Universität Münster

Lena Maria Leeners

# Neurobehavioural Evidence for Co-Occurring Positive and Negative Prediction Errors During Contextual Model Updating

- 2024 -

Psychologie

# Neurobehavioural evidence for co-occurring positive and negative prediction errors during contextual model updating

Inaugural-Dissertation
Zur Erlangung des Doktorgrades
im Fachbereich Psychologie und Sportwissenschaft
der Universität Münster

Vorgelegt von
Lena Maria Leeners, geb. Schliephake
aus Münster
- 2024 -

Dekan: Prof. Dr. Gerald Echterhoff

Erste Gutachterin: Prof. Dr. Ricarda I. Schubotz
Zweiter Gutachter: Prof. Dr. Joachim Groß

Datum der mündlichen Prüfung: ________________________

Datum der Promotion: ________________________

# Acknowledgements

The work behind this thesis comprises years of dedication and perseverance and would not have been possible without the people who supported me throughout this journey.

First and foremost, I want to thank my supervisor, Ricarda Schubotz, who gave me the opportunity to pursue a PhD in her lab. Her expertise, encouragement and belief in my abilities have been fundamental in shaping this research and guiding me through its challenges. Furthermore, I am very grateful to Prof. Dr. Joachim Groß for co-supervising my PhD and reading my dissertation.

I am also immensely grateful to my colleagues at the Biological Psychology Group for their diverse perspectives, constructive feedback and team spirit. A special thanks goes to Ima Trempler, Jennifer Pomp, Nina Heins, Nadiya El-Sourani, Amelie Hübner, and Malli Roehe who made me feel comfortable in the group from day 1.

To my friends, who have provided support, laughter, and a sense of normality, I am profoundly thankful. Thank you, Lisa Nobis, for accompanying me through all the ups and downs of the last years. Thank you, Amelie Haugg, Julia Brehm, and Pia Schröder who were always my biggest inspiration regarding scientific research. Thank you, Tessa Söbbeke and Ewa Ritters for everything outside of academia. Thank you to many others who I met on this path and who have touched my life in meaningful ways.

Above all, I want to thank my family who always had my back when I felt like I have to work during rather inconvenient times. Marcus, words cannot fully capture my gratitude for everything you have done for me during the past years. Your unconditional support, love, and understanding have been my biggest source of motivation and made me feel like I can actually make it till the finish line. Jonte, even though you are only two years old, you teached me patience, reminded me of the simple joys in life and showed me what unconditional love means. I am also grateful to my brother Philipp; whose loyalty and calm demeanour are a true source of inspiration.

Finally, I want to thank my parents, Tom and Gabi. Your trust, encouragement, and good genes have made this path possible for me. Thank you for always being there for me, for believing in my dreams, and for instilling in me the confidence to pursue them. This thesis is dedicated to you.

# Table of Contents

# Summary

In recent years, the predictive coding account has emerged as the dominant framework in cognitive neuroscience. According to predictive coding, the brain is constantly generating and updating a model of the environment to predict sensory input. At each stage of the processing hierarchy, only the difference between the actual and the expected sensory input – termed the prediction error – is transmitted to higher brain regions for further processing. This process allows the brain to efficiently process sensory information, to prioritise relevant inputs, and to effectively adapt to changes in the environment.

This dissertation includes three studies that together aim to elucidate on the temporal markers of the generation and development of cued expectations, on the one hand, and the simultaneous processing of positive and negative prediction error signals resulting from expectation violations on the other hand. In *Study 1*, we used the temporal advantages of electroencephalography (EEG) to investigate the specific time window underlying the formation and maintenance of predictions that have been acquired through probabilistic learning. As early as the onset of the predictive cue, neural signals indicate the pre-activation of expected sensory input initiated by top-down processing. These signals persisted even after the presentation of the expected stimulus. In *Study 2*, we employed an experimental design that was able to differentiate between positive and negative prediction error signals using the high spatial resolution of fMRI. We revealed simultaneous increases in activation of FFA and PPA, representing concurrent processing of the two prediction error types. While one area represented the error response to the unexpected omission of one stimulus (i.e. a face) the other occurred in response to the unexpected presentation of another stimulus (i.e. a place). *Study 3* was able to replicate part of the results from *Study 2* by additionally controlling for effects that have been shown to influence expectation in previous studies. Again, we found compelling evidence for co-occurring positive and negative prediction error signals that were independent of stimulus repetition effects. We suggest that this co-activation supports the idea of at least two different prediction error circuits that can both process unexpectedly omitted stimuli and unexpectedly presented stimuli at the same time. We showed that prediction errors are processed not only at different levels of the hierarchy but also simultaneously at the same level, a finding that informs future advances in neuroscience by highlighting the complexity and importance of sensory error processing in the brain.

# List of Original Publications

This thesis is based on the following original research articles:

***Study 1*** Roehe, M.A., Kluger, D.S., Schroeder, S.C.Y., Schliephake, L.M., Boelte, J., Jacobsen, T., Schubotz., R.I. (2021). Early alpha/beta oscillations reflect the formation of face-related expectations in the brain. *PLoS ONE* 16*(7):* e0255116

***Study 2*** Schliephake, L. M., Trempler, I., Roehe, M.A., Heins, N., Schubotz,R.I. (2021). Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA. *NeuroImage*, 236, 11802.

***Study 3*** Leeners, L. M., Pomp, J., Zahedi, A., Roehe, M. A., Schubotz, R. I. (2024). Co-occurring positive and negative prediction error signals are independent of stimulus repetition effects. *Cerebral Cortex.* In revision.

# 1. Theoretical and Empirical Background

Imagine walking down a familiar street, surrounded by the sights and sounds of the city. As you navigate your way through the bustling environment, your brain seamlessly processes an abundance of sensory information: the chirping of the birds, the scent of freshly brewed coffee, the sight of familiar faces passing by.

As we take in our surroundings, it is remarkable that our brain not only perceives the world around us but it actively anticipates it. Every familiar face we encounter triggers a cascade of predictions within our brain, based on past experiences and learned patterns. These predictions that are processed beneath the surface of conscious awareness, play a crucial role in guiding our actions and shaping our perception of reality. The fundamental challenge – how the brain reconciles predictions with incoming sensory information lies at the heart of cognitive neuroscience. In the pages that follow, we embark on a journey to unravel some parts of this mystery, exploring the mechanisms of predictive coding and, specifically, prediction errors in the brain. By shedding light on these processes, we aim to deepen our understanding of how the brain constructs our perception of the world and support future advances in neuroscience and beyond.

## 1.1 The human brain as a prediction machine

For a long time, the predominant theoretical discussion guiding research into the neural underpinnings of perception has been based on hierarchical feedforward models. These models posit that sensory information undergoes processing via a series of progressively complex feature detectors (DiCarlo et al., 2012; Hubel & Wiesel, 1968; Poggio & Riesenhuber, 1999). These feature detectors represent individual or groups of neurons in the brain that code for perceptually significant stimuli (Marr, 1982). Neurons in the primary visual cortex (V1) for example, are sensitive towards rather simple lines and edges, whereas neurons in higher levels of the visual hierarchy (i.e. V2 and V4) respond to more complex features such as whole shapes and contours (Huang & Rao, 2011; Hubel & Wiesel, 1968). Over the past decade, however, this notion has been challenged by a more pro-active view of the brain (Von Helmholtz, 1867). While hierarchical feedforward models were effective in explaining certain aspects of neural processing, they did not fully capture the dynamic and interconnected nature of neural networks. More specifically, they do not account for the extensive feedback connections present in the brain. Neurophysiological evidence shows that the brain's architecture is replete with reciprocal connections, suggesting that higher-order cognitive areas influence lower-level sensory processing (Bastos et al., 2012; Galloway, 2004). Moreover, feedforward models fail to explain how the brain uses prior knowledge and contextual cues to interpret ambiguous or noisy sensory input (Friston, 2005).

The emergence of predictive processing theories in the late 20th century marked a paradigm shift, challenging the prevailing view of perception as a passive process driven solely by bottom-up sensory input (Clark, 2013). Building upon the foundational work of Von Helmholtz (1867) and the insights of subsequent generations of researchers, modern predictive processing theories propose a dynamic interplay between top-down predictions and bottom-up sensory input (Friston, 2005; Lee & Mumford, 2003; Rao & Ballard, 1999). This framework offers a new lens through which we can explore perception as an ongoing process of inference, with prediction errors serving as crucial signals for updating internal models and optimising perceptual hypotheses (Hohwy, 2012; Huang & Rao, 2011). There are, however, different versions of predictive processing models that diverge in terms of how they compute the prediction error and how representation and prediction error neurons are conceptualised (Spratling, 2008, 2017).

What all theories have in common is that they suggest that the brain develops a generative model of the world which is used to predict incoming sensory information (Barlow, 1961; Craik, 1943; Gregory, 1980). The brain then compares the predicted with the actual input and updates its internal representation of the world. In a hierarchical manner, a brain area at a higher level sends top-down signals to a lower-level area so that the input of that bottom-up area will be predicted. The actual sensory input is then compared to the predictions, computing the difference in form of a prediction error (Bar, 2009). It is suggested that this process requires at least two functional classes of neurons, i) internal representation neurons that project downward the hierarchy to encode predictions about bottom-up input and ii) prediction error neurons that send information upward and encode the difference between predictions and the actual sensory bottom-up input (Keller & Mrsic-Flogel, 2018). Thus, when bottom-up information matches the information from internal representation neurons, the responses in prediction error neurons decrease. Within the sensory brain areas, internal representation neurons as well as prediction error neurons are expected to not only represent an unspecific surprise response but to be selective for specific stimulus features (Friston, 2005; Rao & Ballard, 1999). These neural signatures of predictive processing highlight the intricate mechanisms through which the brain differentiates between expected and unexpected input, ensuring precise and adaptive responses to the environment (Weiss et al., 2002).

## 1.2 Contextual learning and the development of predictions

To effectively harness these predictive processes, the brain must continually refine its anticipatory capabilities. This refinement hinges on the brain’s ability to learn about and exploit the statistical patterns embedded in the sensory input it receives (Seriès & Seitz, 2013). Because these patterns can vary in complexity and form, the underlying neural mechanisms must adapt to the specific nature of each regularity (de Lange et al., 2018). The simplest form of such a regularity is, arguably, that particular features in our environment appear more often and are therefore generally more likely than others.

Next to simple frequency distributions, the expectation of sensory data can also emerge from contextual probabilities. The current state of our natural surroundings can have a huge impact on the processing of incoming sensory data along the hierarchy of the generative models (Clark, 2013, 2017; Limanowski et al., 2020). Looking back at the scenario mentioned in the first section of this introduction, the day and time might serve as a cue for what we anticipate to perceive along the street we are walking. In the early morning, we might expect to meet a friend at the coffee place like every morning., whereas we would not have such an expectation in the late evening. Consequently, in order to deal with the ever-changing surroundings, the brain utilises the present context to formulate hypotheses about the most probable visual information that is either present or forthcoming.

Looking at Figure 1., it is no problem for us to read the words “the cat” even though the letters in the middle of each word are actually written with an identical set of lines. The context of the surrounding letters and their suggestion of a meaningful word let us interpret the middle letter as “H” in the first word and as “A” in the second word.

**Figure 1**. Example of how contextual probabilities can influence visual perception. Image adapted from Selfridge (1955).

A context emerges through the continuous comparison between sensory input and top-down predictions generated by the brain. The predictions are influenced by various kinds of factors including past experiences, personal preferences or the current state of the environment.

The ongoing refinement of predictions through model updating allows the brain to construct a context that is dynamically adjusted to fit the current situation best.

## 1.3 How expectations modulate sensory processing

A central mechanism of predictive processing that has been repeatedly observed is the dampening or reduction of an incoming signal in response to sensory input that matches its predictions about the environment (Egner et al., 2010; Summerfield & De Lange, 2014; Todorovic & de Lange, 2012). When sensory input aligns closely with these predictions, the brain attenuates the neural response associated with that input, effectively reducing its impact on perception. Consequently, the brain optimises processing by saving resources thereby enhancing its sensitivity to novel or unexpected stimuli. Importantly, the reduction of neural responses to expected events has behavioural consequences as research could show reduced response times and higher accuracy levels to expected as compared to unexpected stimuli (Pinto et al., 2015; Stein & Peelen, 2015; Turk-Browne et al., 2010; Wyart et al., 2012).

Several functional magnetic resonance imaging (fMRI) studies confirmed that expected as compared to unexpected visual stimuli result in a reduced neural response in brain areas associated with the processing of the relevant stimuli (Egner et al., 2010; Grotheer & Kovács, 2014; Kok et al., 2013; Kovács & Vogels, 2014). For example, Egner et al. (2010) employed fMRI and an experimental paradigm in which the colour of the frame surrounding the stimuli was predictive of whether a face or house would appear within this frame. The participants had to detect trials that were made up of occasionally inverted target stimuli and respond via a button press. Therefore, participants were unaware of the probabilistic pairing of frame colour and the stimulus type. The results showed that activation of the fusiform face area (FFA) was significantly lower for highly expected face images when compared to highly unexpected face images. The suppression of expected stimuli could also be underlined by other neurophysiological studies using electroencephalography (EEG) and magnetoencephalography (MEG), as well as spiking activity in nonhuman primates (de Lange et al., 2018). In the context of EEG signals, the event-related potential (ERP) N170 has been identified as a reliable temporal biomarker reflecting the predictability of face-related sensory input. More specifically, studies that investigated face perception underlined that the amplitude of the N170 component significantly decreased for expected faces when compared to unexpected faces (Johnston et al., 2017; Ran et al., 2014).

Regarding the underlying neural mechanism of expectation suppression, there are two main perspectives that have been discussed in the literature (de Lange et al., 2018). The "dampening" account suggests that brain activity associated with anticipated stimuli is reduced, allowing the selective processing of the residual prediction error (Strange et al., 2005). On the other hand, the "sharpening" account posits that anticipated stimuli trigger increased neural responses due to the attenuation of incongruent information, which diminishes the overall amplitude of activity while sharpening the representation of anticipated input (de Lange et al., 2018; Kok et al., 2012; Press & Yon, 2019).

Even though many studies have demonstrated an expectation suppression effect, some findings indicate the opposite phenomenon, referred to as expectation enhancement (Doherty et al., 2005; Jaramillo & Zador, 2011). One possible explanation for these contradictory results is that in some studies expectation and attention covaried, making an interpretation of the findings more complex (Summerfield and De Lange, 2014). Another explanation could be the "sharpening" account itself, which suggests that expectations enhance the neural representation of expected stimuli by sharpening the tuning of neural populations. This leads to more efficient processing and can result in higher neural responses in certain contexts (Press et al., 2020; Press & Yon, 2019). Lastly, in numerous studies employing probabilistic cueing designs, stimulus repetition and stimulus expectation were conflated. This is because repetitions were generally more likely and thus more predictable than alternations of stimuli, exerting similar effects on their targets as expectations (for more detailed description see paragraph 1.5).

Although the fundamental principles of adaptive perception have been thoroughly established, there still is a key gap in our knowledge regarding several aspects of the origin and development of top-down driven expectations. Previous studies investigating expectation effects, for example, predominantly focused on a narrow time window of the time period preceding predicted events instead of looking at the entire interval between the predictive and predictable stimulus (Cao et al., 2017; Spaak et al., 2016).

## 1.4 Adaptation when the context changes – positive and negative prediction errors

Building on the understanding of reduced neural signals for expected stimuli, it is crucial to explore what happens when these expectations are violated. The previously addressed aspects of expectation and repetition suppression lead to the hypothesis that if sensory predictions result in decreased neural responses and are only updated to minimise the prediction error signal, then processing stimuli that do not match previous predictions should result in increased neural responses compared to expected stimuli. Prediction errors have been extensively studied and are commonly characterised as neural responses evoked by unexpected or oddball stimuli. Several fMRI studies could show increased blood-oxygenation-level dependent (BOLD) responses to unexpected stimuli when compared to expected stimuli (Alink et al., 2010; Strange et al., 2005). Additionally, prediction errors in EEG and MEG studies manifested as increased gamma band oscillations (Todorovic et al., 2011; van Pelt et al., 2016) or mismatch negativity responses (Herholz et al., 2009; Huotilainen et al., 1998). Moreover, so-called omission designs have shown that the omission of an expected stimulus results in an error response that is similar to other prediction error responses representing the omitted stimulus (den Ouden et al., 2012; Todorovic et al., 2012).

Expanding on these findings, Keller & Mrsic-Flogel (2018) described two distinct types of prediction error signals, namely, positive and negative prediction errors. Positive prediction errors result from increased bottom-up input when for example, an unpredicted stimulus appears, whereas negative prediction errors follow from sensory input that is weaker than predicted, for example, when an expected stimulus does not appear or is omitted. In the following, I discuss evidence for both types of prediction errors and elaborate on the question whether both, positive and negative prediction error signals can temporally co-occur.

