; Kok et al., 2013). The present results extend these findings by showing that the omission of an expected stimulus that is replaced by an unexpected stimulus results in a similar negative prediction error signal as found in omission paradigms. Thus, it seems that omission responses do not depend on a stimulus-free period during which the specific stimulus was expected, but rather on the experience that an expected stimulus does not appear.

As opposed to our hypothesis, we did not find a significant positive prediction error effect in our univariate analysis. Instead, we found significantly larger responses for expected preferred when compared to unexpected preferred stimuli representing a profound expectation enhancement effect. This finding stands in contrast with studies reporting a stimulus-specific BOLD *decrease* for expected compared to unexpected stimuli (Egner et al., 2010; Meyer and Olson, 2011). It has been suggested that top-down expectation of stimulus (feature) repetitions reduces the prediction error, which in turn results in a reduced BOLD response in the relevant areas, also called *fMRI adaptation* (Grill-Spector et al., 2006; Krekelberg et al., 2006; Miller et al., 1991; Segaert et al., 2013).

A possible explanation for the prediction enhancement we found might, however, be derived from recent "sharpening" models of prediction (Press and Yon, 2019; Press et al., 2020). These suggest that

expectations generate increased activation of predicted sensory units and a relative suppression of unexpected ones. In the context of this paradigm, these models might provide a mechanistic explanation for the positive prediction enhancement effect. If for example the observer expects a stimulus of a certain category (e.g., a face), this expectation generally pre-activates units tuned to these stimuli (i.e., pre-activating FFA, thereby suppressing PPA). This pre-activation would lead to the observed prediction enhancement effect. Additionally, studies that used expectation omission designs (e.g., Kok et al., 2014), might explain the negative prediction error effect we found in the current study. If we assume that expectations pre-activate neuronal templates of the predicted stimulus, a negative prediction error signal, as found in this study, might result from the pre-activated sensory units.

Nevertheless, there is an important difference between previous expectation omission designs and the current study design. Traditional omission designs simply withhold an expected stimulus and do not replace it with another (unexpected) stimulus (Kok et al., 2014; den Ouden et al., 2012). Therefore, omission studies are only able to investigate responses to expected stimuli on the one hand and unexpectedly omitted stimuli that might result in negative prediction errors on the other hand. Therefore, sharpening models do not explain the positive prediction error effect that we found in the context of the PPI analysis.

Another possible explanation for the prediction enhancement would be that the particular transition probabilities gave rise to enhanced stimulus-specific activations during expected trials as suggested by previous findings (den Ouden et al., 2012; Keller and Mrsic-Flogel, 2018). In the current study, the most likely stimulus transition was a change in stimulus category (i.e., face-place; place-face). In line with predictive coding models in cognitive neuroscience (Friston et al., 2005; Keller et al., 2018), the unequal stimulus category transitions might have triggered low-level positive prediction errors during expected trials. These models assume that representation units encode the brain's current hypothesis about the outside world (e.g., "I am looking at a face now") independent of expectations about the future or the present. Furthermore, prediction error neurons in this framework can be understood as simple input units that signal incoming information from lower levels. In the context of unequal stimulus category transitions, this means that when a place appears after the observer looked at a face, the activation of prediction error neurons (that are tuned to places) lead to an adjustment in the representation neurons.

As a consequence, the constant category switches and corresponding low-level prediction errors, might have drawn upon attentional resources thereby increasing activation in the stimulus-specific areas (Blondin and Lepage, 2005; Chen et al., 2012; Maunsell and Treue, 2006; Posner et al., 1980). Even though the motivational considerations that guide attention are principally orthogonal to visual expectations (guided by perceptual regularities), expectation and attention often coincide and interact (Summerfield and Egner, 2009).

In future studies, it would be of interest to explicitly address the interplay of low-level prediction errors and expectation effects and to further investigate, under which circumstances expectation leads to fMRI suppression or prediction enhancement.

