OXFORD

Co-occurring positive and negative prediction error signals are independent of stimulus repetition effects

Sophie Siestrup [1,2,*,†], Lena Maria Leeners[1,†], Jennifer Pomp[1,2], Marlen A. Roehe[3], Anoushiravan Zahedi [1,2], and
Ricarda I. Schubotz[1,2]

[1]Department of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany
[2]Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Münster, Fliednerstraße 21, 48149 Münster, Germany
[3]Research Unit for Curriculum Development, Teaching Center Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria
*Corresponding author: Sophie Siestrup, Department of Psychology, University of Münster, Fliednerstraße 21, 48149 Münster, Germany.
Email: s.siestrup@uni-muenster.de
†Sophie Siestrup and Lena Maria Leeners contributed equally and share first authorship.

Both the absence of a predicted stimulus and the unexpected presentation of another stimulus result in increased activation in the areas processing the stimuli. These signals are termed negative and positive prediction errors, respectively. Here, we showed that both types of prediction errors can occur simultaneously and independently of stimulus repetition effects. Participants performed a reaction time task in a magnetic resonance scanner while being exposed to face and place stimuli with a distinct probabilistic distribution resulting in unexpected omissions and unexpected presentations of those stimuli. Participants' responses were significantly faster for expected as compared to neutral or unexpected trials, showing that they learned the statistical regularities inherent to the task. Moreover, the region of interest analysis of beta estimates extracted from the fusiform face area and the parahippocampal place area revealed co-occurring negative and positive prediction error signals. This was evidenced by increases in brain activation for unexpected omissions and unexpected presentations of visual stimuli when compared to expected stimuli. Our results also underlined that these effects occur independently of stimulus repetition effects. Altogether, these findings support a predictive coding model of cognition, highlighting the importance of considering the potential dual nature of expectation violations.

Keywords: FFA; fMRI; PPA; prediction error; predictive coding; stimulus repetition.

Introduction

We live in a highly consistent world in which our brains have the extraordinary capacity to extract the statistical regularities in our environment to predict the future. However, the brain also has to revise expectations when the context changes. According to predictive coding models of sensory processing, the sensory cortices operate under a hierarchical structure where higher-level areas send predictions about sensory inputs to lower-level areas, which then compute the difference between predictions and the actual sensory input (Rao and Ballard 1999). A mismatch between a prior expectation and reality is referred to as a prediction error (Den Ouden et al. 2012). To re-estimate and update the predictions, the prediction error is forwarded from lower- to higher-level areas of the processing system (Kiebel et al. 2008). Previous studies have shown that different prediction error signals can be identified in the brain depending on the nature of the expectation violation (Friston 2005; Den Ouden et al. 2009; Keller and Mrsic-Flogel 2018). While the unexpected presentation of a visual stimulus results in the increased bottom–up input in the sensory area processing that unexpectedly appearing stimulus (Egner et al. 2010; Meyer and Olson 2011; Amado et al. 2016), the surprising omission of an expected stimulus has been shown to result in a robust cortical response in the area processing the missing stimulus (Eliades and Wang 2008; Den Ouden et al. 2012; Fiser et al. 2016). These two different kinds of expectation violation signal increase are called positive and negative prediction errors, respectively (Keller and Mrsic-Flogel 2018).

In a recent study, we were the first to scrutinize positive and negative prediction errors in the same experimental paradigm (Schliephake et al. 2021). We used a speeded forced-choice functional magnetic resonance imaging (fMRI) task where unpredicted cross-category stimulus transitions of faces and places would evoke simultaneous activation of the fusiform face area (FFA) and parahippocampal place area (PPA), representing the unexpected omission of a face stimulus (negative prediction error) on the one hand and the unexpected presentation of a place stimulus (positive prediction error) on the other hand. However, this study could not resolve whether these expectation-related effects were independent of stimulus repetition effects. Both the repetition of the same stimulus and the expectation of a specific stimulus because of the probabilistic distribution of stimuli within a trial sequence have been shown to trigger decreases in brain responses (ie Summerfield et al. 2008; Grotheer and Kovács 2015). Until now, the exact neural mechanisms of these effects are still unclear. The repetitive and simultaneously predictable nature of paradigms involving repeating stimuli has been suggested to be partly driven by expectation effects, underlining the importance of a clear distinction between repetition and expectation effects (Summerfield et al. 2008). Previous research revealed mixed findings regarding the potentially interactive effects of stimulus repetition and stimulus expectation, sometimes due to different approaches of experimental designs. In several studies (Summerfield et al. 2008; Grotheer and Kovács 2015; Kronbichler et al. 2018), stimulus transitions were not manipulated

orthogonally to the statistical regularities that were implemented in the task at hand. An unequal stimulus category transition was, for example, more likely than an equal stimulus category transition, and the results were, therefore, inconclusive. While independently manipulating stimulus repetition and expectation, Grotheer and Kovács (2015) identified additive rather than interactive effects of both factors, showing that repetition suppression was more pronounced for expected as compared to unexpected trials and vice versa. However, there is also evidence that repetition and expectation effects may interact (Todorovic et al. 2011), even though they are expressed at different time scales (Todorovic and de Lange 2012).

This latter finding supports the idea that alternating stimulus categories generate lower-level prediction errors that generate larger neural responses when a stimulus is different from its directly preceding neighbor and smaller responses when stimuli are identical (Hosoya et al. 2005). Stimulus expectation effects, on the other hand, might exert effects at higher levels of the processing hierarchy and during later time points. Therefore, hierarchical predictive coding models inspire the hypothesis that stimulus repetition and stimulus expectation effects may be manifestations of prediction errors on different time scales and hierarchy levels (Friston 2005; Kiebel et al. 2008; Wacongne et al. 2012). During the first stages of cortical processing, alternating stimulus categories would generate sensory prediction errors, while higher-order expectations that are based on more complex statistical regularities may shape subsequent neural processing of the stimulus, giving rise to higher-order contextual prediction errors (Todorovic and de Lange 2012).