### 1.4.1 Positive prediction errors

In line with positive prediction error processing, single neuron recordings in the inferior frontal cortex of primates showed increased firing rates to unexpected when compared to expected visual input (Meyer & Olson, 2011). Alink et al. (2010) used fMRI and visual stimuli that induced illusory motion to show reduced responses in V1 when the predictability of the motion direction increased. Moreover, other neuroimaging studies using more complex visual stimuli could show increased BOLD signal to unexpected visual stimulus categories in brain regions that specifically code for these categories (e.g. FFA) for face stimuli and parahippocampal place area (PPA) for place stimuli) (Den Ouden et al., 2010; Egner et al., 2010). Altogether, positive prediction errors have been identified as an increase in neural responses to unexpected visual stimuli in higher as well as lower level processing areas suggesting that positive prediction errors presumably mirror local feature tuning that is unique to each (visual) cortical area (Richter et al., 2023). That is, positive prediction error signals should signal the type of deviation from expectation and not just the fact that there was a deviation (Keller & Mrsic-Flogel, 2018).

### 1.4.2 Negative prediction errors

The prediction error signal that is more difficult to explain and to test than the positive prediction errors mentioned previously, is the response to the absence of predicted stimuli or sensory input, the so-called negative prediction error (Keller & Mrsic-Flogel, 2018). To assess top-down prediction error signals, researchers have tried to eliminate any confounding bottom-up sensory input by investigating neural responses to unexpected stimulus omission. Negative prediction errors have been intensively studied in the auditory domain using EEG where participants were exposed to single omissions of tones of previously learned sound sequences (Bendixen et al., 2009; SanMiguel et al., 2013). Moreover, using comparable omission paradigms, several previous studies could show robust cortical responses to unexpected stimulus omissions in the relevant sensory cortical area (e.g. Den Ouden et al. 2009; Kok et al. 2012; Todorovic and de Lange 2012; Fiser et al. 2016). Remarkably, it has been shown that neural responses to unexpected tone omissions were not distinguishable from neural activation patterns to the actual tones (Bendixen et al., 2009). Hence, comparable to positive prediction error signals, the activity patterns of omissions contain information about the identity of the absent stimulus, underlining its representational content (Peelen & Kastner, 2011).

However, up until today, research investigating prediction errors either focused on positive or negative prediction errors neglecting the fact that during our every-day life we often encounter situations in which we experience the absence of a predicted stimulus while being presented with an unexpected stimulus replacing the omitted stimulus. This might represent a situation where positive and negative prediction errors need to be computed simultaneously, giving rise to the idea of separate prediction error circuits in the brain.

## 1.5 Stimulus repetition and stimulus expectation effects

Expanding on the influence of expectations and prediction errors on neural processing, it is important to consider that not only the expectation of a specific stimulus but also the repetition of it can lead to decreased neural responses (Grotheer and Kovacs, 2015; Summerfield et al., 2008). However, the evidence regarding whether and how stimulus repetition and expectation interact is inconsistent. In an often-cited study, Summerfield et al., (2008) conducted an experiment where participants were exposed to pairs of faces where one face was either repeated or alternated with another face image. The stimuli were organised in blocks with high (75%, repetition block) or low (25%, alternation block) repetition probabilities. In repetition blocks, the results showed larger repetition suppression in the FFA than in alternation blocks. The authors suggested that higher-order contextual expectations modulated the repetition-related process. However, this study did not allow to test for independent expectation and repetition effects because of the unbalanced use of high and low repetition blocks to manipulate participants expectation. To investigate the relationship between repetition and expectation effects independently, Todorovic and de Lange (2012) conducted an auditory cue-target MEG experiment, in which participants were confronted with two consecutive tones, which could either be identical (repetition) or different (alternation). Orthogonal to this, the first tone that was presented in each pair was indicative of whether the second tone would be a repetition or an alternation. Importantly, for the first time, this experimental design allowed the clear separation of stimulus repetition and stimulus expectation effects. The authors revealed differential temporal dynamics of the two effects with repetition suppression occurring shortly after stimulus onset (40-60 ms) and expectation suppression during a later time point (100-200 ms). This finding suggests both factors to be independent. To adapt these findings to the visual modality, Grotheer and Kovacs (2015) introduced pairs of images in their experimental design in which the first face could be repeated (repetition) or a different second face was presented (alternation). Similar to Todorovic and de Lange (2018) and in contrast to Summerfield et al. (2008), the authors found independent stimulus repetition and stimulus expectation effects in the occipito-temporal cortex.

Another important factor that needs consideration when investigating the potentially differential effects of stimulus repetition and expectation is the fact that in numerous previous studies the main comparison was between expected and unexpected trials (Grotheer & Kovács, 2014, 2015; Summerfield et al., 2008). This comparison alone cannot clarify whether there was an increase in neural activity in one condition or a decrease in the other. For instance, Feuerriegel et al. (2021) emphasised the importance of including a neutral condition to disambiguate the observed effects.

In the three following studies we investigate the temporal constraints inherent in the generation of cued face-related expectations. Building up on these findings, we then focus on the neural interplay of positive and negative prediction error signals elicited by the unexpected omission and presentation of face and place stimuli. The thesis ends with a study that accounts for potential confounds and enhances the validity of our findings.

## 1.6 Research questions and objectives

The concept of predictive processing has been extensively studied and received support from many domains in cognitive neuroscience in the past years (see Clark, 2013, for a comprehensive review). However, there are still open questions regarding the temporal aspects of prediction formation and the computational mechanisms of prediction errors that we aim to address in the upcoming three studies. Previous investigations into visual predictions have provided limited insights into the temporal dynamics of cued prediction generation. Consequently, it remains unclear whether the neural signatures that indicate the formation of cued expectations are activated at the same time as the onset of the predictive cue and then diminish, or whether this process occurs gradually and persists until the expected event occurs. Therefore, in *Study 1* we made use of the temporal advantages offered by EEG to explore the temporal parameters that define the origin of cued face-related expectations. In *Study 2*, we built up on the findings of the first study by investigating how pre-activations that result from predictive stimuli and violated expectations are connected to different types of prediction errors signals using fMRI. More specifically, we aimed at showing that both positive and negative prediction error signals can co-occur in situations in which an expected visual stimulus is omitted but replaced by another unexpected stimulus. Further, we were interested in changes in functional connectivity between positive and negative prediction error signals and higher order brain areas known to play a role in model updating and error processing. Lastly, we conducted *Study 3* to underline the idea of co-occurring positive and negative prediction error signals while controlling for two aspects that have been shown to influence the interpretation of error processing in previous studies. First, we aimed at showing the independence of expectation effects and stimulus repetition effects. Second, we introduced a neutral control condition in which no clear expectations could be formed, to clarify that prediction error signals induce activation increases and are not the result of its comparison to expected events that usually result in diminished responses.

Taken together, the three research articles contained in this thesis aim to address the following research questions:

1) What are the specific temporal aspects involved in the formation and progression of cued face-related expectations?
2) Can positive and negative prediction error signals co-occur and how are they related to the formation of expectations?
3) Are positive and negative prediction error signals still evident when comparing unexpected with neutral events and are those signals independent of stimulus repetition effects?

# 2. Research Articles

## 2.1 Study 1: Early alpha/beta oscillations reflect the formation of face-related expectations in the brain

Marlen A. Roehe[1,2*], Daniel S. Kluger[2,3], Svea C. Y. Schroeder[1,2], Lena M. Schliephake[1], Jens Boelte[1,2], Thomas Jacobsen[4] and Ricarda I. Schubotz[1,2]

[1]Department of Psychology, University of Münster, Münster, Germany

[2]Otto-Creutzfeldt-Centre for Cognitive and Behavioural Neuroscience, University of Münster, Münster, Germany

[3]Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany

[4]Experimental Psychology Unit, Helmut-Schmidt-University/University of the Federal Armed Forces Hamburg, Hamburg, Germany

*PLOS ONE*, *16*(7), e0255116

**Keywords:** face perception, EEG, visual expectation, N170, alpha/beta oscillations

# Abstract

Although statistical regularities in the environment often go explicitly unnoticed, traces of implicit learning are evident in our neural activity. Recent perspectives have offered evidence that both pre-stimulus oscillations and peri-stimulus event-related potentials are reliable biomarkers of implicit expectations arising from statistical learning. What remains ambiguous, however, is the origination and development of these implicit expectations. To address this lack of knowledge and determine the temporal constraints of expectation formation, pre-stimulus increases in alpha/beta power were investigated alongside a reduction in the N170 and a suppression in peri-/post-stimulus gamma power.

Electroencephalography was acquired from naive participants who engaged in a gender classification task. Participants were uninformed, that eight face images were sorted into four reoccurring pairs which were pseudorandomly hidden amongst randomly occurring face images. We found a reduced N170 for statistically expected images at left parietal and temporo-parietal electrodes. Furthermore, enhanced gamma power following the presentation of random images emphasized the bottom-up processing of these arbitrary occurrences. In contrast, enhanced alpha/beta power was evident pre-stimulus for expected relative to random faces. A particularly interesting finding was the early onset of alpha/beta power enhancement which peaked immediately after the depiction of the predictive face. Hence, our findings propose an approximate timeframe throughout which consistent traces of enhanced alpha/beta power illustrate the early prioritisation of top-down processes to facilitate the development of implicitly cued face-related expectations.

# Introduction

Our environment is of a highly dynamic nature, veiling a cascade of statistical regularities. Explicitly, such regularities often go unnoticed, although traces of implicit learning are evident in the brain's neural activity. These regularities are extracted as sensory input and projected 'bottom-up' over multiple cortical levels in order to establish associative neural representations reflecting external influences [1]. These internal representations are henceforth frequently updated and revised to optimise their reliability. This accumulation of knowledge regarding statistically predictable external recurrences can then be drawn upon when the external input is less informative or lacks certainty [2]. To reduce this ambiguity, predictions based on prior knowledge regarding an external stimulus are sent 'top-down' along the cortical hierarchy and are compared with equivocal sensory-driven input to draw relevant inferences. According to predictive processing frameworks, if, for instance, the top-down prediction and bottom-up input carry dissimilar information, a mismatch in the form of a prediction error is propagated upwards to update a subsequent higher level. On the contrary, no revision of any given level would be necessary if the bottom-up signal is congruent with the top-down prediction. In the context of predictive processing, this bidirectional interplay between incoming sensory signals and top-down projected predictions is the underlying mechanism assisting perception [3,4].

Emerging principles within this field of research have highlighted several biomarkers which support this bidirectional predictive framework. For instance, the face-sensitive event-related potential (ERP), N170, is a reliable temporal marker which, amongst other factors, reflects the level of predictability of a face-related sensory input. Specifically, selected studies investigating face perception conveyed that the amplitude of the N170 component was significantly diminished for expected compared to unexpected faces [5–7]. In line with predictive processing, this suggests that the sensory-driven information of an expected face is met by fairly accurate top-down prediction. To establish such predictions, the brain draws upon prior information of expected events in preparation for their actual occurrence [8]. Recent studies have shown that pre-activation of sensory information, and subsequent sensory priors, are mediated by low frequency oscillations encompassing alpha and beta frequency ranges [8-12]. These oscillations, primarily alpha, are believed to enhance the signal-to-noise ratio in task-related networks by carefully selecting relevant and simultaneously silencing irrelevant populations of neurons to establish a more focused access to representations of expected stimuli [13]. Due to the early access to relatively precise prior information, less cognitive resources are required to process anticipated perceptual input and in turn visual event-related potentials are modulated [7,9,13]. Additionally, updating and optimising this given neural representation would be unnecessary, hence, the forward projection of prediction errors is downregulated. Bottom-up processing as well as the projection of prediction errors functionally relate to high gamma frequency (60 – 100Hz) synchronisation, which requires a greater energetic cost than lower frequencies [14,15]. Based on these findings, a reduction in gamma power would be presumed to proceed the onset of expected events. In contrast, due to the limited access to pre-activated prior information, more cognitive resources would be allocated to processing unexpected occurrences [10,15,16]. Unexpected events could, therefore, be distinguishable from expected occurrences by enhanced post-stimulus gamma-band activity (GBA), whereas expected images are preceded by an enhancement in pre-stimulus alpha/beta power and a suppression in GBA post-stimulus onset [11]. In turn, whilst less cognitive resources are devoted to processing sensory information of expected targets, subsequently evoking a diminished ERP response, the opposite would be expected for novel or surprising occurrences [7,9].

Although the fundamental principles of adaptive perception have been well established, various aspects relating to the genesis and development of top-down driven expectations remain underexplored. Several studies have investigated the presence of pre-stimulus alpha/beta power as an indicator of expectation [9,12,17], as well as examined the pre-activation, maintenance, and transfer of prior face-related knowledge [8]. Yet, these studies primarily focus on a small fragment of the pre-stimulus timeframe immediately prior to the onset of expected events. Therefore, it remains unclear at which point facilitatory processes aiding the development of cued face-related expectations commence and how this development evolves over time. The main aim of the present study was, therefore, to locate the point within the pre-stimulus period at which the enhancement in alpha/beta power is initiated for expected relative to random images. Moreover, we meant to investigate whether this enhancement in alpha/beta power either (i) fluctuates, (ii) shows a gradual and steady increase until the expected event occurs, or (iii) shows an accelerated increase just prior to stimulus onset. To our knowledge, the current study, therefore, provides new insight into the evolution of implicitly cued face-related expectations.

Through employing a statistical learning paradigm during a short training session, participants acquired implicit knowledge of the statistical relationships and, hence, predictable nature of certain stimuli. More specifically, the participants completed an explicit gender classification task whilst implicitly learning and predicting the statistically predictable occurrences of certain face images. Participants consequently relied on a previously establish representation of the interrelationships between certain images to form subsequent perceptual expectations. As such, the formation of these expectations was dependent on memory. Foremost, we aimed to replicate findings verifying the presence of implicit expectations. In line with previous studies, we expected a faster and more accurate behavioural response for expected faces alongside an attenuation of the N170 amplitude [1,7]. Furthermore, we assumed that whilst an enhancement in GBA should succeed the depiction of randomly occurring faces [11], the statistically expected images should be met with a prior elevation of alpha/beta activity

[8,11,12,17]. A systematic relationship between this increase in alpha/beta activity and the amplitude reduction of the N170 in response to the stimulus would subsequently support the suggestion that the increase in alpha/beta activity reflects predictive processes. Thus, our primary motivation was to examine pre-stimulus alpha/beta power, indicating the development of face-related expectations, in order to determine the initial onset and offset confining the formation process of implicitly cued expectations.

# Materials and methods

## Participants

A total of 33 individuals participated in this study (23 women; 23.1 ± 3.51 years of age [mean ± SD]) after having signed informed consent based on the principles expressed in the declaration of Helsinki. All participants were right-handed as assessed by the Edinburgh Handedness Inventory [18], reported (corrected-to-) normal visual acuity and had no history of neurological and psychiatric disorders. For compensation, participants were either accredited with class credits or reimbursed for their participation (25 Euros). Four additional participants, whose EEG data contained excessive sweat artefact contamination (severe drifts in the signal), were excluded from further analyses. The study was approved by the Ethics Committee of the University of Münster (Department of Psychology).

## Stimulus material

Participants were presented with 25 neutral face images (12 women) chosen from the Radboud Faces Database (RaFD) [19]. Since visual information to process faces is extracted using sequences of eye fixations over mainly eye regions (the mouth region being second), all images were scaled so that those facial features (especially eyes) aligned [20]. This was done to reduce the amount of eye movements.

To generate statistical regularities, eight of these images were sorted into four reoccurring pairs under the following sequential guidelines: i) a male face invariably preceded a female face, ii) a male face invariably preceded another male face, iii) a female face invariably preceded a male face, and iv) a female face invariably preceded another female face. Each individual participant was assigned a unique set of four pairs which were pseudorandomly embedded amongst reoccurring arbitrary face images. The face images (W = 9.5cm, H = 14cm) were depicted individually in the centre of a black background for 500ms (subtending visual angles of approx. 9° vertically and 6° horizontally). These were immediately followed by a 17ms white noise mask and a fixation period of 2483ms. Each of these trials was, therefore, a total length of 3000ms (Figure 1).