During the PPI analyses, we examined whether activation in the right IFG was directly associated with positive and/or negative prediction error signals. From a hierarchical perspective, it has been suggested that sensory input not accounted for by impending predictions is carried from inferior to superior levels (Alexander and Brown, 2018). Our PPI results support this claim by demonstrating increased functional coupling of the right IFG with those stimulus-specific areas showing a positive prediction error. Given our clear a priori hypothesis for activity in both PPA and FFA, we consider the FDR correction to be appropriate and, therefore do not expect FFA activity to reflect false positive activation.

Thus, we found positive coupling of the seed region with right FFA for unexpected faces as well as significant functional coupling of the seed region with the right PPA for unexpected places. We take this finding to reflect that the right IFG receives information from lower level areas, leading to a subsequent revision of the current internal model (Chao et al., 2018; Trempler et al., 2020). Former studies found activity in the IFG not only on occasions of expectation violation (El-Sourani et al., 2019) but more generally upon presentation of important cues (Hampshire et al., 2010, 2009) reflecting the costs of processing this information (Alexander and Brown, 2018; El-Sourani et al., 2019; Keller and Mrsic-Flogel, 2018). Building on these suggestions, our findings indicate that positive prediction errors, even when not significantly pronounced on the univariate level, play an important role during predictive processes and significantly influence the hierarchical processes to keep our internal models updated.

Notably, we found IFG coupling with positive prediction error signals but not for negative ones, implying that unpredicted appearances are driving the coupling with the right IFG but omissions do not. This nicely corroborates previous findings suggesting the IFG to co-vary with the impact that a stimulus has for updating the predictive model (El-Sourani et al., 2019). It is plausible to assume that the non-occurrence of an expected stimulus is in many cases not as informative as the occurrence of an unexpected stimulus: the latter determines the state of the environment whereas the former does only exclude one possible state of the environment.

Taken together, we observed a negative prediction error for the absent expected stimulus as revealed by the univariate ROI analysis, evidence for a positive prediction error for the present unexpected stimulus as implicated by the PPI analyses, and a prediction enhancement effect for expected stimuli. These different analyses and corresponding findings suggest that the error computations in the brain might be carried out by separate prediction error circuits, that process prediction error signals via different mechanisms (Keller and Mrsic-Flogel, 2018; Rao and Ballard, 1999). This assumption may be further supported by the asymmetric engagement of IFG-PPA and IFG-FFA coupling for positive but not negative prediction errors (as outlined above). While previous studies investigated either positive or negative prediction error signals, we obtained evidence that both can occur at the same time. Due to the limited temporal resolution of fMRI, however, we cannot clearly distinguish between predictive processes and processes purely resulting from prediction errors (Kok et al., 2013). Therefore, in future, it would be valuable to investigate the temporal progression from the emergence of expectations and expectation violation using for example electroencephalography.

Overall, the present study provides evidence that positive and negative prediction errors modify our current internal representations to best approximate and predict our ever-changing environment. Moreover, the functional connectivity between positive prediction error signals and the right IFG suggests the crucial involvement of higher-level prefrontal regions during the revision of the current internal model, especially with regard to stimuli defining the present environmental state rather than providing evidence for non-states. These results give new insights into prediction error processing while taking into account the dual nature of expectation violations that involve the unexpected appearance of a stimulus that can be inevitable connected to the unexpected omission of another stimulus.

Data availability statement

The datasets generated and analysed during the current study including additional analyses are available in OSF at https://osf.io/en97c/?view_only=f1d048597c51441bb9e93711ddd7d92a.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Credit authorship contribution statement

Lena M. Schliephake: Conceptualization, Methodology, Investigation, Software, Formal analysis, Writing - review & editing, Visualization. **Ima Trempler:** Formal analysis, Data curation, Writing - review & editing. **Marlen A. Roehe:** Software, Formal analysis, Writing - review & editing, Visualization. **Ricarda I. Schubotz:** Conceptualization, Writing - review & editing, Visualization, Supervision.

Acknowledgements

We would like to thank Monika Mertens, Marie Kleinbielen, and Annika Garlichs for helping to recruit participants and/or collect data. In addition, we would like to thank all the members of the Schubotz Lab at the University of Muenster for the constructive discussions and feedback.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118028.