The aim of the current fMRI study was to investigate positive and negative prediction errors resulting from higher-order implicit expectation violations while controlling for the effects of lower-level stimulus repetition effects. We used an adapted version of the paradigm of our previous study (Schliephake et al. 2021), employing three different block conditions in a speeded forced choice task where participants categorized two different types of stimuli. As in the previous study, we employed face and place stimuli that have been shown to differentially activate FFA and PPA, respectively (Epstein and Kanwisher 1998; Haxby et al. 2000). Introducing two distinct stimulus categories enabled us to investigate brain responses resulting from an omitted stimulus category (eg a face) and, at the same time, looking at brain activation resulting from an unexpectedly presented stimulus from the other category (eg a place). In order to control for stimulus repetition effects, each block condition had a different distribution of stimulus category transitions. In the first block condition, unequal transitions (ie face–place; place–face) were more likely than equal transitions (ie face–face; place–place). The opposite pattern was used in the second block condition. Lastly, we employed a third block condition where equal and unequal transitions were equally likely since we wanted to include a "neutral" condition in which no biased expectations were induced (Amado et al. 2016).

On the behavioral level, we expected that participants would implicitly learn the regularities implemented in the task showing decreased response times for expected as compared to either neutral or unexpected stimuli. Additionally, we expected the percentage of correct responses (CRs) to increase for expected stimuli when compared to either neutral or unexpected stimuli. Moreover, we hypothesized to find positive evidence for an independence of the factor expectation and stimulus repetition.

On the neural level, we hypothesized to find both positive and negative prediction error signals for unexpected events. On the one hand, our hypothesis posited the occurrence of a positive prediction error when a stimulus category appeared unexpectedly (eg a face) as compared to the expected stimulus category (eg a place), reflected by increased brain activation in the region preferentially processing the unexpected stimulus category (here FFA). On the other hand, we expected the simultaneous manifestation of a negative prediction error, indicated by an increased blood oxygenation level–dependent (BOLD) signal within the brain area processing the unexpectedly omitted stimulus category (here PPA), thereby representing the absence of a preferentially processed expected stimulus category (eg a place). Specifically, when considering face stimuli, "preferred" activity should be observed in the FFA, and "non-preferred" in the PPA. Likewise, for place stimuli, "preferred" activity was expected in the PPA and "non-preferred" activity in the FFA. Additionally, we hypothesized that the same pattern would hold if we compare unexpected with neutral events since activation should increase for unexpected events and decrease for expected events. Importantly, we also expected to find positive evidence for the null hypothesis that the two factors' stimulus repetition and expectation do not interact.

Materials and methods
Participants

Thirty-three volunteers (25 identified as women, 8 identified as men) participated in the experiment. All participants were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield 1971). One of them was excluded because of excessive movement during scanning (>4 mm). Two further participants were excluded after initial inspection of the behavioral data due to low performance levels with an accuracy of more than 2 SDs below the mean accuracy level of all participants. Additionally, one participant had to be excluded as they did now show any significant PPA activity in the localizer. The mean age of the remaining 29 participants was 23.31 years (SD = 4.07 years), and the sample included 22 women and 7 men. All participants had normal or corrected-to-normal vision and reported no prior history of psychiatric or neurological disorders. They provided written informed consent to all procedures and data usage before the study started. The experimental procedures were approved by the ethics committee at the University of Münster (approval ID: 2017-2031-MR).

Stimuli

We used 36 digital photographs that consisted of 18 face images (9 female, 9 male) and 18 place images (8 indoor, 8 outdoor). The face images were drawn from the Radboud Faces Database (Langner et al. 2010), and the place images were collected from the internet. The place stimuli were cropped and resized in order to match the face stimulus set using the GNU Image Manipulation Program (GIMP, https://gimp.org, freeware). All photographs were displayed in the middle of a uniform gray background at a visual angle of approximately 4.20° × 6.17°.

Trial sequence and task design

The stimulus sequence was programmed using Matlab (Version R2022a; The MathWorks Inc., Natick, MA, USA). The Presentation Software (Version 19.0, Neurobehavioral Systems, Inc., Berkeley, CA) was used to present the stimuli and to record participants' behavioral responses. The paradigm used was an adapted version of the paradigm employed in Schliephake et al. (2021). For each participant, the 36 stimuli were randomly assigned to three different stimulus sets (blocks). In all three sets, 12 images including three photographs of each category (female and male faces,

indoor and outdoor places) were repeated 10 times, which resulted in 120 trials per block. Further, we manipulated the transitional probabilities of the stimuli to introduce statistical regularities. The stimulus transitions used in the current paradigm were either within- or between-category, meaning that either a face (place) followed another face (place) stimulus (within-category transition) or a face (place) followed a place (face) stimulus (between-category transition) or vice versa.

In three different block conditions, we varied the probability of these transitions. In the "unequal dominant" block, 12 images were shown in random succession with 70% of the transitions being between-category and 30% of the transitions being within-category. Therefore, in this block, the between-category transitions were the expected trials and the within-category transitions the unexpected ones. The "equal dominant" block consisted of another set of 12 images. Here, 70% of the transitions were within-category and expected and 30% between-category and unexpected. The third block was used as a control condition, including 50% within-category and 50% between-category transitions with yet again another set of images (Fig. 1). Because within- and between-category transitions were equally likely, these trials represented the neutral condition in which no clear expectations about the upcoming stimulus category could be made. Introducing the unequal and equal dominant conditions, we ensured that participants would be able to implicitly predict the most likely upcoming stimulus category when presented with a given stimulus.

Task

Participants were told that we were investigating how the brain responds to different types of images. This cover story was used to ensure that the transitional probabilities were only learned implicitly. During the main task in the MR scanner, stimuli were centrally back-projected onto a screen. Participants viewed the screen through a mirror above the head coil. All photographs were displayed for 350 ms each followed by a jittered intertrial interval (ITI) of 3,500, 4,000, 4,500, or 5,000 ms during which a fixation cross was presented. To indicate whether the depicted image was a face or place, participants were asked to respond as quickly and accurately as possible by pressing one of two response keys with their right index or middle finger. The stimulus–response mapping was counterbalanced across participants. The task consisted of five experimental blocks including two repetitions of the "unequal-dominant" and "equal-dominant" block conditions and one "neutral" block. The order of the block conditions was counterbalanced across participants so that half of the participants started the experiment with "unequal-dominant" block and the other half started with the "equal-dominant" block. The neutral block was always the third block. After each block, a short break of 8 s was introduced informing the participants that a block had ended and introducing the start of the new block at the end of the break. All in all, the task included 600 trials and lasted around 47 min.

To define the individual anatomical location of the FFA and the PPA, we let participants subsequently perform a localizer task (Kanwisher et al. 1997). The localizer consisted of a 1-back task during which face and place stimuli were presented block-wise on a gray background. Participants were asked to press the response button whenever two identical stimuli were presented in a row. Each block consisted of 15 stimuli (including one to two repetitions), with each stimulus presented for 750 ms followed by 250-ms fixation. Between the blocks, there was a short break with 10 s fixation. The task consisted of 12 blocks with alternating face and place blocks.