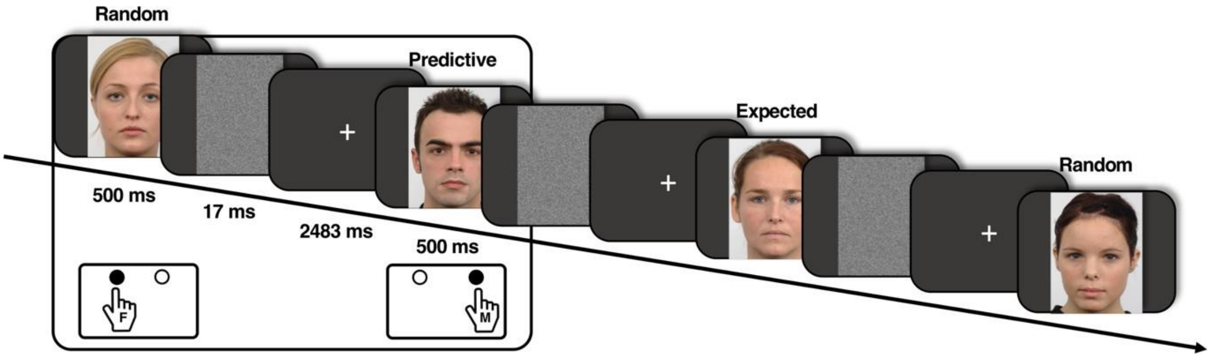


**Figure 1 Schematic illustration of the experimental task.** Each image was depicted for 500ms followed by a white noise mask (17ms) and a darkened fixation screen (2483ms). The participants were instructed to press either the left (right index finger) or right (right middle finger) button on a response box to discriminate between female (F) and male (M) images, respectively.

## Task

The participants engaged in a gender classification task without having prior knowledge of the presence of the embedded pairs. They were given a response box and were instructed to respond as fast and accurately as possible – via a right-hand button press – respective to the gender of the face presented on screen. Here a left button press (right index finger) classified the presented image as a female face, whilst a right button press (right middle finger) classified the depicted face image as male (Figure 1).

## Experimental procedure

Participants were tested on two consecutive days. The first day consisted of a short, 18-minute behavioural training session, providing a chance for the participants to gain implicit knowledge regarding the presence of the paired images and to familiarise themselves with the classification task at hand. Four image pairs were pseudorandomly hidden amongst 17 arbitrary face images, with each image depicted five times to form sequences (blocks) of approximately six to seven minutes (125 images within each block). The first image within each pair served as a predictive event for the second image and subsequently enhanced its predictability. Focusing on this predictability regarding the occurrence of a certain image, all images could be sorted into two image categories – *expected* and *random*. Given the confounding informative nature of a predictive image, these images were removed from analyses determining the predictability of a shown image (behavioural, ERP and gamma-band analyses). The timeframe following the onset of predictive images was, however, examined in the alpha/beta analysis, since it also served as the pre-stimulus timeframe for the expected images.

The EEG session on the following day comprised of an elongated replica of the training session. All images were equally distributed throughout eight blocks with self-determined breaks separating them (yielding a total of 1000 trials). Whilst the four image pairs remained consistent throughout the training and EEG sessions for each individual participant, the combinations of paired images differed and were counterbalanced across participants. The participants were seated comfortably in a dimly lit EEG booth and advised to keep general movement to a minimum. Overall, the EEG session took approximately 50 – 60 minutes to complete depending on the length of the breaks.

A questionnaire following immediately after the EEG session tested the participants' explicit awareness of the predictive nature underlying the classification task. The participants were asked to state whether they had noticed face images which invariably ensued certain predictive images and asked to identify them seriatim.

The experiment was programmed and performed using Presentation 18.1 (Neurobehavioral Systems, San Francisco, CA, USA).

## Behavioural data analysis

The statistical analysis of the behavioural response time (RT) and accuracy (percentage of correct responses) was performed in R (version 3.6.0; R Foundation for Statistical Computing, Vienna, Austria; Rstudio Team, 2015). Premature and prolonged responses (occurring 3 SDs faster/slower than the aggregated group mean), in addition to incorrect answers, were excluded from the behavioural analysis. Random images were arbitrarily selected to create a sample size equal to the number of expected images (~160 of expected and ~170 of random image trials per participant). Since we hypothesised that the responses for the expected faces would display an increase in accuracy aligned with a decrease in RT, these two aspects of performance were subjected to individual dependent, one-tailed *t*-tests for the two image categories (expected versus random).

# EEG data analyses

**EEG data acquisition.** Scalp EEG was acquired using 62 Ag/AgCl-electrodes mounted to the actiCAP snap electrode cap in combination with the BrainVision Recorder software (Brain Products, Gilching, Germany). The electrodes were placed according to the 10-20 system and additional electrooculogram (EOG) electrodes were attached below and next to the right eye to account for vertical and horizontal eye movement, respectively. An online bandpass filter (0.1 – 1000Hz) was applied to the EEG data recorded at a sampling rate of 1kHz. Electrodes FCz and FPz served as online reference and ground, respectively, and were disregarded from all analyses. Electrode impedance was maintained below 10 kΩ.

**EEG signal processing.** EEG data was pre-processed offline using the EEGLAB toolbox (version 14.1.1b) [21] in MATLAB (R2017b). The raw data was down-sampled to 500Hz and bandpass filtered by applying a 0.1Hz high-pass and 30Hz low-pass Butterworth filter (12 db/octave) for the ERP analysis, whereas a 0.5Hz high-pass and 100Hz low-pass Butterworth filter was implemented for the time-frequency analysis (TFA) [22]. Line noise was suppressed at the source through a carefully designed set-up (as recommended by [23]). Continuous data was segmented into epochs extending from -200ms pre- to 600ms post-stimulus onset for ERPs. The 200ms prior to stimulus onset served as a baseline. For the TFA, data was epoched from -2000 to approximately 1500ms, time-locked to stimulus onset. These time segments of 3500ms framing image onset were used with the intention to allow edge artefacts to subside before and after our points of interest [22]. Consecutive epochs overlapped by approximately 500ms to minimise loss of data during convolution. Ocular correction was applied using the Gratton plug-in for EEGLAB [24]. Noisy channels (kurtosis criterion: $z > 6$) were manually inspected and replenished by an interpolation of neighbouring electrodes (ERPs: 1.14% and TFA: 0.66% of electrodes were interpolated). For the ERP analysis, semiautomatic artefact inspection discarded epochs contaminated by artefacts exceeding an amplitude threshold of ± 75 µV or voltage fluctuations greater than 50 µV with regard to the previous sample point (3.9% of trials removed). For the TFA, epochs with artefacts exceeding an amplitude threshold of ± 200 µV and voltage fluctuations greater than 50 µV were rejected (6.2% of trials removed). A dataset was disregarded when more than 2 SDs of trials were removed during the semiautomatic rejection (ERP mean: 943 trials; TFA mean: 941 trials). Henceforth, four out of the 37 participants were dismissed from all further processing. During the final pre-processing step, datasets were re-referenced to a common average. For all EEG analyses, the number of expected and random trials was equalised across participants (expected: ~160 per participant; random: ~170 per participant).

**Event-related potentials.** The epochs framing the event of interest were averaged across each image category (expected and random) for each individual participant. The N170 was quantified by measuring the mean amplitude within the timeframe of 150 – 200ms in relation to a pre-stimulus baseline of 200ms. Based on former literature, we restricted the attributing electrode sites to exclusively P7/P8 and TP7/TP8 for our analysis [25-29]. In line with our directional hypothesis, the mean amplitudes for expected and random images were subjected to dependent, one-tailed *t*-tests for each set of electrodes (left hemisphere: TP7/P7; right hemisphere: TP8/P8; see S1 File for an alternative repeated measures cluster permutation test approach).

**Time-frequency.** The spectral analysis was performed using the MATLAB toolbox FieldTrip [30]. Spectral power was estimated by applying FFT to a sliding window passing through averaged trials (for both low and high frequencies). A Hanning taper was used for low frequencies (2 – 30Hz) by centring a 500ms fixed sliding window that moved in time steps of 50ms and 1Hz increments. This process subsequently constructed trial windows extending from -1750 to 1200ms (stimulus locked), as 250ms on either side of the original epoch frames (-2000 – approx. 1500ms stimulus locked) were discarded due to convolution. For high frequencies (40 – 100Hz), an adaptive DPSS (discrete prolate spheroidal sequences) multitaper approach was applied [31]. Estimates were acquired using a 500ms fixed sliding window maintaining identical stepwise motion over time and frequency axes as previously stated (± 4Hz smoothing).

Cluster-based permutation tests were computed in three dimensions (frequency, channel, and time) to correct for multiple comparisons. Hence, voxels of the two image categories were subjected to Monte Carlo randomisation tests with 1000 iterations and a significance level of $\alpha = .05$. With reference to our hypotheses regarding the enhancement of alpha/beta power pre-stimulus and the diminution of gamma power post-stimulus onset for expected versus random faces, dependent, one-tailed *t*-tests were computed for these permutation tests. Statistical tests were performed on the normalised difference in raw power estimates between expected and random images (difference$_{\text{expected vs. random}} = (X-Y)/(X+Y)$). This normalisation was also applied to all spectral data used for time-frequency representations. Lastly, data-driven analyses were carried out to assess the relationship between the observed effect in pre-stimulus alpha/beta power and the modulation of the N170. For each participant, the difference in normalised alpha/beta power between expected and random images was calculated and averaged over channels, frequency, and time. In this case, power was only averaged across those channels that contributed to the positive cluster. The difference in mean amplitude for the N170 was computed and averaged over time and left electrodes for each participant. The modulation of alpha/beta power within the immediate pre-stimulus timeframe (1250 – 3000ms) was z-standardised and correlated with the z-standardised magnitude of the difference in mean amplitude for the N170. To examine the functionality of the last two peaks more closely, their underlying alpha/beta power was segregated and correlated individually with the modulation of the N170. The timeframe proceeding the onset of the predictive images was not included in these analyses because of the images' informative and 'cue-like' nature.

# Results

## Behavioural results

Throughout the EEG session, participants engaged in a classification task with the instructions to identify the gender of the depicted faces as fast and as accurately as possible. The performance for both image categories conveyed the participants' close engagement with the task at hand (expected: 96% accuracy; random: 96% accuracy; one-tailed: $t(32) = 0.51$, $p = .305$). The RT for the two observed image categories showed a mean of 478ms (*SD* = 49ms) for expected faces and 479ms (*SD* = 46ms) for random faces. The dependent, one-tailed *t*-test showed no significant difference between the two categories ($t(32) = 0.38$, $p = .352$).

Notably, the answers of the questionnaires revealed that only a single participant became explicitly aware and was capable of correctly identifying merely one out of the four confronted pairs.

## Event-related potential results

The modulation of the N170 played a fundamental role in identifying whether the participants had gained implicit knowledge regarding the predictability of the paired images. Our approach examined four relevant channels – TP7 and P7 in addition to their right lateral counterparts –for a substantial reduction in mean amplitude (150 – 200ms) for expected faces. Supporting our hypothesis, a significant reduction in mean amplitude was observed for expected (-1.31 ± 2.40 µV) versus random images (-1.51 ± 2.25 µV) at left parietal and temporal-parietal channels (Bonferroni-corrected: $p = .039$; Figure 2). In contrast, no significantly reduced amplitude for expected faces was evident at right-lateralised channels (expected: -1.80 ± 3.02 µV; random: -1.86 ± 3.05 µV; Bonferroni-corrected: *p* = .569).

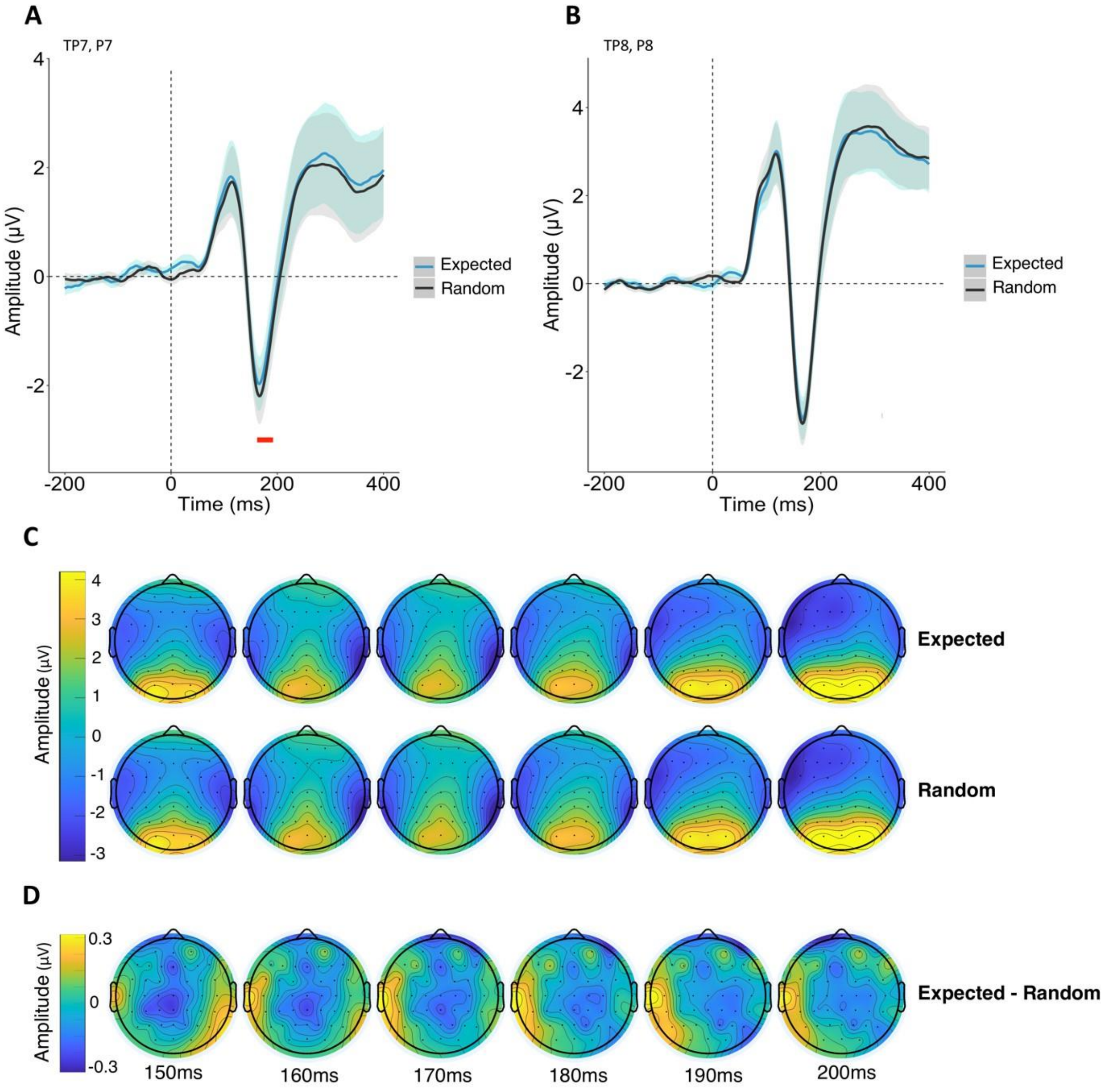


**Figure 2 (A)** A significant reduction in the N170 amplitude for expected (blue line) in comparison to random (black line) faces was found across the averaged electrodes TP7 and P7 (left hemisphere). The timeframe of the significant mean amplitude difference is marked by a red dash. The shaded area illustrates within-participants confidence intervals for the expected (blue) and random (grey) faces. **(B)** No vast differences in the N170 were evident across electrodes TP8 and P8 (right hemisphere). **(C)** Voltage topographies (µV) for expected and random images show activity in a timeframe from 150 – 200ms following stimulus onset. **(D)** Topographies show voltage differences between expected and random face images.

## Time-frequency results: gamma oscillations

An underlying implication of predictive processing is the notion that the feeding forward of sensory information is upregulated for novel or unexpected as compared to expected occurrences [14]. In accordance with this conception, our time-frequency analysis showed a significant reduction in gamma power for expected versus random images within the first 1000ms upon stimulus onset (*p* = .04, cluster corrected; Figure 3). In other words, a significant enhancement in GBA was found in response to random compared to expected images. No particular frequencies within the broad gamma-band were singled out a priori, meaning, gamma power (40 – 100Hz) was treated as a singular entity. Interestingly, we observed that the gamma-related spectral difference between random and expected images appeared to reside in an early and late enhancement (Figure 3A). Visual inspection of the cluster revealed that the channels contributing to these elevations in GBA were predominantly located over posterior scalp regions (Figure 3B).