References

Alexander, W.H., Brown, J.W., 2018. Frontal cortex function as derived from hierarchical predictive coding. Sci. Rep. 8 (1), 1–11. doi:10.1038/s41598-018-21407-9.

Amado, C., Hermann, P., Kovács, P., Grotheer, M., Vidnyánszky, Z., Kovács, G., 2016. The contribution of surprise to the prediction based modulation of fMRI responses. Neuropsychologia 84, 105–112. doi:10.1016/j.neuropsychologia.2016.02.003.

Bentin, S., Allison, T., Puce, A., Perez, E., McCarthy, G., 1996. Electrophysiological studies of face perception in humans. J. Cogn. Neurosci. 8 (6), 551–565. doi:10.1162/jocn.1996.8.6.551.

Blondin, F., Lepage, M., 2005. Decrease and increase in brain activity during visual perceptual priming: an fMRI study on similar but perceptually different complex visual scenes. Neuropsychologia 43 (13), 1887–1900. doi:10.1016/j.neuropsychologia.2005.03.021.

Bukowski, H., Dricot, L., Hanseeuw, B., Rossion, B., 2013. Cerebral lateralization of face-sensitive areas in left-handers: only the FFA does not get it right. Cortex 49 (9), 2583–2589. doi:10.1016/j.cortex.2013.05.002.

Chao, Z.C., Takaura, K., Wang, L., Fujii, N., Dehaene, S., 2018. Large-scale cortical networks for hierarchical prediction and prediction error in the primate brain. Neuron 100 (5), 1252–1266. doi:10.1016/j.neuron.2018.10.004, e3.

Chen, A.J.W., Britton, M., Turner, G.R., Vytlacil, J., Thompson, T.W., D'Esposito, M, 2012. Goal-directed attention alters the tuning of object-based representations in extrastriate cortex. Front. Hum. Neurosci. (6) 1–11. doi:10.3389/fnhum.2012.00187, June 2012.

Clark, A., 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav. Brain Sci. 36 (3), 181–204. doi:10.1017/S0140525x12000477.

den Ouden, H.E.M., Kok, P., de Lange, F.P, 2012. How prediction errors shape perception, attention, and motivation. Front. Psychol. 3, 1–12. doi:10.3389/fpsyg.2012.00548, Dec.

Egner, T., Monti, J.M., Summerfield, C., 2010. Expectation and surprise determine neural population responses in the ventral visual stream. J. Neurosci. 30 (49), 16601–16608. doi:10.1523/JNEUROSCI.2770-10.2010.

El-Sourani, N., Trempler, I., Wurm, M.F., Fink, G.R., Schubotz, R.I., 2019. Predictive impact of contextual objects during action observation: evidence from functional magnetic resonance imaging. J. Cogn. Neurosci. 32 (2), 326–337. doi:10.1162/jocn_a_01480.

Eickhoff B., S., Stephan E., K., Mohlberg, H., Grefkes, C., Fink R., G., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. Neuroimage 25 (4), 1325–1335. doi:10.1016/j.neuroimage.2004.12.034.

Eliades, S.J., Wang, X., 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. Nature 453 (7198), 1102–1106. doi:10.1038/nature06910.

Epstein, R., Kanwisher, N., 1998. The parahippocampal place area: a cortical representation of the local visual environment. NeuroImage 7, 6–9. doi:10.1016/s1053-8119(18)31174-1, 4 Part II.

Esterman, M., Yantis, S., 2010. Perceptual expectation evokes category-selective cortical activity. Cereb. Cortex 20 (5), 1245–1253. doi:10.1093/cercor/bhp188.

Fiser, A., Mahringer, D., Oyibo, H.K., Petersen, A.V., Leinweber, M., Keller, G.B., 2016. Experience-dependent spatial expectations in mouse visual cortex. Nat. Neurosci. 19 (12), 1658–1664. doi:10.1038/nn.4385.

Friston, K., 2005. A theory of cortical responses. Philos. Trans. R. Soc. B: Biol. Sci. 360 (1456), 815–836. doi:10.1098/rstb.2005.1622.