FMRI data acquisition

Whole-brain MRI data were acquired using a 3 T Siemens Magnetom Prisma scanner (Siemens, Erlangen, Germany) using a 20-channel head coil. A 3D-multiplanar rapidly acquired gradient-echo (MPRAGE) sequence was employed to obtain high resolution T1-weighted images before functional scanning, with isotropic voxel size ($1 \times 1 \times 1$ mm) in a 256 mm field of view (FOV) (256×256 matrix, 196 slices, repetition time (TR) = 2,130 ms, echo time (TE) = 2.28, slice thickness of 1 mm and a flip angle of 8°). Functional BOLD images were acquired parallel to the anterior commissure/posterior commissure line using a T2*-weighted gradient echo planar imaging sequence (64×64 matrix, 192 mm FOV, 90° flip angle, TR = 2,000 ms, TE = 30 ms). Each volume consisted of 33 adjacent axial slices with 3 mm slice thickness and a gap of 1 mm, resulting in a voxel size of $3 \times 3 \times 4$ mm. All visual images were projected on a screen behind the scanner bore. Stimuli were presented in the center of the field of vision using a video projector. Participants were able to view the screen via a mirror that was fixated on top of the head coil and adjusted for each participant to provide a good view of the entire screen. To record behavioral responses, a response box was positioned on the right thigh of the participants so that they could place their right index and middle finger on two response buttons. To minimize motion artifacts, foam paddings were applied around the participants' head; whereas earplugs as well as noise-canceling headphones were provided to reduce scanner noise.

Data analysis
Behavioral data analysis

The behavioral data analysis was conducted using RStudio (2024 September 1, R Core Team 2024). The focus of the analysis was on response times and accuracy (percentage of CRs), which were measured to examine participants' performance in the task. For the analysis, trials without a response and trials with incorrect responses were categorized as error trials. To establish a time frame for valid responses, a maximum response window of 1,500 ms was defined, starting from trial onset. Any responses recorded after this phase were considered error trials as well. Task performance was determined by calculating the mean percentage of correctly answered trials of all correct and error trials per participant and condition. For the analysis of response times, we only considered correct responses according to the above-described criteria and calculated mean response times (RTs) per participant and condition. We confirmed the normality of RTs using the Shapiro–Wilk test ($P > 0.5$).

For each participant, RTs and CR were entered as dependent variable into two 2×3 repeated-measures ANOVA with the factors: Transition (equal, unequal) and Expectation (expected, neutral, unexpected). We report *P*-values for the two main effects as well as their interaction. In the case of nonsphericity, we report Greenhouse–Geisser-corrected *F*- and *P*-values. Afterward, planned one-tailed *t*-tests (Bonferroni–Holm-corrected) evaluate the expectation effect to assess whether participants possessed a robust knowledge of the contextual regularities of the task that would be evident in decreased RTs and increased CRs for expected as compared to unexpected and neutral trials.

fMRI data analysis

All preprocessing and statistical analyses were carried out with SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) and custom Matlab scripts (Version R2020b; The MathWorks Inc., Natick, MA, USA). All functional images were slice time–corrected
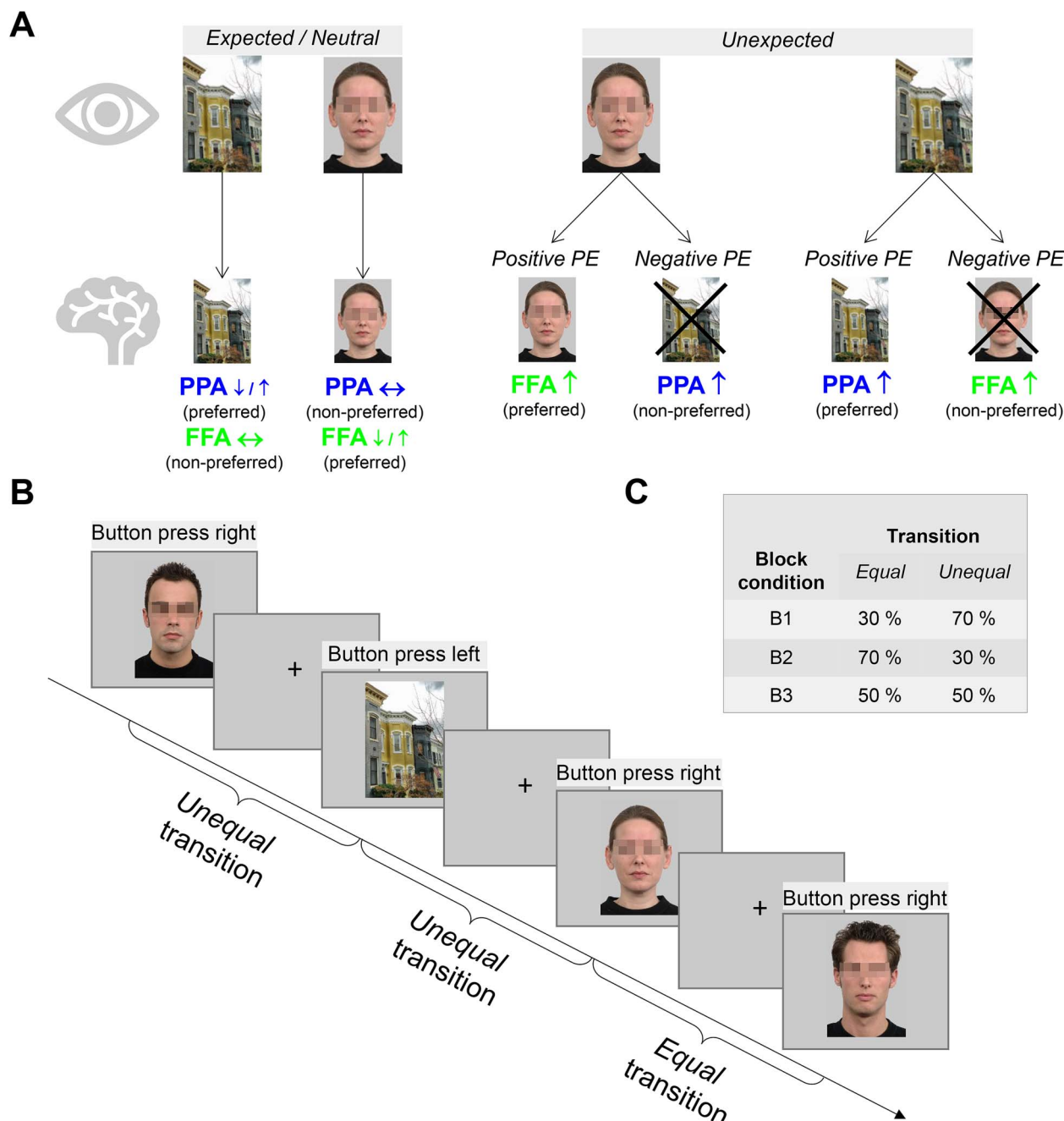
Fig. 1. Task design. A) General hypothesis. When a neutral or expected stimulus is presented, the brain region that preferentially processes this type of stimulus (FFA for faces, PPA for places) will be active or even deactivated for expected stimuli. For unexpected stimuli of a specific type, the brain area that preferentially processes this stimulus type will be even more active due to the positive prediction error. At the same time, a different type of stimulus is unexpectedly omitted, leading to increased activation in the brain area that preferentially processes stimuli of the omitted category (ie the brain region that non-preferentially processes the presented stimulus) due to the negative prediction error. B) Diagram of the experimental task. Each image was presented for 350 ms, and the interstimulus interval (ITI) was jittered between 3 and 5 s. For each trial, participants had to indicate the stimulus category (face or place) via a button press (left or right) and C) an overview of the three block conditions with their corresponding transitional probabilities. The block order was counterbalanced across participants so that half of the participants started the experiment with B1 and the other half started with B2. B3 was always the third block. This resulted in two different block orders: B1, B2, B3, B1, B2 or B2, B1, B3, B2, B1. Please note that the eyes in the example stimuli were blurred here due to privacy reasons, but not in the original stimulus material. Face images were drawn from the Radboud Faces Database (Langner et al. 2010).