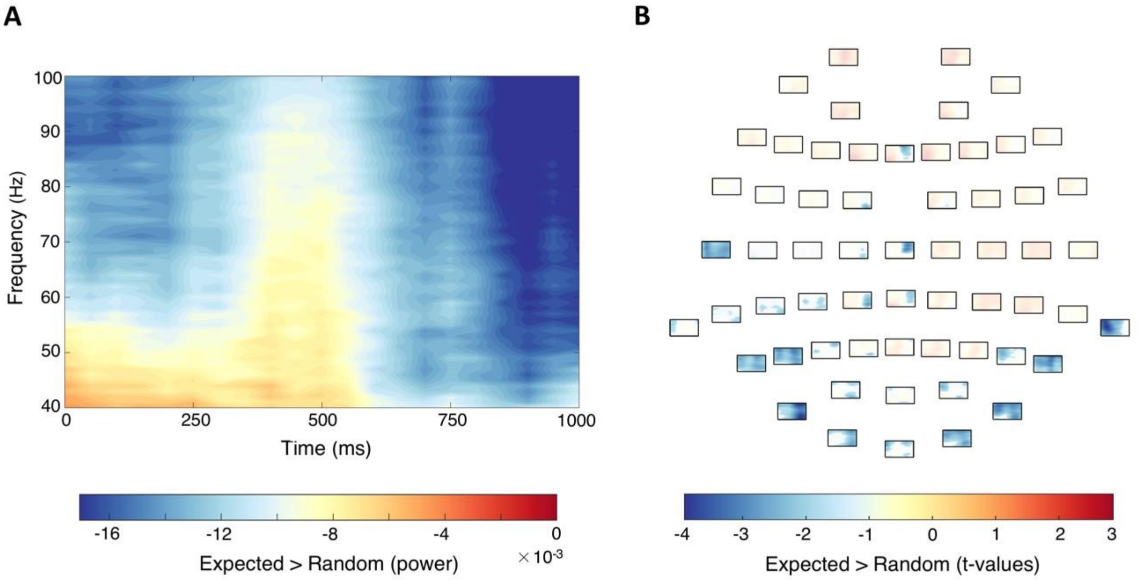


**Figure 3 (A)** Time-frequency representation (TFR) showing the normalised difference in gamma-band activity (GBA) between expected and random images peri- and post-stimulus onset (40 – 100Hz; 0 – 1000ms). Power was averaged across electrodes contributing to the negative cluster (*p* = .04, cluster corrected). **(B)** TFRs for each individual channel illustrate the topographical distribution of the negative cluster. The significant time and frequency points composing the negative cluster stand out as opaque; insignificant differences are transparent.

## Time-frequency results: alpha and beta oscillations

The above analyses determined several neural traces which suggest that the predictive nature underlying the classification task was learned implicitly. Building upon this finding, we examined alpha and beta frequencies to determine a confined time-window preceding an expected event, in which the development of an expectation was reflected by its power distribution pattern. Hence, this analysis focused on the presence and distribution of low frequencies, primarily associated with top-down processes, in the short timeframe preceding the expected stimuli [12,17,32]. In line with our hypothesis, we found a significant enhancement in alpha and beta power for expected images (in comparison to random images) ranging from approximately 1250ms after the presentation of the predictive image to onset of the expected stimulus (*p* = .017, cluster corrected; Figure 4B). However, as displayed in Figure 4A, the positive cluster seems to commence prior to this enclosed timeframe. Thus, we additionally contrasted the time course extending from stimulus onset until 1200ms post-stimulus onset between predictive and random images. Interestingly, a further positive cluster enclosing alpha/beta frequency bands was observed in this timeframe (*p* = .044, cluster corrected). Visual inspection of the cluster suggested that the channels corresponding to the largest power differences between predictive and random images were predominantly located over central electrodes within occipital, parietal, and frontal regions (Figures 4Ai and 4Aii). These topographical distributions then spread to primarily occipital and frontal regions for the largest difference in alpha/beta power between expected and random images (Figure 4Bi). In contrast to the previous scalp maps, the final enhancement in alpha/beta power was mostly lateralised bilaterally over parietal regions (Figure 4Bii). Collectively, these findings suggest that the facilitation of the development of a cued expectation is initiated by the onset of the predictive image and ends shortly prior to depiction of the expected image.

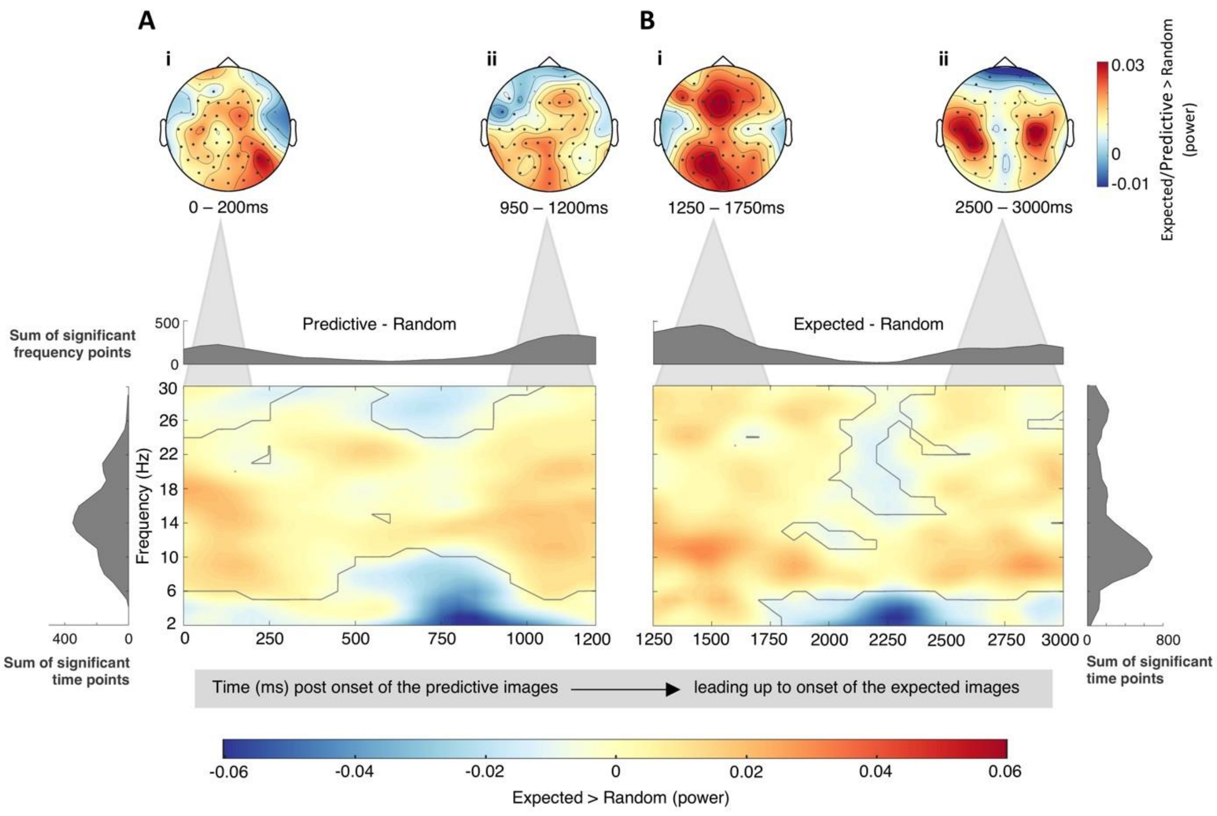


**Figure 4** TFRs of the normalised differences in low frequency power (2 – 30Hz) averaged across channels contributing to the positive clusters (outlined in grey). The timeframe shown extends from stimulus onset (predictive/random) to stimulus onset (expected/random). The histograms along the x-axes show the sum of significant frequency points per time point across cluster contributing channels. The reversal is shown in the histograms along the y-axes (sum of significant time points for each frequency across cluster contributing channels). **(A)** Significant cluster for the alpha/beta power differences between predictive minus random images. Scalp maps illustrate the topographical distribution of the greatest power differences (predictive - random) within **(i)** 0 – 200ms (14 – 19Hz), and **(ii)** 950 – 1200ms (10 – 18Hz). **(B)** Significant cluster for the alpha/beta power differences between expected minus random images. The scalp maps illustrate the topographical distribution of power differences (expected - random) within **(i)**1250 – 1750ms (10 – 14Hz) and **(ii)** 2500 – 3000ms (6 – 11Hz). A 50ms rift disjoins the 3000ms interstimulus timeframe as a result of the chosen epoch size and Fourier transform parameters (see *Materials and methods*). Both **4A** and **4B**, however, provide supportive indications to assume that in place of the 50ms rift, a steady increase in significant frequency points (histograms along the x-axes) would link the gradual increase in **4A** with the peak seen in **4B**.

*Post hoc* correlations demonstrated that the modulations of alpha/beta power (8-30Hz) within neither the entire pre-stimulus timeframe immediately prior to stimulus onset (1250 – 3000ms) nor the middle peak (1250 – 1750ms) significantly correlated with the magnitude of the reduction of the N170 (Spearman's rho = .19, *p* = .838, 95% CI [-0.16 0.50]; Spearman's rho = .04, *p* = 1, 95% CI [-0.31 0.38], respectively). A significant positive relationship was, however, observed between the modulation of alpha/beta power underlying the final peak (8-30Hz; 2500 – 3000ms) and the modulation of the N170 (Spearman's rho = .46, *p* = .021, 95% CI [0.14 0.69]; Figure 5). All above *p*-values were Bonferroni-adjusted to correct for multiple comparisons. Collectively, these findings suggest that the final peak could reflect a relatively precise expectation of the upcoming stimulus. The continuous enhancement of alpha/beta power extending throughout the entire interstimulus interval may, on the other hand, provide an elongated favourable state optimal for expectation formation.

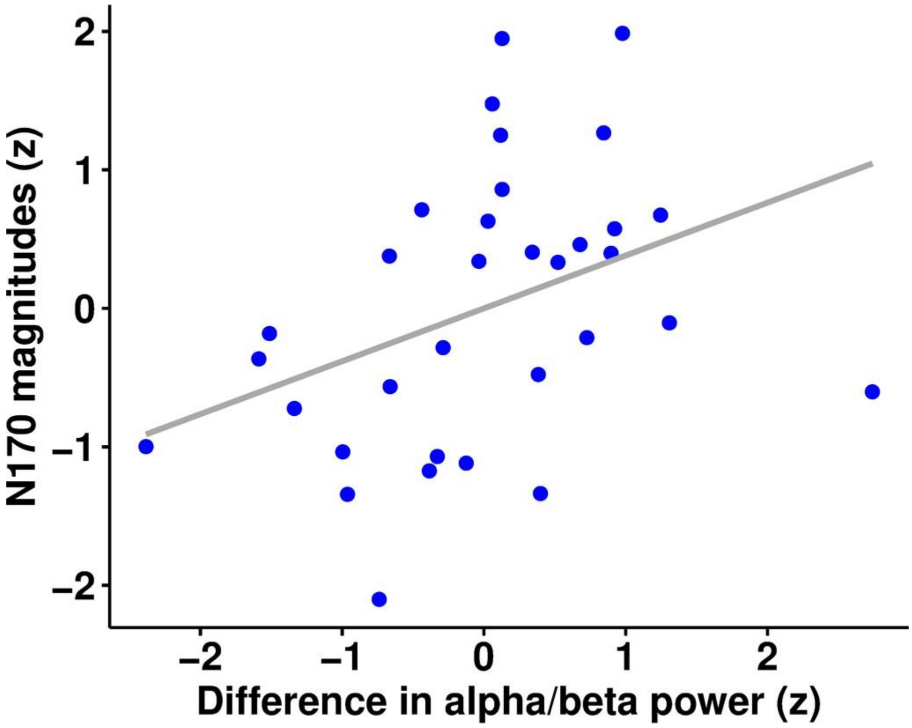


**Figure 5** Correlation between the z-standardised modulations of pre-stimulus alpha/beta power (8-30 Hz; 2500 – 3000ms) and of the left-lateralised N170 (N = 33).

# Discussion

The present study provides findings which suggest that the development of implicitly cued expectations is optimised by the early prioritisation of top-down processes. In turn, predictable visual events are met by relatively accurate implicit expectations to allow the brain to reserve cognitive resources. These processes were reflected by enhancements in pre-stimulus alpha/beta power for expected relative to randomly occurring faces. Intriguingly, this enhancement commenced as early as the onset of the predictive image and prevailed until the expected stimulus occurred. A correlation between the final elevation in alpha/beta power and the reduction of the N170 revealed a positive relationship between these two modulations. Ultimately, a reduction in bottom-up processing for expected relative to random images appeared to be reflected by a suppression in post-stimulus gamma power (Figure 6).

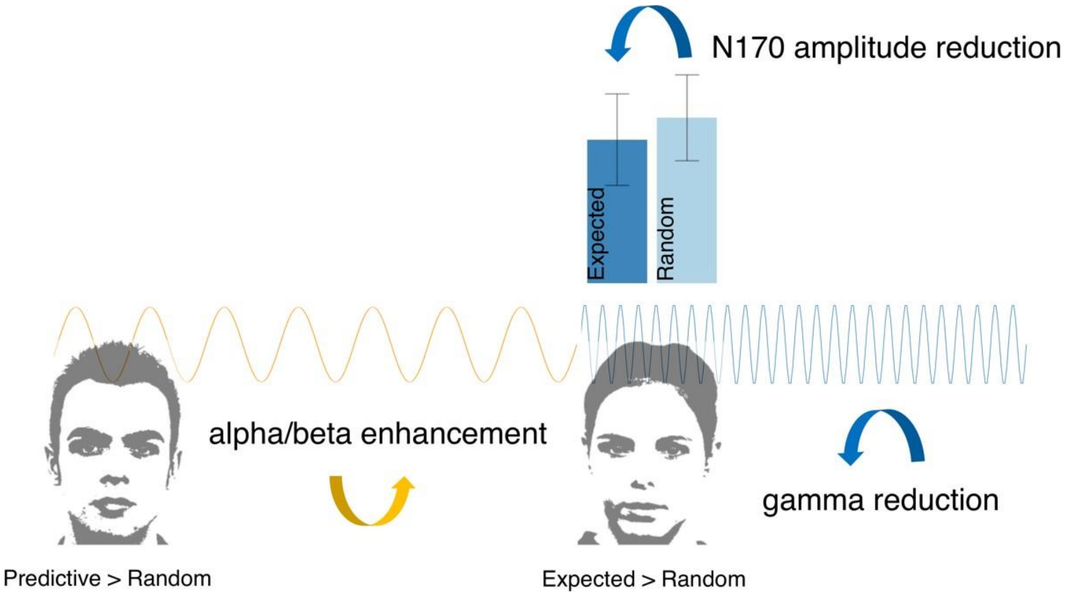


**Figure 6** Schematic overview of observed electrophysiological modulations in response to predictive/expected relative to random images.

Through employing a short statistical learning test (training) prior to the EEG session, participants were given the chance to acquire implicit knowledge regarding the predictive relationship between paired face images. At first glance, the gain of neither a significant decrease in response time nor a significant increase in accuracy for statistically expected images seems at odds with previous studies [1,7]. Considering the simplicity and repetitiveness of our task, however, the lack of behavioural effects could be caused by a ceiling effect. Turk- Browne and collegues [1], for instance, took into account that signs of statistical learning become evident after merely 2 to 3 repetitions. In their experiment, novel images and paired image combinations were introduced in each new block to eliminate the likelihood of reaching a plateau in response time and accuracy across expected and unexpected images. Noting that our focal point of interest lay with the origination of implicit expectations and not the statistical learning process as such, our task (the 25 images and paired-up faces remained the same throughout the experiment) may have permitted participants to quickly reach optimal proficiency. A further point to consider is that the images were presented for an entirety of 500ms which may have buffered a speeded reaction. A reduced presentation period could, thus, help encourage participants to give a more speeded response.

Whilst this study does not allow us to draw a strong conclusion regarding the lateralisation of the N170 effect, the observed modulation of the N170 does, however, appear to support the notion that participants implicitly differentiated between statistically expected and random images (Figure 2). This observation is in line with the established premise that a smaller fraction of cognitive resources is devoted to expected in comparison to somewhat unexpected or surprising events [5,7]. This in turn results in an attenuated electrophysiological response. An assumption as to why this attenuation was solely observed over left-lateralised electrodes is based on the principles that the right-lateralised N170 seems to be more sensitive to familiarity than the left-lateralised N170. That is, studies regarding the role of the N170 during face identity processing have shown that consecutive presentation of identical face images leads to a reduction in the N170 [33,34]. On the contrary, Jemel, Schuller & Goffaux [35] observed an enhanced N170 amplitude for familiar (famous) in comparison to unfamiliar faces during overt face recognition. These habituation effects or memory-driven modulations were found to be predominantly right-lateralised, irrespective of gender [33-35]. Although consecutive depictions of the same face were not permitted within our pseudorandomisation parameters, each individual image was presented 55 times (regardless of its assigned category) throughout the two experimental sessions (training and EEG). Thus, it seems plausible that the N170 components for expected and random images may have, to some extent, been influenced by habituation. This right-lateralised habituation effect may, therefore, have dampened a considerable right-lateralised, expectation-related modulation.