Gardner, M.P.H., Schoenbaum, G., Gershman, S.J, 2018. Rethinking dopamine as generalized prediction error. Proc. R. Soc. B: Biol. Sci. 285 (1891). doi:10.1098/rspb.2018.1645.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. (Regul. Ed.) 10 (1), 14–23. doi:10.1016/j.tics.2005.11.006.

Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. NeuroImage 50 (3), 1313–1319. doi:10.1016/j.neuroimage.2009.12.109.

Hampshire, A., Thompson, R., Duncan, J., Owen, A.M., 2009. Selective tuning of the right inferior frontal gyrus during target detection. Cognit. Affect. Behav. Neurosci. 9 (1), 103–112. doi:10.3758/CABN.9.1.103.

Haxby, J.V., Hoffman, E.A., Gobbini, M.I., 2000. 08_Haxbytics2000. Trends Cogn. Sci. 4, 223–233. doi:10.1016/S1364-6613(00)01482-0, *4*(6)223–233.

Horvitz, J.C., 2000. Mesolimbocortical and nigrostriatal dopamine responses to salient non- reward events. Neuroscience 96 (4), 651–656. doi:10.1016/S0306-4522(00)00019-1.

Ishai, A., Ungerleider, L.G., Martin, A., Haxby, J.V., 2000. The representation of objects in the human occipital and temporal cortex. J. Cogn. Neurosci. 12 (Suppl. 2), 35–51. doi:10.1162/089892900564055.

Keller, G.B., Mrsic-Flogel, T.D., 2018. Predictive processing: a canonical cortical computation. Neuron 100 (2), 424–435. doi:10.1016/j.neuron.2018.10.003.

Kok, P., Failing, M.F., de Lange, F.P., 2014. Prior expectations evoke stimulus templates in the primary visual cortex. J. Cogn. Neurosci. 26 (7), 1546–1554. doi:10.1162/jocn_a_00562.

Kok, P., Brouwer, G.J., van Gerven, M.A.J., de Lange, F.P., 2013. Prior expectations bias sensory representations in visual cortex. J. Neurosci. 33 (41), 16275–16284. doi:10.1523/JNEUROSCI.0742-13.2013.

Krekelberg, B., Boynton, G.M., van Wezel, R.J.A., 2006. Adaptation: from single cells to BOLD signals. Trends Neurosci. 29 (5), 250–256. doi:10.1016/j.tins.2006.02.008.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D.H.J., Hawk, S.T., van Knippenberg, A, 2010. Presentation and validation of the radboud faces database. Cognit. Emot. 24 (8), 1377–1388. doi:10.1080/02699930903485076.

Liu, J., Harris, A., Kanwisher, N., 2002. Stages of processing in face perception: an MEG study. Nat. Neurosci. 5 (9), 910–916. doi:10.1038/nn909.

Lorenz, S., Weiner, K.S., Caspers, J., Mohlberg, H., Schleicher, A., Bludau, S., Amunts, K., 2017. Two new cytoarchitectonic areas on the human mid-fusiform gyrus. Cereb. Cortex 27 (1), 373–385. doi:10.1093/cercor/bhv225, *(New York, N.Y. : 1991)*.

Maunsell, J.H.R., Treue, S, 2006. Feature-based attention in visual cortex. Trends Neurosci. 29 (6), 317–322. doi:10.1016/j.tins.2006.04.001.

McKendrick M., Allison, Jonathan, Denniss, Andrew, Turpin, 2014. Response times across the visual field: Empirical observations and application to threshold determination. Vision Research 101, 1–10. doi:10.1016/j.visres.2014.04.013.

Menegas, W., Babayan, B.M., Uchida, N., Watabe-Uchida, M., 2017. Opposite initialization to novel cues in dopamine signaling in ventral and posterior striatum in mice. Elife 6, 1–26. doi:10.7554/eLife.21886.

Meyer, T., Olson, C.R., 2011. Statistical learning of visual transitions in monkey inferotemporal cortex. In: Proceedings of the National Academy of Sciences of the United States of America, 108, pp. 19401–19406. doi:10.1073/pnas.1112895108.