and spatially realigned to the mean image. Structural images were coregistered to the mean functional image and then segmented into native space tissue components. Then, functional images (resampled at 3 mm³) were spatially normalized to the standard Montreal Neurological Institute template using tissue-segmented T1-weighted anatomical scans. Normalized functional images as well as those in individual subject space were spatially

smoothed with a Gaussian kernel of full width half-maximum of 6 mm³.

To identify FFA and PPA activity, we used anatomical masks of fusiform gyrus (FG) and parahippocampal gyrus (PHG) to restrict our analyses. These masks were created by extracting the respective regions from a probabilistic atlas (Amunts et al. 2020; Caspers et al. 2012; Caspers et al. 2019; Lorenz et al. 2015;

Lorenz et al. 2021; Stenger et al. 2022a, 2022b, 2022c, 2022d) and implementing a threshold of >0.2. As in our previous study (Schliephake et al. 2021), we decided to only include the right FFA and right PPA in our analyses since it has been shown that face images are dominantly processed by the right hemisphere (Bentin et al. 1996; Rossion 2014).

To then obtain FFA and PPA regions of interest (ROIs) for the main task, we convolved the block regressors coding for onsets and durations of face and place blocks of the localizer task with the canonical hemodynamic response function (HRF). This model was computed both in individual subject space and normalized space. For each participant, we contrasted face versus place blocks for FFA activation and place versus face blocks for PPA activation. As a first step, all participants' faces > places and places > faces contrasts (in individual space) were inspected for significantly activated voxels within our predefined anatomical regions (FG and PHG, respectively), using a threshold of $P < 0.05$ (uncorrected). One participant who did not show activation for place stimuli in the PHG was excluded from further analyses. Data from the remaining 29 participants were entered into a second-level one-sample t-test (normalized space) and faces > places and places > faces contrasts were obtained on group level. Significant activation for face and place stimuli within the predefined anatomical regions was determined using a threshold of $P < 0.05$ (uncorrected). Activation clusters were extracted and inverse-normalized to each subject's individual space to serve as ROIs for the analysis of the main task.

The statistical model of the main task (computed in individual subject space) included 12 regressors coding for onset time and duration (0 s) of each trial of each experimental condition in our 2 (Transition: equal, unequal) × 3 (Expectation: expected, neutral, unexpected) repeated measures factorial design, separately for face and place stimuli. Additionally, six regressors of no interest for the motion parameters (three translations and three rotations) were entered to the model. As the first trial of each block was by definition neutral, ie no expectation could have been formed for this trial beforehand, we excluded these five trials from the analysis, as well as trials with incorrect or missing answers. The stimulus functions were then convolved with a canonical HRF and regressed against the observed fMRI data to form our condition regressors.

Employing two different stimulus categories, we could investigate positive and negative prediction errors within two different category-specific areas, FFA and PPA. For unexpected as compared to neutral stimuli, we expected a positive prediction error that would be reflected in a BOLD activation increase in the brain region preferentially processing the stimulus category at hand. At the same time, we suggested that a negative prediction error would result in increased BOLD activation in the area that preferentially processed the omitted stimulus category (ie non-preferentially processed the presented category). As such, activation elicited by a face stimulus represents "preferred" beta activity in the FFA, whereas activation in the PPA signifies "non-preferred" beta activity and vice versa. This coding was adapted from our previous study (Schliephake et al. 2021).

The obtained ROIs were then used to extract data estimates (mean beta parameters) of the activity associated with each condition in the task from each individual subject. Subsequently, beta values of our two ROIs were subjected to a 2 × 3 × 2 repeated-measures ANOVA with the factors: Transition (equal, unequal), Expectation (expected, neutral, unexpected), and Preference (preferred, non-preferred). We report P-values for the three main effects and their interactions. To assess whether

unexpected stimuli modulated brain responses in the preferred and non-preferred trials, we implemented planned post hoc contrasts as Bonferroni–Holm-corrected paired one-sided t-tests of the conditions unexpected_preferred vs. neutral_preferred, unexpected_preferred vs. expected_preferred, unexpected_non-preferred vs. neutral_non-preferred, and unexpected_non-preferred vs. expected_non-preferred.

Bayesian analysis

In order to test whether nonsignificant results provided evidence for the corresponding null hypotheses, we conducted Bayesian analyses. In summary, the Jeffreys-Zellner-Siow (JZS) Bayes (Rouder et al. 2009) implemented in JASP 0.18.1 (Love et al. 2019) (www.jasp-stats.org; RRID:SCR_015823) with default parameters (Cauchy prior width of 0.707 for t-tests, uniform priors, and a random seed of 1 for ANOVAs) was employed for Bayesian tests. The interpretations of resultant Bayes factors, which quantify evidence supporting the null hypotheses ($B01$), adhered to established classification criteria (Lee and Wagenmakers 2013).