On a different note, past studies have shown sex-related differences in face processing and the lateralisation of the N170. These findings suggest a dominating right lateralisation of the N170 in men and a more bilateral tendency in women [36,37]. Intriguingly, Proverbio and colleagues (2012) showed that sex-coding studies revealed a slightly different pattern in hemispheric lateralisation [38]. Here, women showed a more dominating left-lateralised response whilst men showed bilateral functioning; thus, suggesting that the involvement of the left hemisphere is essential during gender classification in both gender groups. Given that in the present study participants performed a gender classification task, the observed left-lateralised modulation of the N170 may have been influenced by the underlying nature of the task at hand and the fact that women outnumbered men (10 men and 23 women). However, this unbalanced sample makes it difficult to draw firm conclusions regarding any sex-related differences impacting hemispheric lateralisation. Ultimately, this question would be interesting to pursue in future, with an adequately designed study that specifically investigates how sex-related difference may impact the origination of face-related expectations.

On a final note, the current study used faces as stimuli because the N170 component is a well-established signature of face processing. We would, however, like to emphasize that the N170 has also been reported for non-face stimuli [5,39]. Whether the reduction of the N170 along with the modulations in alpha/beta power observed here for expected faces generalises across other stimulus categories remains to be investigated.

Acknowledging that the occurrence of the face images without a preceding predictive image lacked the predictability of the paired images, random images were deemed to require more cognitive resources and elicit an enhanced gamma-band response. In other words, since all images were task relevant and required a specific behavioural response, it seems likely that more cognitive resources were necessary for processing randomly occurring images, for which the gender was not foretold by a predictive image. In line with previous findings, we observed enhancements in GBA for the somewhat unexpected random images within peri- and post-stimulus periods (Figure 3A; for a review see [40]). Drawing on previous studies, gamma synchronisation has been shown to play a facilitatory role during specific neural functions such as feature binding of incoming visual information [16], the projection of prediction errors [10,14], and influencing synaptic strength during memory encoding and retention [15,41]. When linking these previous findings to our observations, the early peri-stimulus enhancement in GBA (~ 0 – 500ms) could reflect the feeding forward of salient visual information which unifies each individual random image. This notion is supported by the observation that this early gamma-band enhancement is predominantly distributed over occipital electrodes (Figure 3B). Namely, regions which are associated with low-level perceptual processing. Initial processing of incoming visual information, therefore, seems to be augmented for the somewhat unexpected in comparison to expected stimuli.

Given that fast gamma frequencies (~ 60 – 100Hz) are deemed optimal for strengthening synapses during the encoding and updating of short-term memories [14], the later post-stimulus enhancement in broadband gamma power (~ 500 – 1000ms) could indicate that neural representations of the random images are encoded, retained, and revised [16]. Since the random images were recurrently presented over the duration of the experiment, it seems plausible that associated representations could be kept “active” whilst being progressively updated by bottom-up sensory input upon depiction. Consistent with previous findings by Arnal, Wyart & Giraud [10] and Bauer and colleagues [11], the late enhancement in broadband GBA could, thus, reflect the augmented projection of prediction errors from early visual areas via low GBA and the revision of higher cortical levels via high GBA for random images (Figure 3B).

Complementary to previous studies, we found that the depiction of expected targets (in comparison to random faces) was met by an enhanced pre-stimulus alpha- and beta-band activity [11,12,17]. Extending previous findings, we observed that this enhanced alpha/beta activity persisted throughout the entire interstimulus interval. Interestingly, this elongated enhancement in alpha/beta power was governed by three peaks that marked the largest differences in power between expected and random images (Figure 4). The first peak, cresting shortly after stimulus onset (~ 0 – 200ms), suggests an elevation in alpha/beta activity for predictive relative to random images (Figure 4A). It appears that the initial activation of underlying processes facilitating expectation formation is subsequently triggered by the informative attribute of these cue-like images. Namely, the predictive image itself marks a pivotal juncture and foretells the approach of a certain expected face. The largest power difference between predictive/expected and random images appears to be primarily located across central electrodes within occipital, parietal, and frontal regions (Figures 4Aii and 4Bi). Even though the corresponding scalp map does not provide the same spatial resolution as magnetoencephalography results, this topographical distribution seems to show the engagement of predominantly dorsal regions, frequently associated with the propagation of top-down processes [10,11]. This continuous modulation in alpha/beta power, thus, seems to suggest that the prioritisation of top-down processes commences much earlier than just immediately prior to the occurrence of the expected target. Several past accounts have provided evidence to suggest that alpha/beta power is an electrophysiological marker for the inhibition of forward feeding networks [13,42]. Arguably, it seems very plausible that a similar neural state is elicited upon the presentation of the predictive image. As such, the predictive image seems to give rise to a favourable condition in which increases in alpha/beta power reflect prioritisation of top-down processes whilst competing forward-feeding representations are suppressed. Especially since each predictive image only cued a single specific face, alternative neural representations were unnecessary to be processed or maintained during this interval. The reverse has been demonstrated recently in a study by Griffith et al., (2019), which showed that a decrease in alpha/beta power (disinhibition of relevant networks) facilitates information processing [42]. Thus, the continuous maintenance of a favourable condition within the timeframe confined by the onsets of the predictive and expected images could appear to aid the development of precise perceptual expectations.

In the context of predictive processing, alpha oscillations are leading modulators of attention and expectation. Yet, the process of how these two means modulate information processing remains controversial. Recent studies have suggested that whilst attention boosts the precision of prediction error by synaptic gain, expectation regulates the precision of top-down predictions [43,44]. In the latter case, a highly predictable event would, thus, yield fewer prediction errors which would be distinguishable by a subsequent attenuation in high frequency neural responses. Given that our predictive images invariably prompted certain face images, the expectation generated should ideally have been fairly accurate. Upon stimulus depiction, less iterative optimisations between hierarchical levels should, therefore, have been necessary to establish a relatively precise representation of the expected stimulus. The observed positive relationship between the final enhancement in alpha/beta activity (2500-3000ms; Figure 4B) and the left-lateralised reduction of the N170 appears to be coherent with this modulation framework. Namely, data-driven observation suggests that there is a systematic relationship between the modulation in alpha/beta power occurring immediately prior to stimulus onset and the modulation of the peri-stimulus N170. This electrophysiological pattern is also in line with the notion that preactivated prior knowledge and subsequent predictions regarding an approaching target must be maintained until this predictable event is encountered [8]. Hence, this *post hoc* observation supports the hypothesis that processes of expectation are reflected in increased alpha/beta activity, which makes processing of expected stimuli more efficient and consequently reduces the amplitude of the N170. In this case, the prioritising of top-down predictions could, therefore, be elucidated in terms of prediction accuracy [11,43].

To summarise, we extended findings substantiating the presence of perceptual expectations. Of particular interest was the observation that the predictive images seemed to be essential for the initiation of the evolution of cued expectations. This was rendered by the enhanced alpha/beta activity cresting shortly after the depiction of the predictive faces. Even though the early peaks did not correlate with the modulation of the N170, the onset of the predictive image seems to initiate an early optimisation of a favourable neural state to boost the development of relatively precise perceptual expectations. The perception of predictable faces is subsequently facilitated through the implementation of these expectations, leading to a suppression in bottom-up information processing reflected by a reduction of the N170 and GBA. The facilitation of the development of implicitly cued face-related expectations, thus, appeared to prevail over the entire interstimulus period with fluctuations in alpha and beta power varying throughout the three second timeframe. One could question why this pre-stimulus enhancement in alpha/beta power fluctuates instead of being a stable and continual increase in power leading right up to the presentation of expected targets. An explanation could be that the spectral distribution pattern within this timeframe is biased by the temporal aspects of the experimental parameters. Given that three seconds are a relatively long interstimulus interval for this particular perceptual task, the gradual decrease in alpha/beta power (approx.

1500 – 2250ms; Figure 4B) could illustrate a progressive conservation of top-down processes before the final power enhancement which marks the imminent approach of the expected target. This would also explain why only the last peak in alpha/beta power positively correlated with the modulation of the N170. Averaging over the entire immediate pre-stimulus timeframe may have concealed a systematic relationship between the aggregated pre-stimulus alpha/beta power and the modulation of the ERP. The precise functional purpose of the second and most prominent peak would, however, benefit from further investigation, which we intend to do in a currently orchestrated study. In addition, although phase analysis is beyond the scope of the present study, it would be an intriguing research question for future studies to investigate alpha/beta phase coherence at the timepoints of each of the observed peaks.

In conclusion, the current study provides new insight into the temporal dynamics and development of face-related expectations. Notably, our findings raise the notion that the formation of cued expectations does not occur at random within the period preceding a statistically expected target. Instead, the facilitation of this developmental process appears to be instigated by the predictive image and proceeds, with fluctuations in growth, until shortly before the depiction of the target. In turn, expected stimuli are met by a relatively precise expectation to allow the brain to reserve cognitive resources. The evolution of implicit face-related expectations, thus, seems to prevail over the entire interstimulus period. From these results we could draw a timeframe confining the genesis and reflecting the developmental nature of cued face-related expectations. As such, these results open up opportunities for future studies to investigate and pinpoint more specific aspects underlying the anticipation of faces. It would, for instance, be of interest to narrow down the precise functional roles – as well as the neural networks – of the observed pre-stimulus peaks in alpha/beta power. Collectively, this would further advance our understanding of how the development of perceptual expectations is shaped in preparation for upcoming expected targets.

# Acknowledgements

We would like to thank Monika Mertens, Marie Kleinbielen, Katherina Thiel, Annika Garlichs, Alina Dette, Corinna Gietmann and Falko Mecklenbrauck for helping to recruit participants and/or collect data. In addition, we are exceedingly grateful for Dr. Axel Kohler's insightful ideas throughout the early stages of this study. Last but not least, for always being willing to engage in constructive discussions, we would like to thank Prof. Dr. Pienie Zwitserlood and all the members of the Schubotz Lab.

## Supporting information

**S1 File. Alternative ERP analysis.** Repeated measures cluster permutation test approach.

# References

1. Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. Implicit Perceptual Anticipation Triggered by Statistical Learning. J Neurosci. 2010;30(33): 11177–87. doi: 10.1523/JNEUROSCI.0858-10.2010.
2. Kluger DS, Quante L, Kohler A, Schubotz RI. Being right matters: Model-compliant events in predictive processing. PLoS One. 2019;14(6): e0218311. doi: 10.1371/journal.pone.0218311.
3. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. Behav Brain Sci. 2013;36(3): 181–204. doi: 10.1017/S0140525X12000477.
4. Friston K. A theory of cortical responses. Philos Trans R Soc B Biol Sci. 2005;360(1456): 815–36. doi: 10.1098/rstb.2005.1622.
5. Johnston P, Robinson J, Kokkinakis A, Ridgeway S, Simpson M, Johnson S, et al. Temporal and spatial localization of prediction-error signals in the visual brain. Biol Psychol. 2017;125: 45–57. doi: 10.1016/j.biopsycho.2017.02.004.
6. Johnston P, Overell A, Kaufman J, Robinson J, Young AW. Expectations about person identity modulate the face-sensitive N170. Cortex. 2016;85: 54-64. doi: 10.1016/j.cortex.2016.10.002.
7. Ran G, Zhang Q, Chen X, Pan Y. The effects of prediction on the perception for own-race and other-race faces. PLoS One. 2014;9(11): e114011. doi: 10.1371/journal.pone.0114011.
8. Brodski-Guerniero A, Paasch GF, Wollstadt P, Özdemir I, Lizier JT, Wibral M. Information-Theoretic Evidence for Predictive Coding in the Face-Processing System.

J Neurosci. 2017;37(34): 8273-8283.
doi: 10.1523/JNEUROSCI.0614-17.2017.

9. Mayer A, Schwiedrzik CM, Wibral M, Singer W, Melloni L. Expecting to see a letter: Alpha oscillations as carriers of top-down sensory predictions. Cereb Cortex. 2016;26: 3146-3160.
doi: 10.1093/cercor/bhv146.

10. Arnal LH, Wyart V, Giraud AL. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. Nat Neurosci. 2011;14(6): 797–801.
doi: 10.1038/nn.2810.

11. Bauer M, Stenner MP, Friston KJ, Dolan RJ. Attentional Modulation of Alpha/Beta and Gamma Oscillations Reflect Functionally Distinct Processes. J Neurosci. 2014;34(48): 16117–25.
doi: 10.1523/JNEUROSCI.3474- 13.2014.

12. Cao L, Thut G, Gross J. The role of brain oscillations in predicting self-generated sounds. Neuroimag. 2017;147: 895–903.
doi: 10.1016/j.neuroimage.2016.11.001.

13. Klimesch W. Evoked alpha and early access to the knowledge system: The P1 inhibition timing hypothesis. Brain Res. 2011;1408: 52-71.
doi: 10.1016/j.brainres.2011.06.003.

14. Bastos AM, Vezoli J, Bosman CA, Schoffelen J M, Oostenveld R, Dowdall JR, et al. Visual Areas Exert Feedforward and Feedback Influences through Distinct Frequency Channels. Neuron. 2015;85(2): 390–401.
doi: 10.1016/j.neuron.2014.12.018.

15. Kaiser J, Lutzenberger W. Human gamma-band activity: A window to cognitive processing. Neuroreport. 2005;16(3): 207–11.

doi: 10.1097/00001756-200502280-00001.

16. Buschman TJ, Miller EK. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. Science. 2007;315(5820): 1860–2.
doi: 10.1126/science.1138071.

17. Spaak E, Fonken Y, Jensen O, de Lange FP. The Neural Mechanisms of Prediction in Visual Search. Cereb Cortex. 2016;26(11): 4327–36.
doi: 10.1093/cercor/bhv210.

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

19. Langner O, Dotsch R, Bijlstra G, Wigboldus DHJJ, Hawk ST, van Knippenberg A. Presentation and validation of the radboud faces database. Cogn Emot. 2010;24(8): 1377–88.
doi: 10.1080/02699930903485076.

20. Blais C, Jack RE, Scheepers C, Fiset D, Caldara R. Culture shapes how we look at faces. PLoS One. 2008;3(8): e3022.
doi: 10.1371/journal.pone.0003022.

21. Delorme A, Makeig S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods. 2004;134(1): 9–21.
doi: 10.1016/j.jneumeth.2003.10.009.

22. Cohen MX. Preprocessing Steps Necessary and Useful for Advanced Data Analysis. In: Analyzing Neural Time Series Data. Massachusetts: The MIT Press; 2014. pp. 73-85.

23. de Cheveigné A, Nelken I. Filters: When, Why, and How (Not) to Use Them. Neuron.

2019;102(2): 280–93.
doi: 10.1016/j.neuron.2019.02.039.

24. Gratton G, Coles MGH, Donchin E. A new method for off-line removal of ocular artifact. Electroencephalogr Clin Neurophysiol. 1983;55(4): 468–84.
doi: 10.1016/0013-4694(83)90135-9.

25. Gao C, Conte S, Richards JE, Xie W, Hanayik T. The neural sources of N170: Understanding timing of activation in face-selective areas. Psychophysiology. 2019;56(6): e13336.
doi: 10.1111/psyp.13336.

26. Kropotov JD. Sensory Systems and Attention Modulation. In: Functional Neuromarkers for Psychiatry. Elsevier; 2016. pp. 137–69.

27. Rossion B, Jacques C. The N170: Understanding the Time Course of Face Perception in the Human Brain. In: Kappenman ES, Luck SJ, editors. The Oxford Handbook of Event-Related Potential Components. New York: Oxford University Press; 2012. pp.115-142.

28. Shen C, Stasch J, Velenosi L, Madipakkam AR, Edemann-Callesen H, Neuhaus AH. Face identity is encoded in the duration of N170 adaptation. Cortex. 2017;86: 55–63.
doi: 10.1016/j.cortex.2016.10.010.

29. Taylor J, Shehzad Z, McCarthy G. Electrophysiological correlates of face-evoked person knowledge. Biol Psychol. 2016;118: 136–46.
doi: 10.1016/j.biopsycho.2016.05.011.

30. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Comput Intell Neurosci. 2011;2011: 1–9.
doi: 10.1155/2011/156869.

31. Cohen MX. Multitapers. In: Analyzing Neural Time Series Data. Massachusetts: The MIT Press; 2014. pp.203-9.

32. van Pelt S, Heil L, Kwisthout J, Ondobaka S, van Rooij I, Bekkering H. Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Soc Cogn Affect Neurosci. 2016;11(6): 973–80.
doi: 10.1093/scan/nsw017.

33. Caharel S, D'Arripe O, Ramon M, Jacques C, Rossion B. Early adaptation to repeated unfamiliar faces across viewpoint changes in the right hemisphere: Evidence from the N170 ERP component. Neuropsychologia. 2009;47(3): 639–43.
doi: 10.1016/j.neuropsychologia.2008.11.016.