Miller, E.K., Li, L., Desimone, R., 1991. A neural mechanism for working and recognition memory. Science 254, 1377–1379. NovemberRetrieved from http://dx.doi.org/10.1126/science.1962197 .

O'Craven, K.M., Kanwisher, N., 2000. Mental imagery of faces and places activates corresponding stimulus-specific brain regions. J. Cogn. Neurosci. 12 (6), 1013–1023. doi:10.1162/08989290051137549.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113. doi:10.1016/0028-3932(71)90067-4.

O'Reilly, J.X., Woolrich, M.W., Behrens, T.E., Smith, S.M., Johansen-Berg, H., 2012. Tools of the trade: psychophysiological interactions and functional connectivity. Soc. Cogn. Affect. Neurosci. 7 (5), 604–609. doi:10.1093/scan/nss055.

Poldrack, R.A., 2007. Region of interest analysis for fMRI. Soc. Cogn. Affect. Neurosci. 2 (1), 67–70. doi:10.1093/scan/nsm006.

Posner, M.I., Snyder, C.R., Davidson, B.J., 1980. Attention and the detection of signals. J. Exp. Psychol.: Gen. 109 (2), 160–174. doi:10.1037/0096-3445.109.2.160.

Press, C., Kok, P., Yon, D., 2020. The perceptual prediction paradox. Trends in Cognitive Sciences 24 (1). doi:10.1016/j.tics.2019.11.003.

Press, C., Yon, D., 2019. Perceptual prediction: rapidly making sense of a noisy world. Curr. Biol. 29 (15), R751–R753. doi:10.1016/j.cub.2019.06.054.

Rangarajan, V., Hermes, D., Foster, B.L., Weiner, K.S., Jacques, C., Grill-Spector, K., Parvizi, J., 2014. Electrical stimulation of the left and right human fusiform gyrus causes different effects in conscious face perception. J. Neurosci. 34 (38), 12828–12836. doi:10.1523/JNEUROSCI.0527-14.2014.

Rao, R.P.N., Ballard, D.H, 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nat. Neurosci. 2 (1), 79–87. doi:10.1038/4580.

Rossion, B., 2014. Understanding face perception by means of prosopagnosia and neuroimaging. Front. Biosci. 6 (258), e307.

Schiffer, A.M., Waszak, F., Yeung, N., 2015. The role of prediction and outcomes in adaptive cognitive control. J. Physiol-Paris 109 (1–3), 38–52. doi:10.1016/j.jphysparis.2015.02.001.

Segaert, K., Weber, K., de Lange, F.P., Petersson, K.M., Hagoort, P., 2013. The suppression of repetition enhancement: a review of fMRI studies. Neuropsychologia 51 (1), 59–66. doi:10.1016/j.neuropsychologia.2012.11.006.

Stanley, J., Miall, R.C., 2007. Functional activation in parieto-premotor and visual areas dependent on congruency between hand movement and visual stimuli during motor-visual priming. NeuroImage 34 (1), 290–299. doi:10.1016/j.neuroimage.2006.08.043.

Summerfield, C., Egner, T., 2009. Expectation (and attention) in visual cognition. Trends Cogn. Sci. (Regul. Ed.) 13 (9), 403–409. doi:10.1016/j.tics.2009.06.003.

Trempler, I., Bürkner, P.C., El-Sourani, N., Binder, E., Reker, P., Fink, G.R., Schubotz, R.I., 2020. Impaired context-sensitive adjustment of behaviour in Parkinson's disease pa-

tients tested on and off medication: an fMRI study. NeuroImage 212 (September 2019), 116674. doi:10.1016/j.neuroimage.2020.116674.

Turk-Browne, N.B., Scholl, B.J., Johnson, M.K., Chun, M.M., 2010. Implicit perceptual anticipation triggered by statistical learning. J. Neurosci. 30 (33), 11177–11187. doi:10.1523/JNEUROSCI.0858-10.2010.

Yon, D., Gilbert, S.J., de Lange, F.P., Press, C., 2018. Action sharpens sensory representations of expected outcomes. Nat. Commun. 9 (1), 1–8. doi:10.1038/s41467-018-06752-7.