Results
Behavioral data

In line with our hypotheses, the results showed a significant main effect of expectation [$F_{(1.65,46.06)} = 6.53$, $P = 0.005$] and no significant interaction effect between expectation and transition [$F_{(1.46,40.83)} = 1.13$, $P = 0.318$, $BF_{01} = 3.592$] for the RT data. The main effect of transition was nonsignificant [$F_{(1,28)} = 3.2$, $P = 0.084$, $BF_{01} = 1.671$]. The same pattern was observed for the CR data that revealed a main effect of expectation [$F_{(2,56)} = 20.51$, $P < 0.001$] as well as a nonsignificant interaction [$F_{(2,56)} = 1.85$, $P = 0.167$, $BF_{01} = 1.608$] and a nonsignificant transition main effect [$F_{(1,28)} = 3.32$, $P = 0.079$, $BF_{01} = 1.471$].

To further examine whether participants implicitly learned the transitional probabilities within and between stimulus categories, we employed paired-sample t-tests for expected category stimuli as compared to neutral and unexpected ones. Indeed, as for the RT data, participants were significantly faster for expected events when compared to unexpected events [$t_{(28)} = -4.23$, $P < 0.001$] and neutral events [$t_{(28)} = -2.87$, $P = 0.008$]. Regarding CR, participants showed elevated accuracy levels for expected as compared to unexpected [$t_{(29)} = 6.14$, $P < 0.001$] and neutral events [$t_{(28)} = 3.19$, $P = 0.004$)](see Fig. 2).

fMRI data

The aggregated FFA and PPA BOLD responses revealed no significant main effects of transition [$F_{(1,28)} = 0.28$, $P = 0.604$, $BF_{01} = 3.752$] or expectation [$F_{(1.05,29.35)} = 0.26$, $P = 0.624$, $BF_{01} > 999$] but a significant main effect of preference [$F_{(1,28)} = 235.84$, $P < 0.001$]. As hypothesized, the results provided evidence for the null hypothesis that there is no interaction between transition and expectation [$F_{(1.11,31)} = 0.59$, $P = 0.466$, $BF_{01} = 106.425$]. In contrast, the effect of transition was dependent on whether the presented stimulus was a region's preferred or nonpreferred stimulus as revealed by a significant two-way interaction between transition and preference [$F_{(1,28)} = 61.80$, $P < 0.001$]. Post hoc paired t-test revealed significantly higher beta values for preferred compared to non-preferred stimuli for equal [$t_{(28)} = 13.34$, $P < 0.001$] and unequal [$t_{(28)} = 14.52$, $P < 0.001$] transitions. However, for preferred stimuli, beta values were higher for unequal vs. equal transitions [$t_{(28)} = -4.15$, $P < 0.001$] and the other way around for non-preferred ones [$t_{(28)} = 3.94$, $P < 0.001$].
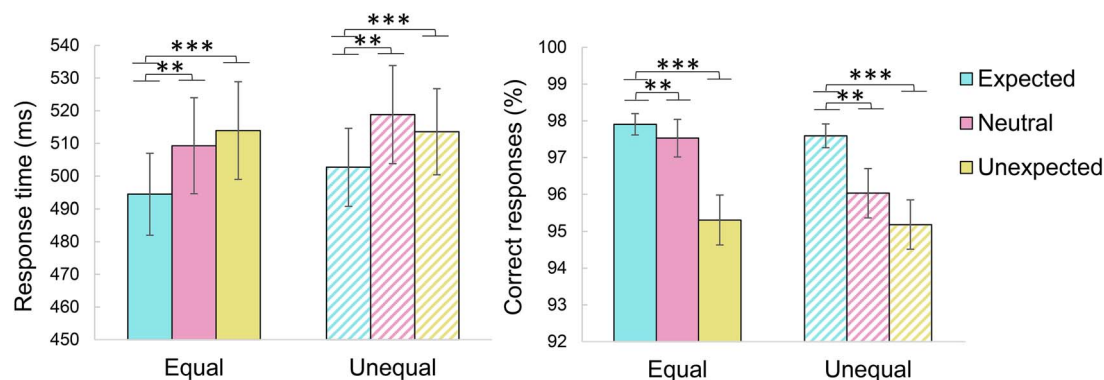
Fig. 2. Behavioral data. Mean response times and mean percentage of correct responses for all experimental conditions. Error bars show standard errors of the means. ** = P < 0.01; *** = P < 0.001.

There was no significant interaction between expectation and preference [$F_{(2,56)} = 2.35$, $P = 0.105$, $BF_{01} = 11.780$]. Additionally, the frequentist calculation of the three-way interaction among expectation, transition, and preference was found to be nonsignificant [$F_{(1,29)} = 1.77$, $P = 0.188$]. The Bayesian equivalent to the ANOVA that was used to better understand nonsignificant effects (see Bayesian Analysis), however, revealed positive evidence in favor of the alternative hypothesis that there is a three-way interaction between the three factors ($BF_{01} < 0.001$). Taken together, it remains unclear whether the combined effect of these factors deviates significantly from what might be predicted based on their individual two-way interactive effects.

To see whether the presence of an unexpected image would robustly increase activation in our ROIs when compared to neutral or expected images, we conducted planned paired-sample t-tests. Regarding a positive prediction error, we expected increased beta values for unexpected events within a stimulus' preferred region when compared to either neutral or expected events. A negative prediction error should be represented by increased activation for unexpected events in the region non-preferentially processing a presented stimulus as compared to expected and neutral events. Indeed, the results showed a significant difference between expected and unexpected events for both preferred [$t_{(28)} = 2.33$, $P = 0.041$] and non-preferred [$t_{(28)} = 2.48$, $P = 0.038$] stimuli.

The frequentist t-tests comparing the unexpected and neutral conditions did not reach significance for the preferred [$t_{(28)} = 0.27$, $P = 0.40$] as well as the non-preferred condition [$t_{(28)} = 0.77$, $P = 0.45$]. To further examine the evidence, a Bayesian t-test was conducted, revealing a Bayes factor $BF_{10} = 0.245$ for the preferred and $BF_{10} = 0.396$ for the non-preferred condition, suggesting positive evidence for the null hypothesis of no difference between unexpected and neutral trials (see Fig. 3).