34. Campanella S, Hanoteau C, Dépy D, Rossion B, Bruyer R, Crommelinck M, et al. Right N170 modulation in a face discrimination task: An account for categorical perception of familiar faces. Psychophysiology. 2000;37(6): 796-806.
doi: 10.1017/S0048577200991728.

35. Jemel B, Schuller AM, Goffaux V. Characterizing the Spatio-temporal Dynamics of the Neural Events Occurring prior to and up to Overt Recognition of Famous Faces. J Cogn Neurosci. 2010;22(10): 2289–305.
doi: 10.1162/jocn.2009.21320.

36. Godard O, Leleu A, Rebaï M, Fiori N. Sex differences in interhemispheric communication during face identity encoding: Evidence from ERPs. Neurosci Res. 2013;76(1–2): 58–66.
doi: 10.1016/j.neures.2013.03.005.

37. Proverbio AM, Riva F, Martin E, Zani A. Face coding is bilateral in the female brain. PLoS One. 2010;5(6): e11242.
doi: 10.1371/journal.pone.0011242.

38. Proverbio AM, Mazzara R, Riva F, Manfredi M. Sex differences in callosal transfer and hemispheric specialization for face coding. Neuropsychologia. 2012;50(9): 2325–32. doi: 10.1016/j.neuropsychologia.2012.05.036.

39. Thierry G, Martin CD, Downing P, Pegna AJ. Controlling for interstimulus perceptual variance abolishes N170 face selectivity. Nat Neurosci. 2007;10(4): 505–511. doi: 10.1038/nn1864.

40. Martinovic J, Busch NA. High frequency oscillations as a correlate of visual perception. Int J Psychophysiol. 2011;79(1): 32-38. doi: 10.1016/j.ijpsycho.2010.07.004.

41. Zheng C, Colgin LL. Beta and gamma rhythms go with the flow. Neuron. 2015;85(2): 236–7. doi: 10.1016/j.neuron.2014.12.067.

42. Griffiths BJ, Mayhew SD, Mullinger KJ, Jorge J, Charest I, Wimber M, et al. Alpha/beta power decreases track the fidelity of stimulus specific information. Elife. 2019;8: e49562. doi: 10.7554/eLife.49562.

43. Gordon N, Tsuchiya N, Koenig-Robert R, Hohwy J. Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. PLoS Biol. 2019;17(4): e300023. doi: 10.1371/journal.pbio.3000233.

44. Smout CA, Tang MF, Garrido MI, Mattingley JB. Attention promotes the neural encoding of prediction errors. PLoS Biol. 2019;17(2): e2006812. doi: 10.1371/journal.pbio.2006812.

# S1 File

## Alternative ERP analysis

The N170 amplitude was quantified by measuring the mean amplitude within the timeframe of 150 – 200ms in relation to a pre-stimulus baseline of 200ms. For this computation, the Mass Univariate ERP Toolbox for MATLAB [1] was employed to analyse the mean amplitude across selected channels (TP7/TP8, P7/P8) and within the respective timeframe. The ERPs for expected and random images were subjected to repeated measures cluster permutation tests with 5000 random iterations to correct for multiple comparisons. In line with our hypothesis that expected faces elicit a modulation in the N170 amplitude in comparison to random faces, a dependent, one-tailed *t*-test was computed for each electrode and timepoint (a total of 100 comparisons). Clusters arose when the one-tailed *t*-tests resulted in a *p*-value of or less than .05. The sum of the *t*-values within each cluster were compared and the maximum values extracted throughout each permutation to derive the null hypothesis distribution. Our test statistic of interest was thus deemed significant, if it grounded outwith the 95% tail of the derived distribution.

A significant reduction in amplitude was observed for expected in comparison to random images ($p = .039$) at left parietal and temporal-parietal channels (Fig 2A). The cluster permutation test also conveyed that this positive, left-lateralised cluster extends from approximately 158 – 188ms peri-stimulus (Fig 2A). In contrast, no significantly altered amplitude for expected faces ($p = .172$) was evident across the equivalent right-lateralised channels (Fig 2B)

# References

1. Groppe DM, Urbach TP, Kutas M. Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. Psychophysiology. 2011;48(12): 1711–25. doi: 10.1111/j.1469-8986.2011.01273.x

## 2.2 Study 2: Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA

Lena M. Schliephake[1], Ima Trempler[1,2], Marlen A. Roehe[1,2], Nina Heins[1,2], Ricarda I. Schubotz[1,2]

[1] Department of Psychology, University of Muenster, Muenster, Germany

[2] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

*NeuroImage, 236*, 118028

**Keywords:** prediction error; predictive coding; psychophysiological interaction; visual processing hierarchy; fMRI

# Highlights

- Trial sequences evoked (and then violated) expectations about upcoming stimulus category (faces and places)
- We used functional magnetic resonance imaging (fMRI) to differentiate between positive and negative prediction error signals in the fusiform face area (FFA) and parahippocampal place area (PPA)
- Neural responses to violated expectations about faces and places were evident in right FFA and PPA but only with regard to FFA and PPA's non-preferred stimulus categories
- Only positive prediction error signals were significantly correlated with right inferior frontal gyrus (IFG) activation during unexpected face and place trials

# 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.

# 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, Monti, & Summerfield, 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 & 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. 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, Schoenbaum, & Gershman, 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, Babayan, Uchida, & Watabe-Uchida, 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 & Olson, 2011) or only negative prediction errors (Fiser et al., 2016; den Ouden et al., 2012; Eliades & Wang, 2008; Stanley & 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 & Mrsic-Flogel, 2018; Rao & 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, Hoffman, & Gobbini, 2000; Liu, Harris, & Kanwisher, 2002), whereas place and house stimuli elicit higher activations in the parahippocampal place area (PPA) (Epstein & Kanwisher, 1998; Ishai, Ungerleider, Martin, & Haxby, 2000). Moreover, both imagery and expectation of faces and places lead to activity increases in FFA and PPA, respectively (Esterman & Yantis, 2010; O'Craven & 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, Takaura, Wang, Fujii, & Dehaene, 2018; Trempler et al., 2020) presumably scaling with the amount of modification needed to adapt the current predictive model accordingly (Alexander & Brown, 2018; El-Sourani, Trempler, Wurm, Fink, & Schubotz, 2019; Keller & 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.

# Materials and Methods

## 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.

## 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.

## 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%. With this manipulation, we ensured that participants would be able to predict the 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.

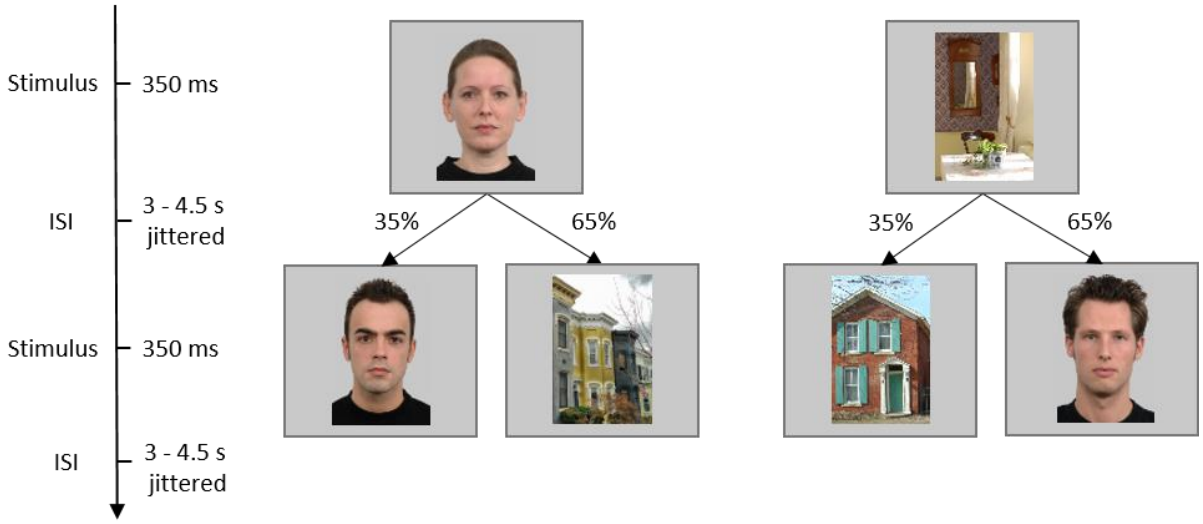


**Fig. 1. Example stimuli including stimulus transitions with their corresponding transitional probabilities and underlying time course.** The "expected" condition included between-category transitions (i.e., Face-Place; Place-Face) whereas the "unexpected" condition included within-category transitions (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).

## Task

All participants completed a 20-minute 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 seconds.

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 minutes in the scanner.

## 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.

## Data analysis

### 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 1500ms starting from trial onset during which participants' responses were classified as valid as for example suggested by McKendrick et al. (2013). Responses recorded after these 1500ms 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.

## 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 2x2x2 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.

# Results

## Behavioural results

**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); $t_{(31)} = 4.17$, $p < 0.001$).

**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; $t_{(32)} = -4.00$, $p < 0.001$) (Figure 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); $t_{(32)} = 4.44$, $p < 0.001$) (Figure 2B).

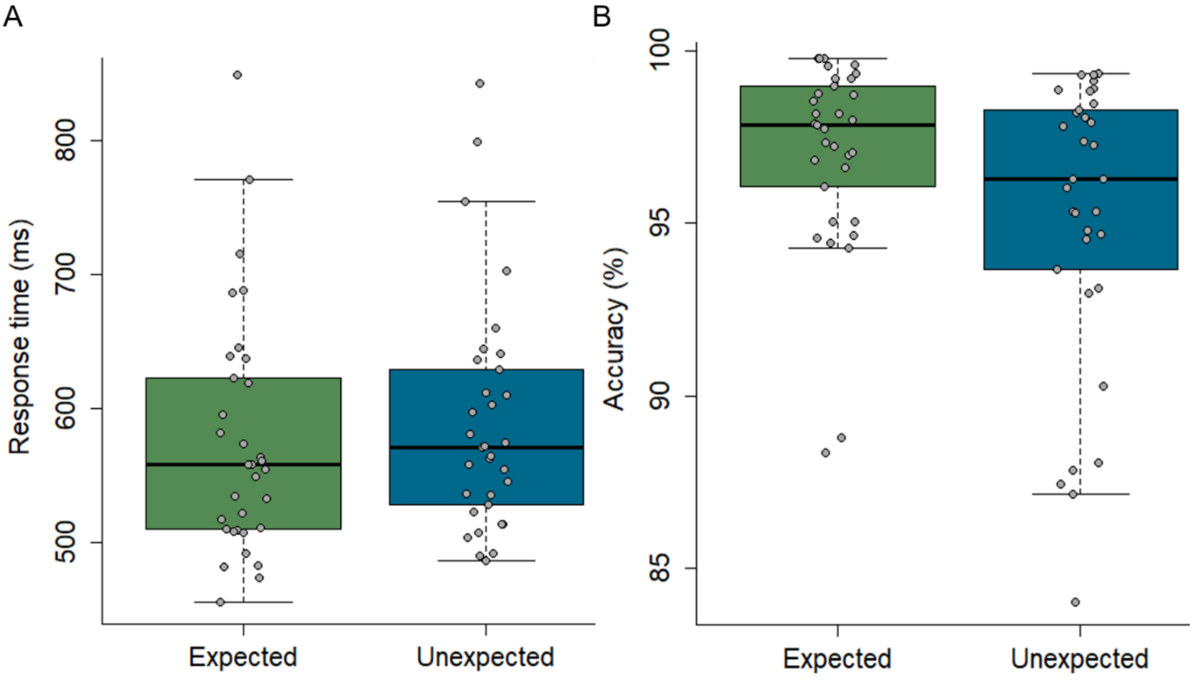


**Fig. 2. Behavioural measures. (A)** Mean response times for expected and unexpected trials and **(B)** mean percentage of correct responses. Black horizontal lines represent the mean values, boxes represent the standard error of the mean (*SEM*), and grey vertical lines represent the standard deviation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

**fMRI.** 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).

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 Figure 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.

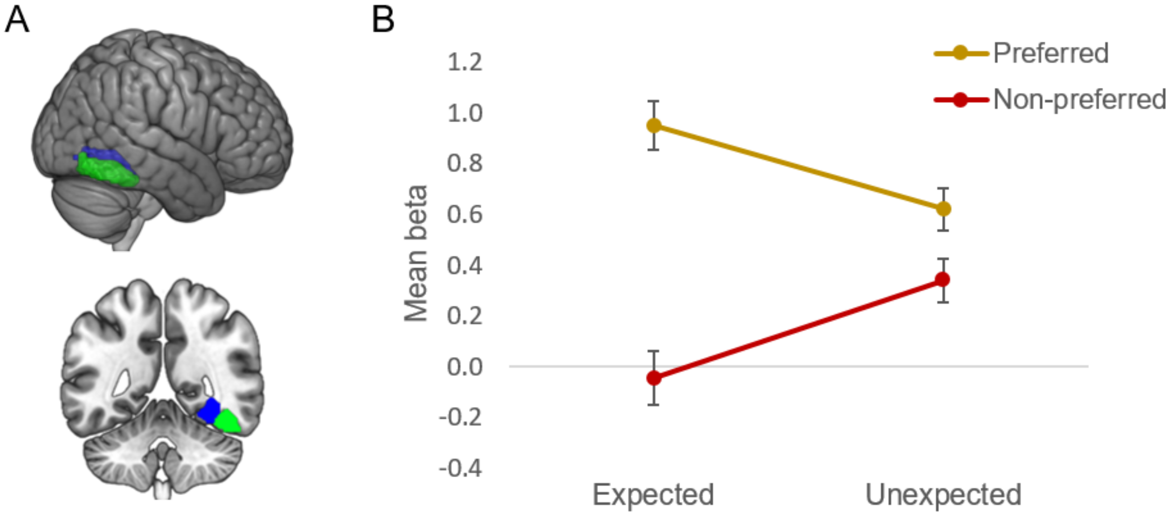


**Fig. 3. Region of interest (ROI) fMRI data. (A)** Right FFA and right PPA ROIs used for beta value extraction. **(B)** Mean group activation estimates ($\beta$ parameters +/- SEM) for expected_preferred, expected_non-preferred, unexpected_preferred, and unexpected_non-preferred trials.

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 Figure 4 and Table 1 for an overview of the PPI results.

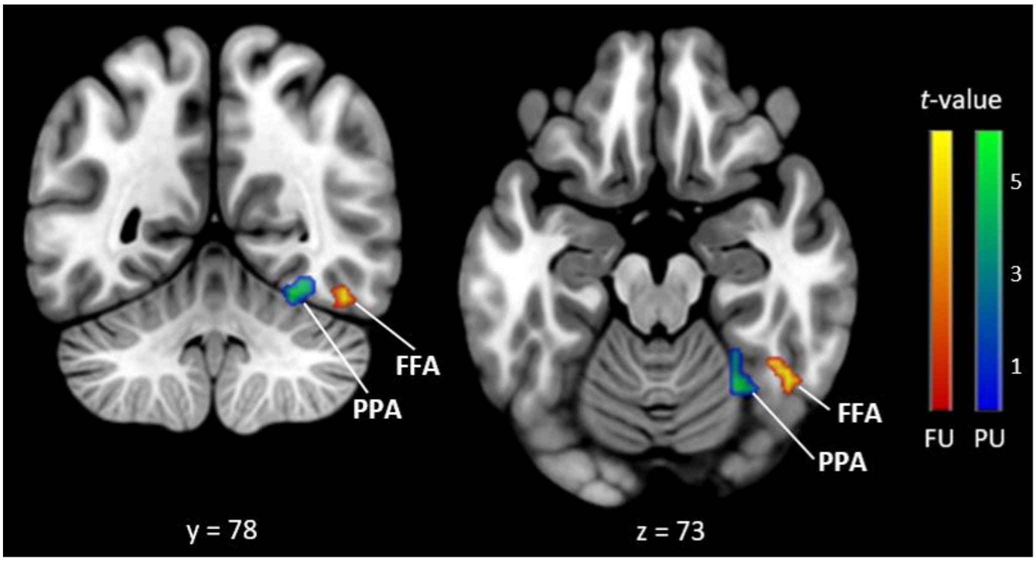


**Fig. 4. Results of PPI analyses for the right IFG as seed region on the right FFA for FU > FE (in red) and on the right PPA for PU > PE (in blue)**. For display purposes, the activation identified by the PPI in the respective ROIs (FFA, PPA) is shown at $p < 0.005$ (*uncorrected*). Significance of 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 |

# 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 & Yantis, 2010; Turk-Browne, Scholl, Johnson, & Chun, 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, Monti, & Summerfield, 2010b). 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, Kok, & de Lange, 2012; Fiser et al., 2016; Kok, Brouwer, van Gerven, & de Lange, 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 & 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, Henson, & Martin, 2006; Krekelberg, Boynton, & van Wezel, 2006; Miller, Li, & Desimone, 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 & 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 & 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 & Lepage, 2005; Chen et al., 2012; Maunsell & Treue, 2006; Posner, Snyder, & Davidson, 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 & 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 & 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, Chamberlain, Monti, Duncan, & Owen, 2010; Hampshire, Thompson, Duncan, & Owen, 2009) reflecting the costs of processing this information (Alexander & Brown, 2018; El-Sourani et al., 2019; Keller & 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 & Mrsic-Flogel, 2018; Rao & 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.