Discussion

To investigate co-occurring positive and negative prediction error signals, we employed an experimental paradigm that was able to dissociate between the BOLD response of unexpectedly appearing stimuli from one category (eg a face) and unexpected omissions from another category (eg a place). Participants implicitly learned the varying probability of stimulus category transitions that were either equal (eg face–face) or unequal (eg face–place). We ensured that the stimulus category transitions were counterbalanced throughout the experiment to distinguish between increases in brain activations due to unequal stimulus category transitions and violated expectations, ie prediction errors.

The behavioral results underline that participants developed implicit predictions that were dependent on whether it was more likely that equal or unequal stimulus categories were following each other. We found faster responses to expected as compared to unexpected and neutral stimuli as well as higher accuracy levels for predicted stimuli suggesting that participants implicitly learned the statistical regularities inherent in the task to optimize performance (Esterman and Yantis 2010; Turk-Browne et al. 2010).

This finding was accompanied by increased BOLD responses for unexpected compared to expected stimuli. Importantly, we found evidence that these effects were independent of the effect of the stimulus repetitions. In most previous studies, stimulus repetitions and stimulus expectation were not independent given that repetitions tended to occur more frequently when they were expected and less frequently when they were unexpected (Summerfield et al. 2008; Summerfeld et al. 2011; Kovács et al. 2012). Even though these studies found significant increases for unexpected trials (ie decreases for expected trials), they could not show that these effects were independent of the stimulus repetitions and alternations. In other studies using associative cues to investigate prediction errors, the stimulus order was often completely random (eg Alilovic et al. 2019) or balanced (eg Kok and Turk-Browne 2018) and irrelevant for analyses, so concrete conclusions about the interplay of stimulus expectation and repetition are often not available. However, this distinction is important since both factors have been shown to have similar neuronal consequences (Henson 2003; Todorovic and de Lange 2012). The current study addressed this issue by presenting two different stimulus categories that either repeated or alternated while warranting an equal repetition and alternation probability of 50% each. Importantly, we found that the effects of repetition and expectation were consistently additive rather than interactive in our ROIs. This suggests that the impact of stimulus alternation and violated expectations can be differentiated from each other in terms of their effects on the neural responses (Tang et al. 2018).

Moreover, the results are compatible with predictive coding models that suggest that predictive processes are organized in multiple hierarchical stages. A study by Wacongne et al (2011), for example, introduced an auditory novelty paradigm showing (i) low-level expectations that were based on local transitions and (ii) higher-level expectation processes based on the overall probability of the stimulation. Hence, a two-stage model of neuronal response on expectation and expectation violation has been suggested by Grotheer and Kovács (2015). On the one hand, low-level mechanisms in posterior cortical areas compute early prediction errors, leading to increases in neuronal responses for alternated stimuli as compared to repeating stimuli. On the other hand,
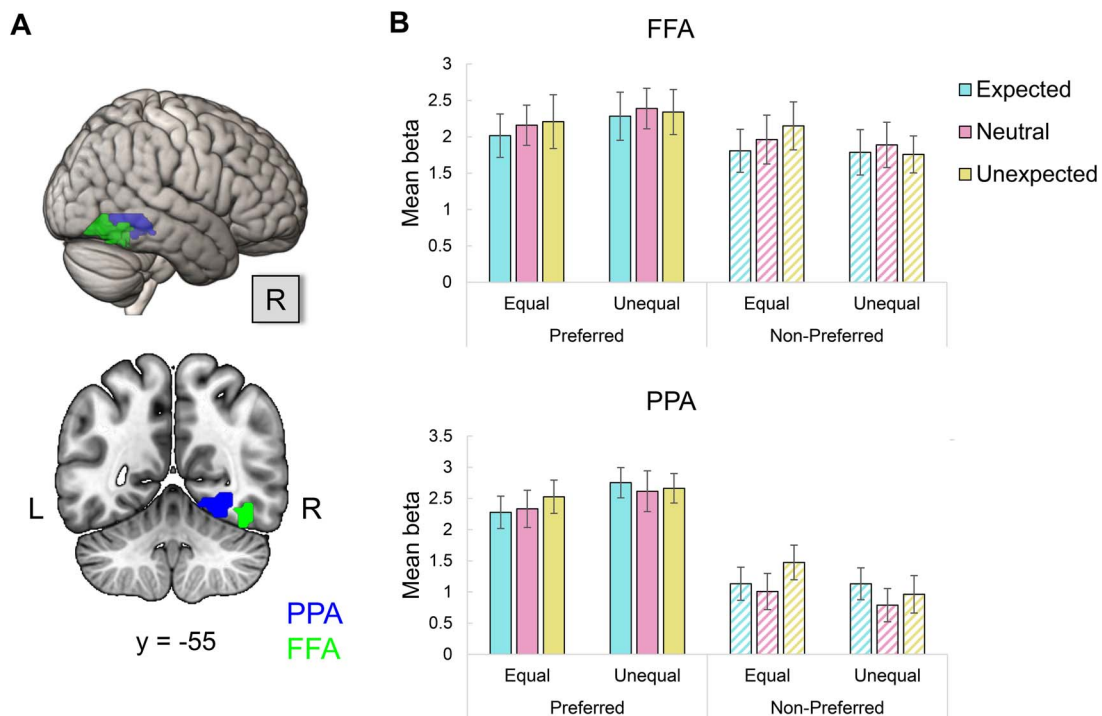
Fig. 3. ROI fMRI data. A) Right FFA and PPA ROIs used for beta value extraction. Depicted are the active clusters from the localizer task that responded to face and place stimuli, within predefined anatomical masks of fusiform and parahippocampal gyrus at $P < 0.05$ (uncorrected). Please note that for beta extraction, ROIs were reverse-normalized to each subject's individual space. B) Mean individual activation estimates (beta parameters +/− standard error of the mean) for each condition of the experiment. Please note that for completeness, we show beta values for FFA (top) and PPA (bottom) separately, while they were aggregated for the statistical analysis.

feedback connections from higher-level areas are responsible for computations of violated expectations which take more time, occur later in time, and are thought to originate in prefrontal (Summerfield et al. 2006) and inferior frontal brain regions (Wig et al. 2005; Horner and Henson 2012; Ferrari et al. 2022) and the hippocampus (Hindy et al. 2016; Kok and Turk-Browne 2018). Please note, however, that the present study cannot make conclusions about the precise timing of prediction error processing, for which a modality with higher time resolution, like electroencephalography (EEG), would be required.