## Acknowledgments

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.

# References

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 1–11. https://doi.org/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. https://doi.org/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. *Journal of Cognitive Neuroscience, 8*(6), 551-565. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.cortex.2013.05.002

Chao, L. L., Weisberg, J., & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cerebral Cortex*, *12*(5), 545–551. https://doi.org/10.1093/cercor/12.5.545

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.e3. https://doi.org/10.1016/j.neuron.2018.10.004

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. *Frontiers in Human Neuroscience*, *6*(JUNE 2012), 1–11. https://doi.org/10.3389/fnhum.2012.00187

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology*, *3*(DEC), 1–12. https://doi.org/10.3389/fpsyg.2012.00548

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of Neuroscience*, *30*(49), 16601–16608. https://doi.org/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. *Journal of Cognitive Neuroscience*, *32*(2), 326–337. https://doi.org/10.1162/jocn_a_01480

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

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

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*(5), 1245–1253. https://doi.org/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. *Nature Neuroscience*, *19*(12), 1658–1664. https://doi.org/10.1038/nn.4385

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Gardner, M. P. H., Schoenbaum, G., & Gershman, S. J. (2018). Rethinking dopamine as generalized prediction error. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1891). https://doi.org/10.1098/rspb.2018.1645

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: Neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, *10*(1), 14–23. https://doi.org/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. https://doi.org/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. *Cognitive, Affective and Behavioral Neuroscience*, *9*(1), 103–112. https://doi.org/10.3758/CABN.9.1.103

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

Horvitz, J. C. (2000). Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience*, *96*(4), 651–656. https://doi.org/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. *Journal of Cognitive Neuroscience, 12*(SUPPL. 2), 35–51. https://doi.org/10.1162/089892900564055

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435. https://doi.org/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. *Journal of cognitive neuroscience*, *26*(7), 1546-1554. https://doi.org/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. *Journal of Neuroscience*, *33*(41), 16275–16284. https://doi.org/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 in Neurosciences*, *29*(5), 250–256. https://doi.org/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. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience, 5*(9), 910–916. https://doi.org/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. *Cerebral Cortex (New York, N.Y. : 1991)*, *27*(1), 373–385.

https://doi.org/10.1093/cercor/bhv225

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

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. https://doi.org/10.7554/eLife.21886

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory. *Science*, *254*(November), 1377–1379. Retrieved 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. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023. https://doi.org/10.1162/08989290051137549

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/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. *Social cognitive and affective neuroscience*, *7*(5), 604-609. https://doi.org/10.1093/scan/nss055

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70. https://doi.org/10.1093/scan/nsm006

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals.

*Journal of Experimental Psychology: General*, *109*(2), 160–174. https://doi.org/10.1037/0096-3445.109.2.160

Yon, D., de Lange, F. P., & Press, C. (2019). The predictive brain as a stubborn scientist. *Trends in Cognitive Sciences*, *23*(1), 6-8. https://doi.org/10.1016/j.tics.2018.10.003

Press, C., & Yon, D. (2019). Perceptual prediction: Rapidly making sense of a noisy world. *Current Biology, 29*(15), R751-R753. https://doi.org/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. *Journal of Neuroscience*, *34*(38), 12828-12836. https://doi.org/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. *Nature Neuroscience, 2*(1), 79–87. https://doi.org/10.1038/4580

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience, 6*(258), e307.

Schiffer, A. M., Waszak, F., & Yeung, N. (2015). The role of prediction and outcomes in adaptive cognitive control. *Journal of Physiology-Paris*, *109*(1-3), 38-52. https://doi.org/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. https://doi.org/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 motorvisual priming. *NeuroImage*, *34*(1), 290–299. https://doi.org/10.1016/j.neuroimage.2006.08.043

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/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 patients tested on and off medication: An fMRI study. *NeuroImage*, *212*(September 2019), 116674. https://doi.org/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. *Journal of Neuroscience*, *30*(33), 11177–11187. https://doi.org/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. *Nature Communications*, *9*(1), 1-8. https://doi.org/10.1038/s41467-018-06752-7

# 2.3 Study 3: Co-occurring positive and negative prediction error signals are independent of stimulus repetition effects

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

[1] Department of Psychology, University of Muenster, Muenster, Germany

[2] Otto Creutzfeldt Center for Cognitive and Behavioral Neuroscience, University of Muenster, Muenster, Germany

[3] Teaching Center Medical University of Vienna, Vienna, Austria

*Cerebral Cortex.* In revision

**Keywords:** predictive coding, fMRI, prediction error, FFA, PPA, stimulus repetition

# Abstract

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 prediction error 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 an MR 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 represented 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.

# 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 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 scrutinise positive and negative prediction errors in the same experimental paradigm (Schliephake et al. 2021). We used a speeded forced-choice 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 (i.e. Grotheer and Kovacs, 2015; Summerfield et al., 2008). 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 & Kovacs (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 neighbour 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, Hence, 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, 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 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 categorised two different stimulus categories. 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 (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 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 (i.e. face-place; place-face) were more likely than equal transitions (i.e. 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 behavioural 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 (CR) to increase for expected stimuli when compared to either neutral or unexpected stimuli. Moreover, we hypothesised to find positive evidence for an independence of the factors expectation and stimulus repetition

On the neural level, we hypothesised 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 (e.g. a face) as compared to the expected stimulus category (e.g. 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 BOLD signal within the brain area processing the unexpectedly presented stimulus category (here PPA), thereby representing the absence of a preferentially processed expected stimulus category (e.g. 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 hypothesised 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 behavioural data due to low performance levels with an accuracy of more than 2 *SD*s below the mean accuracy level of all participants. The mean age of the remaining 30 participants was 23.21 years (*SD* = 4.03 years), and the sample included 22 women and 8 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 Muenster.

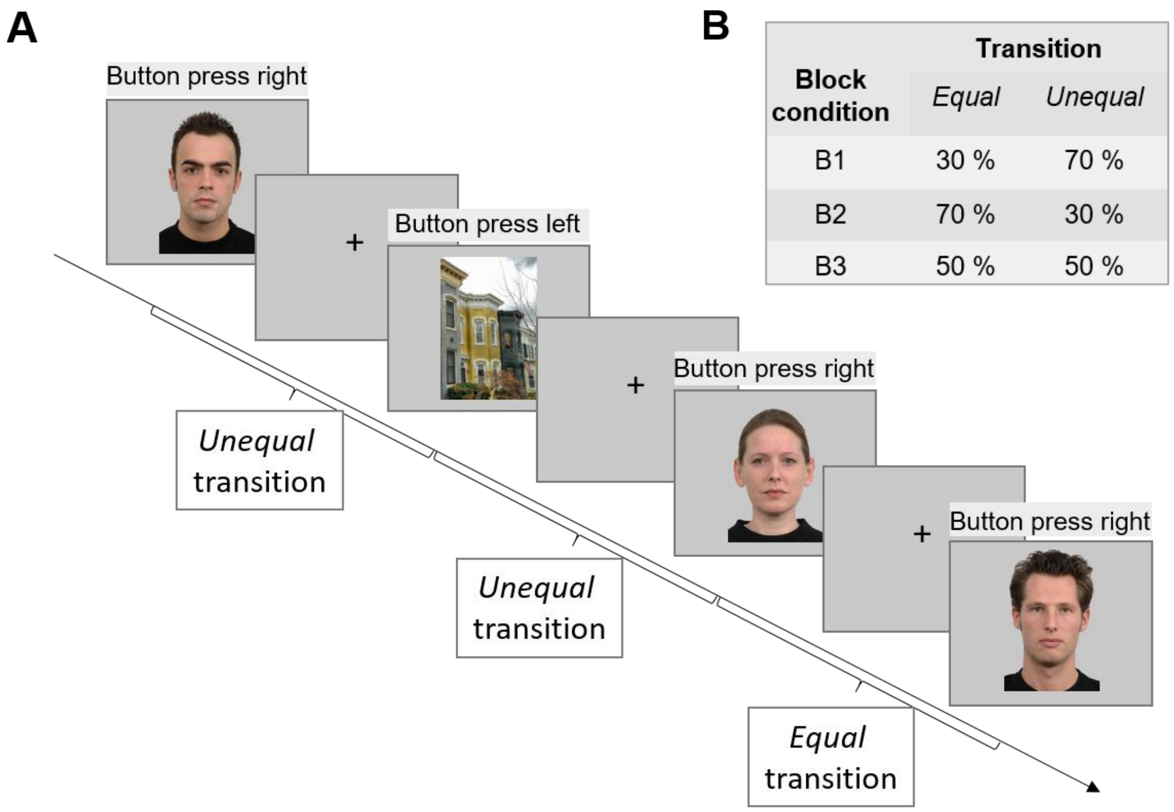
## Stimuli

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 Face 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 GIMP. All photographs were displayed in the middle of a uniform grey background.

## 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' behavioural 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 (Figure 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.



| Block condition | Transition |  |
|---|---|---|
|  | *Equal* | *Unequal* |
| B1 | 30 % | 70 % |
| B2 | 70 % | 30 % |
| B3 | 50 % | 50 % |

**Figure 1. Task design. (A)** Diagram of the experimental task. Each image was presented for 350 ms and the inter-stimulus-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 **(B)** 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 results in two different block orders: B1, B2, B3, B1, B2 or B2, B1, B3, B2, B1.

## 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 inter-trial-interval (ITI) of 3500, 4000, 4500, or 5000 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. 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 localiser task (Kanwisher et al., 1997). The localiser consisted of a 1-back task during which face and place stimuli were presented block-wise on a grey 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 x 1 x 1 mm) in a 256 mm field of view (FOV) (256 x 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 blood oxygenation level-dependent (BOLD) images were acquired parallel to the anterior commissure/posterior commissure line using a T2*-weighted gradient echo planar imaging (EPI) sequence (64 x 64 matrix, 192 mm FOV, 90° flip angle, TR = 2000 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 x 3 x 4 mm. All visual images were projected on a screen behind the scanner bore. Stimuli were presented in the centre 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 behavioural 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 minimise motion artefacts, foam paddings were applied around the participants’ head; whereas earplugs as well as noise-cancelling headphones were provided to reduce scanner noise.

# Data analysis

## Behavioural data analysis

The behavioural data analysis was conducted using Matlab. The focus of the analysis was on response times and accuracy (percentage of correct responses, CR), which were measured to examine participants' performance in the task. For the analysis, trials without a response and trials with incorrect responses were categorised as error trials. To establish a time frame for valid responses, a maximum response window of 1500 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.

For each participant, mean response times (RTs) and CR were entered as dependent variable into two 2 x 3 repeated measures ANOVA with the factors: Transition (equal, unequal) and Expectation (expected, neutral, unexpected). We report two-tailed *p*-values for the two main effects as well as their interaction. Afterwards, planned two-tailed *t*-tests evaluate the expectation effect to assess whether participants possessed a robust knowledge of the contextual regularities of the task which 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 R2022a; The MathWorks Inc., Natick, MA, USA). All functional images were slice scan time corrected and spatially realigned to the first volume for each participant. Structural images were co-registered to the mean functional image and then segmented into native space tissue components. Then, functional images (resampled at 3 $mm^3$) were spatially normalised to the standard Montreal Neurological Institute (MNI) template using tissue segmented T1-weighted anatomical scans and spatially smoothed with a Gaussian kernel of full width half-maximum (FWHM) of 6 $mm^3$.

The statistical model of the main task consisted of 12 regressors coding for onset time and duration (1 s) of each trial of each experimental condition (face-unequal-expected, face-equal-expected, face-unequal-unexpected, face-equal-unexpected, face-equal-neutral, face-unequal-neutral, place-unequal-expected, place-equal-expected, place-unequal-unexpected, place-equal-unexpected, place-equal-neutral, place-unequal-neutral). Additionally, six regressors for the motion parameters (three translations and three rotations) and one regressor coding for the error trials were entered to the model. The stimulus functions were then convolved with a canonical hemodynamic response function (HRF) and regressed against the observed fMRI data to form our condition regressors. To obtain FFA and PPA group ROIs, we convolved the block regressors coding for onsets and durations of face and place blocks of the localiser task with the canonical HRF. For each participant, we contrasted face versus place blocks for FFA activation and place versus face blocks for PPA activation. The resulting contrast images were then employed in a second-level random-effects group analysis. Subsequently, we used the probabilistic masks of FFA and PPA from the Anatomy toolbox (Lorenz et al., 2017) to restrict the radius of the group ROIs. As in our previous study (Schliephake et al., 2021), we decided to only include the right FFA and right PPA to our analyses, since it has been shown that face images are dominantly processed by the right hemisphere (Bentin 1996; Rossion 2014).

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 non-preferentially processed the omitted stimulus 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. 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. Subsequently, beta values of our two ROIs were subjected to a 2 x 3 x 2 repeated measures ANOVA with the factors: Transition (equal, unequal), Expectation (expected, neutral, unexpected) and Preference (preferred, non-preferred). We report two-tailed *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 corrected paired two-sided *t*-tests of the conditions neutral_preferred vs. unexpected_preferred, expected_preferred vs. unexpected_preferred, neutral_non-preferred vs. unexpected_non-preferred, and expected_non-preferred vs. unexpected_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 JZS Bayes (Rouder, Speckman, Sun, Morey, & Iverson, 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) was employed for all Bayesian tests. The interpretations of resultant Bayes factors, which quantify evidence supporting the null hypotheses (B01), adhered to established classification criteria (Lee & Wagenmakers, 2013).

# Results

## Behavioural data

In line with our hypotheses, the results showed a significant main effect of expectation ($F_{(1,29)} = 6.65$, $p = .005$) and no significant interaction effect between expectation and transition ($F_{(1,29)} = 0.87$, $p = .394$, $\mathrm{BF}_{01} = 4.99$) for the RT data. The main effect of transition was non-significant ($F_{(1,29)} = 3.5$, $p = .071$, $BF_{01} = 1.4$). The same pattern was observed for the CR data that revealed a main effect of expectation ($F_{(1,29)} = 19.91$, $p < .001$) as well as a non-significant interaction ($F_{(1,29)} = 1.73$, $p = .186$, $BF_{01} = 1.85$) and a non-significant transition main effect ($F_{(1,29)} = 3.56$, $p = .069$, $BF_{01} = 1.33$).

To further examine whether participants implicitly learned the transitional probabilities within and between stimulus categories, we employed paired samples *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 ($t_{(29)} = -4.21$, $p < .001$) and neutral events ($t_{(29)} = -2.91$, $p = 007$). Regarding CR, participants showed elevated accuracy levels for expected as compared to unexpected ($t_{(29)} = 5.97$, $p < .001$) and neutral events ($t_{(28)} = 3.42$, $p = .002$) (see Figure 2).

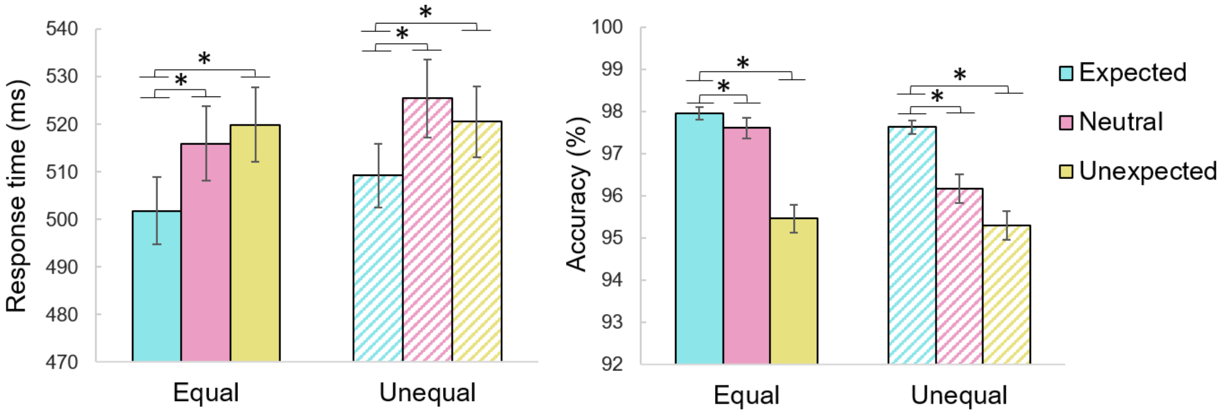


**Figure 2. Behavioural data.** Mean response times and mean percentage of correct responses (CR) for all experimental conditions.