With regard to our hypothesis on co-occurring positive and negative prediction errors, the results of our fMRI analyses replicate the findings from our previous study (Schliephake et al. 2021). Moreover, we provide evidence for positive prediction errors, represented by increases in brain activation (eg in the FFA) for unexpectedly appearing stimuli as compared to expected stimuli, and negative prediction errors, represented by elevated activation levels (eg in the PPA) for unexpectedly omitted stimuli as compared to expected stimuli. The positive prediction error effects in the current study are especially interesting as previous research in this field consistently implemented paradigms during which expected stimuli were withheld and robust cortical responses to such surprising stimulus omissions were found in the relevant cortical area (Den Ouden et al. 2009; Todorovic and de Lange 2012; Wacongne et al. 2012). In the current study, however, there was no blank period but the unexpectedly omitted stimulus was replaced by another unexpected stimulus. Yet, we found similar cortical responses representing the omitted stimulus as studies employing real omissions. This finding supports the idea that increased activation in the area processing the omitted stimulus category does not depend on a stimulus-free time window during which the stimulus was expected but rather on the experience

that the expected stimulus category did not appear. In our previous study (Schliephake et al. 2021), we only found the positive prediction error effect in a PPI analysis but not a basic ROI analysis in FFA and PPA. The difference between our two studies could be driven by the fact that in the previous study, the most likely stimulus transition was a change in stimulus category, which we explicitly controlled in the present work as it was central to our research question.

Another important factor that has been considered in the current study is the introduction of a neutral stimulus category in which the probability of equal and unequal stimulus category transitions was 50% each and participants could not form strong expectations about which stimulus category would appear next (Grotheer and Kovács 2015; Den Ouden et al. 2023). In the past, studies compared brain activation between unexpected and expected events to assess elevated activation levels attributed to unexpected events. However, a limitation inherent in the comparison of unexpected and expected events lies in its inability to differentiate whether the observed effect signifies an increase in activation for unexpected stimuli or a decrease for expected ones. The finding that activation in the neutral condition was not significantly different from activation in the unexpected condition, and instead situated in between activation for expected and unexpected stimuli, hints at both a decrease in activation for expected events and an increase for unexpected ones.

We expected to show that prediction error signals represented by activation increases would still be evident when comparing unexpected to neutral trials. However, this contrast did not reach significance for both unexpected omissions and unexpected presentations of stimuli. It is plausible that the difference between the neutral and unexpected condition that can be observed descriptively in our results was too small to induce significant

activation differences. This is because the probability of a specific stimulus category transition was 50% in the neutral condition and 30% in the unexpected condition. In order to strengthen the evidence, more data are needed in the future.

To further advance our understanding of the interplay between positive and negative prediction error signals, future studies could employ advanced neuroimaging techniques, such as high-density EEG or multimodal imaging approaches. These methods can capture the temporal dynamics and network-level interactions associated with these prediction error signals.

In conclusion, the investigation of co-occurring positive and negative prediction errors is essential for advancing our understanding of sensory processing, perception, and learning. By studying these prediction errors simultaneously, we gain valuable insights into the dynamic interplay between expectations and sensory input, enriching our knowledge of the underlying mechanisms. Our study showed for the first time how to disentangle simultaneous positive and negative visual prediction errors and future research will extend this differentiation for other domains.

References

Alilovic J, Timmermanns B, Reteig LC, van Gaal S, Slagter HA. 2019. No evidence that predictions and attention modulate the first feedforward sweep of cortical information processing. *Cereb Cortex*. 29:2261–2278. https://doi.org/10.1093/cercor/bhz038.

Amado C et al. 2016. The contribution of surprise to the prediction based modulation of fMRI responses. *Neuropsychologia*. 84: 105–112. https://doi.org/10.1016/j.neuropsychologia.2016.02.003.

Amunts K, Mohlberg H, Bludau S, Zilles K. 2020. Julich-brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*. 369:988–992. https://doi.org/10.1126/science.abb4588.

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

Caspers J et al. 2012. Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Struct Funct*. 218:511–526. https://doi.org/10.1007/s00429-012-0411-8.

Caspers J, Zilles K, Eickhoff SB, Schleicher A, Mohlberg H, Amunts K. 2019. Probabilistic cytoarchitectonic map of area FG2 (FusG) (v1.4) [data set]. *Human Brain Project Neuroinformatics Platform*. https://doi.org/10.25493/F2JH-KVV.

Den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE. 2009. A dual role for prediction error in associative learning. *Cereb Cortex*. 19:1175–1185. https://doi.org/10.1093/cercor/bhn161.

Den Ouden HEM, Kok P, de Lange FP. 2012. How prediction errors shape perception, attention, and motivation. *Front Psychol*. 3:1–12. https://doi.org/10.3389/fpsyg.2012.00548.

Den Ouden C et al. 2023. Stimulus expectations do not modulate visual event-related potentials in probabilistic cueing designs. *NeuroImage*. 280:120347. https://doi.org/10.1016/j.neuroimage.2023.120347.

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

Eliades SJ, Wang X. 2008. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 453:1102–1106. https://doi.org/10.1038/nature06910.

Epstein R, Kanwisher N. 1998. A cortical representation of the local visual environment. *Nature*. 392:598–601. https://doi.org/10.1038/33402.

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

Ferrari A, Richter D, de Lange FP. 2022. Updating contextual sensory expectations for adaptive behavior. *J Neurosci*. 42:8855–8869. https://doi.org/10.1523/JNEUROSCI.1107-22.2022.

Fiser A et al. 2016. Experience-dependent spatial expectations in mouse visual cortex. *Nat Neurosci*. 19:1658–1664. https://doi.org/10.1038/nn.4385.

Friston K. 2005. A theory of cortical responses. *Philos Trans R Soc B Biol Sci*. 360:815–836. https://doi.org/10.1098/rstb.2005.1622.

Grotheer M, Kovács G. 2015. The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia*. 67:175–182. https://doi.org/10.1016/j.neuropsychologia.2014.12.017.

Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci*. 4:223–233. https://doi.org/10.1016/S1364-6613(00)01482-0.

Henson RNA. 2003. Neuroimaging studies of priming. *Prog Neurobiol*. 70:53–81. https://doi.org/10.1016/S0301-0082(03)00086-8.

Hindy NC, Ng FY, Turk-Browne NB. 2016. Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat Neurosci*. 19:665–667. https://doi.org/10.1038/nn.4284.

Horner AJ, Henson RN. 2012. Incongruent abstract stimulus–response bindings result in response interference: FMRI and EEG evidence from visual object classification priming. *J Cogn Neurosci*. 24:760–773. https://doi.org/10.1162/jocn_a_00163.

Hosoya T, Baccus SA, Meister M. 2005. Dynamic predictive coding by the retina. *Nature*. 436:71–77. https://doi.org/10.1038/nature03689.

Kanwisher N, McDermott J, Chun MM. 1997. The Fusiform Face Area: A module in human extrastriate cortex specialized for

face perception. *J Neurosci*. 17:4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997.

Keller GB, Mrsic-Flogel TD. 2018. Predictive processing: a canonical cortical computation. *Neuron*. 100:424–435. https://doi.org/10.1016/j.neuron.2018.10.003.

Kiebel SJ, Daunizeau J, Friston KJ. 2008. A hierarchy of time-scales and the brain. *PLoS Comput Biol*. 4:e1000209. https://doi.org/10.1371/journal.pcbi.1000209.

Kok P, Turk-Browne NB. 2018. Associative prediction of visual shape in the hippocampus. *J Neurosci*. 38:6888–6899. https://doi.org/10.1523/JNEUROSCI.0163-18.2018.

Kovács G, Iffland L, Vidnyánszky Z, Greenlee MW. 2012. Stimulus repetition probability effects on repetition suppression are position invariant for faces. *NeuroImage*. 60:2128–2135. https://doi.org/10.1016/j.neuroimage.2012.02.038.

Kronbichler L, Said-Yürekli S, Kronbichler M. 2018. Perceptual expectations of object stimuli modulate repetition suppression in a delayed repetition design. *Sci Rep*. 8:1–8. https://doi.org/10.1038/s41598-018-31091-4.

Langner O et al. 2010. Presentation and validation of the radboud faces database. *Cogn Emot*. 24:1377–1388. https://doi.org/10.1080/02699930903485076.

Lee MD, Wagenmakers E-J. 2013. Bayesian cognitive modeling A practical course. Cambridge University Press, pp. 101–117. https://doi.org/10.1017/CBO9781139087759.009.

Lorenz S et al. 2015. Two new cytoarchitectonic areas on the human mid-fusiform gyrus. *Cereb Cortex*. 27:373–385. https://doi.org/10.1093/cercor/bhv225.

Lorenz S et al. 2021. Probabilistic cytoarchitectonic map of area FG4 (FusG) (v7.2) [data set]. *EBRAINS*. https://doi.org/10.25493/MTWF-X4V.

Love J et al. 2019. JASP: graphical statistical software for common statistical designs. *J Stat Softw*. 88:1–17. https://doi.org/10.18637/jss.v088.i02.

Meyer T, Olson CR. 2011. Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci USA*. 108:19401–19406. https://doi.org/10.1073/pnas.1112895108.

Oldfield RC. 1971. The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. 9:97–113. https://doi.org/10.1016/0028-3932(71)90067-4.

R Core Team. (2024). R: A language and environment for statistical computing. https://www.R-project.org/.

Rao RPN, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*. 2:79–87. https://doi.org/10.1038/4580.

Rossion B. 2014. Understanding face perception by means of prosopagnosia and neuroimaging. *Front Biosci*. E6:258. https://doi.org/10.2741/706.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G. 2009. Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychon Bull Rev*. 16:225–237. https://doi.org/10.3758/PBR.16.2.225.

Schliephake LM, Trempler I, Roehe MA, Heins N, Schubotz RI. 2021. Positive and negative prediction error signals to violated

expectations of face and place stimuli distinctively activate FFA and PPA. *NeuroImage*. 236:118028. https://doi.org/10.1016/j.neuroimage.2021.118028.

Stenger S, Bludau S, Mohlberg H, Amunts K. 2022a. Cytoarchitectonic parcellation and functional characterization of four new areas in the caudal parahippocampal cortex. *Brain Struct Funct*. 227:1439–1455. https://doi.org/10.1007/s00429-021-02441-2.

Stenger S, Bludau S, Mohlberg H, Amunts K. 2022b. Probabilistic cytoarchitectonic map of area Ph1 (PhG) (v7.2) [data set]. *EBRAINS*. https://doi.org/10.25493/WWD5-2JX.

Stenger S, Bludau S, Mohlberg H, Amunts K. 2022c. Probabilistic cytoarchitectonic map of area Ph2 (PhG) (v7.2) [data set]. *EBRAINS*. https://doi.org/10.25493/JFQA-6EV.

Stenger S, Bludau S, Mohlberg H, Amunts K. 2022d. Probabilistic cytoarchitectonic map of area Ph3 (PhG) (v7.2) [data set]. *EBRAINS*. https://doi.org/10.25493/TA54-1N8.

Summerfield C et al. 2006. Predictive codes for forthcoming perception in the frontal cortex. *Science*. 310:1311–1314. https://doi.org/10.1126/science.1132028.

Summerfeld C, Wyart V, Johnen VM, de Gardelle V. 2011. Human scalp electroencephalography reveals that repetiti suppression varies with expectation. *Front Hum Neurosci*. 5:1–13. https://doi.org/10.3389/fnhum.2011.00067.

Summerfield C, Trittschuh EH, Monti JM, Mesulam MM, Egner T. 2008. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci*. 11:1004–1006. https://doi.org/10.1038/nn.2163.

Tang MF, Smout CA, Arabzadeh E, Mattingley JB. 2018. Prediction error and repetition suppression have distinct effects on neural representations of visual information. *elife*. 7:1–21. https://doi.org/10.7554/eLife.33123.

Todorovic A, de Lange FP. 2012. Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci*. 32:13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012.

Todorovic A, van Ede F, Maris E, de Lange FP. 2011. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J Neurosci*. 31:9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011.

Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010. Implicit perceptual anticipation triggered by statistical learning. *J Neurosci*. 30:11177–11187. https://doi.org/10.1523/JNEUROSCI.0858-10.2010.

Wacongne C, Changeux JP, Dehaene S. 2012. A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci*. 32:3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012.

Wacongne C et al. 2011. Evidence for a hierarchy of predictions and prediction errors in human Cortex. *Biol Sci*. 108:20754–20759. https://doi.org/10.1073/pnas.111780710.

Wig GS, Grafton ST, Demos KE, Kelley WM. 2005. Reductions in neural activity underlie behavioral components of repetition priming. *Nat Neurosci*. 8:1228–1233. https://doi.org/10.1038/nn1515.