## fMRI data

The aggregated FFA and PPA BOLD responses revealed no significant main effects of transition ($F_{(1,29)} = 0.40$, $p = .532$, $BF_{01} = 3.55$) or expectation ($F_{(1,29)} = 0.35$, $p = 0.707$, $BF_{01} > 999$) but a significant main effect of preference ($F_{(1,29)} = 188.760$, $p < .001$). As hypothesised, the results provided evidence for the null hypothesis that there is no interaction between transition and expectation ($F_{(1,29)} = 0.34$ , $p = .712$, $BF_{01} > 999$). In contrast, the effects of transition and expectation were dependent on whether the presented stimulus was a region’s preferred or non-preferred stimulus as revealed by a significant two-way interaction between transition and preference ($F_{(1,29)} = 64.62$ , $p < .001$) and between expectation and preference ($F_{(1,29)} = 4.152$ , $p = .021$).

The frequentist calculation of the three-way interaction among expectation, transition, and preference was found to be non-significant ($F_{(1,29)} = 1.58$ , $p = .215$). The Bayesian equivalent, however, revealed positive evidence in favour of the alternative hypothesis that there is a three-way interaction between the three factors ($BF_{01} = 0.016$). 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 samples *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 ($t_{(29)} = -3.82$, $p < .001$) and non-preferred ($t_{(29)} = -2.71$, $p = .011$) stimuli.

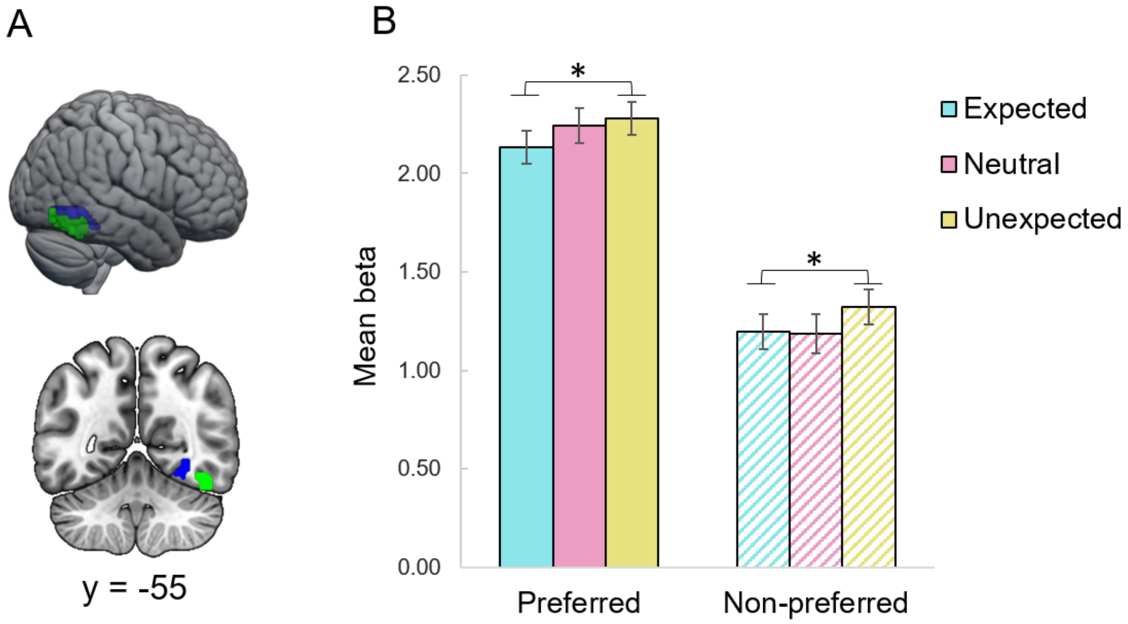


**Figure 3. Region of interest (ROI) fMRI data. (A)** Right FFA and PPA ROIs used for beta value extraction. **(B)** Mean group activation estimates (*β* parameters +/- SEM) for each condition of the experiment.

The frequentist *t*-tests comparing the unexpected and neutral conditions did not reach significance for the preferred ($t_{(29)} = 0.18$, $p = .86$) as well as the non-preferred condition ($t_{(29)} = 0.68$, $p = .50$). To further examine the evidence, a Bayesian *t*-test was conducted, revealing a Bayes factor $BF_{10} = 3.22$ for the preferred and $BF_{10} = 355.78$ for the non-preferred condition suggesting positive evidence for the alternative hypothesis of a difference between unexpected and neutral trials (see Figure 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 (e.g. a face) and unexpected omissions from another category (e.g. a place). Participants implicitly learned the varying probability of stimulus category transitions that were either equal (e.g. face – face) or unequal (e.g. 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 (i.e., prediction errors).

The behavioural 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 optimise 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 (i.e. decreases for expected trials) they could not show that these effects were independent of the stimulus repetitions and alternations. 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 organised 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 Kovacs (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 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), inferior frontal brain regions (Wig, Grafton, Demos, & Kelley, 2005, Horner, 2012), and the hippocampus (Hindy et al.; Kok et al., 2018).

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 (e.g. in the FFA) for unexpectedly appearing stimuli as compared to expected stimuli, and negative prediction errors, represented by elevated activation levels (e.g. 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.

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 50:50 and participants could not form strong expectations about which stimulus category would appear next (Grotheer & 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 addition of the neutral condition allowed for the identification of activation increases specifically associated with unexpected stimuli.

We expected to show that prediction error signals represented by activation increases would still be evident when comparing unexpected to neutral trials. While this contrast did not reach significance for both unexpected omissions and unexpected presentations of stimuli for the frequentist analyses, the Bayesian equivalent revealed positive and strong evidence in favour of higher activation levels for unexpected as compared to neutral events. It is plausible, for the frequentist analyses, the difference between the neutral and unexpected condition 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 multi-modal imaging approaches. These methods can capture the temporal dynamics and network-level interactions associated with these prediction error signals. Additionally, incorporating computational modelling approaches could help refine our understanding of the precise computations and learning mechanisms involved in processing co-occurring prediction errors.

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

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. http://dx.doi.org/10.1016/j.neuropsychologia.2016.02.003.

Bentin S. 1996. 1996_Bentin_Et_Al.Pdf. :551–565.

Egner T, Monti JM, 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.

Eliades SJ, 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(4 PART II):6–9. doi:10.1016/s1053-8119(18)31174-1.

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 HK, Petersen A V., Leinweber M, Keller GB. 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.

Grotheer M, Kovács G. 2015. The relationship between stimulus repetitions and fulfilled expectations. Neuropsychologia. 67:175–182.

doi:10.1016/j.neuropsychologia.2014.12.017. http://dx.doi.org/10.1016/j.neuropsychologia.2014.12.017.

Haxby J V, Hoffman EA, Gobbini MI. 2000. 08_Haxbytics2000. Trends Cogn Sci 4223–233. 4(6):223–233.

Henson RNA. 2003. Neuroimaging studies of priming. Prog Neurobiol. 70(1):53–81. doi:10.1016/S0301-0082(03)00086-8.

Hosoya T, Baccus SA, Meister M. 2005. Dynamic predictive coding by the retina. Nature. 436(7047):71–77. doi:10.1038/nature03689.

Keller GB, Mrsic-Flogel TD. 2018. Predictive Processing: A Canonical Cortical Computation. Neuron. 100(2):424–435. doi:10.1016/j.neuron.2018.10.003. 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(11). doi:10.1371/journal.pcbi.1000209.

Kovács G, Iffland L, Vidnyánszky Z, Greenlee MW. 2012. Stimulus repetition probability effects on repetition suppression are position invariant for faces. Neuroimage. 60(4):2128–2135. doi: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):1–8. doi:10.1038/s41598-018-31091-4.

Langner O, Dotsch R, Bijlstra G, Wigboldus DHJ, Hawk ST, van Knippenberg A. 2010. Presentation and validation of the radboud faces database. Cogn Emot. 24(8):1377–1388. doi:10.1080/02699930903485076.

Meyer T, Olson CR. 2011. Statistical learning of visual transitions in monkey inferotemporal cortex. Proc Natl Acad Sci U S A. 108(48):19401–19406. doi:10.1073/pnas.1112895108.

Den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE. 2009. A dual role for prediction error in associative learning. Cereb Cortex. 19(5):1175–1185. doi:10.1093/cercor/bhn161.

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

Rao RPN, Ballard DH. 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. E6(2):258. doi:10.2741/706.

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(February):118028. doi:10.1016/j.neuroimage.2021.118028. https://doi.org/10.1016/j.neuroimage.2021.118028.

Summerfeld C, Wyart V, Johnen VM, de Gardelle V. 2011. Human scalp electroencephalography reveals that repetiti suppression varies with expectation. Front Hum Neurosci. 5(JULY):1–13. doi: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(9):1004–1006. doi: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. doi: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(39):13389–13395. doi:10.1523/JNEUROSCI.2227-12.2012.

Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM. 2010. Implicit perceptual anticipation triggered by statistical learning. J Neurosci. 30(33):11177–11187. doi: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(11):3665–3678. doi:10.1523/JNEUROSCI.5003-11.2012.

Wig GS, Grafton ST, Demos KE, Kelley WM. 2005. Reductions in neural activity underlie behavioral components of repetition priming. Nat Neurosci. 8(9):1228–1233. doi:10.1038/nn1515.

# 3. General Discussion and Future Research

## 3.1 Summary of the presented studies

The main objective of the presented studies was to explore various temporal and spatial aspects related to the anticipatory mechanisms involved in visual perception and to investigate how these preparatory mechanisms are related to simultaneous positive and negative prediction error processing.

In *Study 1*, we aimed at answering the question whether the formation of expectations can be confined within a specific timeframe. We employed an associative learning paradigm, where participants acquired implicit knowledge of re-occurring stimulus pairs that were hidden in the stimulus sequence through probabilistic learning. During the task, participants had to press a button as fast and accurately as possible indicating whether the presented image depicted a female or male person. Making use of the temporal precision of EEG, we scrutinised alpha/beta power and the N170 response to pinpoint the on- and offset of top-down processes related to expectation formation. Consistent with previous investigations, we found a diminished response in the N170 and increased gamma-band power after the onset of the expected image. This is thought to be induced by reduced bottom-up processes for expected as compared to random trials.  Moreover, we found the beginning of anticipatory processes to start right after the presentation of the predictive image. The interval between the onset of the predictive image and the expected image was characterised by a continuous increase in alpha/beta power. The latest peak of this increase correlated positively with the modulation of the N170 amplitude for expected as compared to random images. These findings support the idea that information about the expected image was provided by early top-down processes. In summary, *Study 1* yielded evidence with regard to the temporal dynamics of early top-down processes before the onset of the expected events.

In *Study 2*, the primary focus was on examining positive and negative prediction error signals using fMRI. Specifically, we intended to demonstrate that the neural processing of both the omission of an expected stimulus and the presentation of another unexpected stimulus can and often does coincide. We employed a speeded forced-choice fMRI paradigm during which participants were exposed to cross-category stimulus transitions with a pre-defined probabilistic distribution. Via a button press, participants had to respond to whether the presented image was a face or a place. The results indicate a behavioural facilitation effect, represented by decreased response times and increased accuracy levels for expected versus unexpected stimuli. Furthermore, we observed negative prediction error signals with differential activation of FFA and PPA during unexpected omissions of places and faces, respectively. Additionally, a psychophysiological interaction (PPI) analysis uncovered positive prediction error signals that manifested as functional coupling between the right inferior frontal gyrus (IFG) and FFA or PPA during unexpected presentations of face and place images. The PPI results suggest the involvement of a frontal network updating the current generative model after the presentation of an unexpected stimulus.

As opposed to our hypotheses, we found increased activation levels for expected trials as compared to unexpected trials for the region of interest’s (ROI’s) preferred stimuli. Nevertheless, this could be due to the fact that the stimulus transitions were not distributed equally. Since unequal stimulus category transitions were always more likely, it could be the case that the stimulus alternations during expected trials increased the brain activation due to low-level sensory prediction errors.

This open question was addressed in *Study 3* where we manipulated stimulus expectation and stimulus repetition orthogonal to each other using an adapted version of the experimental design employed in *Study 2*. Moreover, we introduced a neutral stimulus condition in which the probability of equal and unequal stimulus category transitions was 50:50. This allowed us to ensure that activation increases found in the unexpected conditions were not due to activation decreases during expected trials but actually the result of prediction errors, as hypothesised. In *Study 3,* we could replicate the behavioural results from *Study 2*. Comparing expected with unexpected trials, we found significantly increased accuracy levels and shortened response times. Additionally, the analysis of beta estimates extracted from the right FFA and PPA revealed simultaneous activations of negative and positive prediction error signals (increases in activation) for unexpected omissions and unexpected presentations of face and place stimuli, respectively. The Bayesian analyses of the beta estimates yielded further evidence that these increases still persisted when comparing the unexpected condition with the neutral condition. Importantly, our findings were independent of stimulus repetition effects.

## 3.2 Consequences of the pre-activation of expected sensory input

In the realm of hierarchical predictive processing models, the brain is postulated to operate as predictive engine, constantly generating and refining predictions about incoming sensory information based on prior knowledge (Clark, 2013; Friston, 2005). It has been suggested that the brain pre-activates existing knowledge about upcoming stimuli to facilitate subsequent processing (Kok et al., 2017; Walsh et al., 2020). The idea of pre-activation is in line with the key finding of *Study 1* where we could show enhanced alpha/beta power for expected as compared to unexpected face stimuli. These signals emerged right after the beginning of the predictive stimulus and prevailed until the expected stimulus was presented. Such pre-stimulus alpha/beta power enhancement could be linked to the anticipatory activation of already acquired knowledge, as indicated in recent studies (Brodski-Guerniero et al., 2017; Mayer et al., 2016).

Further evidence was provided by fMRI studies using omission designs, which have shown that the omission of an expected stimulus results in stimulus-specific activation patterns in those sensory areas processing the withheld stimulus (Kok et al., 2012; SanMiguel et al., 2013; Todorovic & de Lange, 2012). Additionally, the activation level in response to the omission is generally higher than the response to the expected and actually presented stimuli. Nevertheless, these studies as well as the results from *Study 1* could not clarify whether the activation increases represented a general surprise signal or had representational content. Furthermore, omission studies were unable to determine whether the stimulus templates appearing after the omission of expected stimuli represented pre-stimulus expectations or stimulus-specific error signals triggered by the unexpected omission. This limitation arises because these studies only examined activity following the omission, not during the period preceding it.

A more recent study by Kok et al. (2017) employed MEG and multivariate decoding methods to investigate the temporal specificities and representational content of neuronal expectation signatures. Their experimental protocol involved an auditory cue that predicted the orientation of an upcoming grating stimulus with varying validity. Additionally, in 25% of the trials, no grating stimulus was presented after the auditory cue. The results showed that expectations can indeed induce sensory templates involving representational content of the expected stimulus even before the onset of the expected stimulus. Furthermore, the researchers could display that the expectation signal remained active throughout the trials and lasted into the period of the presentation of the expected stimulus. This finding by Kok et al. (2017) is supported by *Study 1*, which shows that the alpha/beta power increase was not only evident when comparing predictive and random images (i.e. the period before expected stimulus appeared) but also for expected and random images (i.e. period after expected stimulus appeared).

Another finding that can be explained with the pre-activation of expected sensory input, is the negative prediction error effect that we revealed in *Study 2*. We defined the negative prediction error as increased activation to an unexpectedly omitted stimulus when compared to an expected stimulus. Assuming that expectations pre-activate neuronal templates of expected stimuli, the negative prediction error signal might result from pre-activation of sensory units (Wacongne et al., 2012).

The pre-activation of expected sensory input can also be related to other parts of the results of *Study 2,* where we found elevated activation levels for expected as compared to unexpected preferred stimuli. According to "sharpening" models (Press & Yon, 2019; Yon et al., 2018), expectations trigger activation enhancement of predicted sensory input while suppressing unexpected units. At the population level, sharpening models suggest that the contrast of activity patterns elicited by a specific stimulus will increase. This occurs because the relative suppression of the least-selective voxels amplifies the differences in activity between voxels, resulting in a more distinct or "sharpened" representation (de Lange et al., 2018). Concerning the study design of *Study 2,* if the participant expected to see a specific stimulus category, for example a face, the expectation pre-activated sensory units that were tuned toward that stimulus. This resulted in a relative pre-activation of the FFA and a suppression of PPA neurons, giving rise to the observed expectation enhancement.

Taken together, the findings provide compelling evidence for the pre-activation of expected sensory input in the brain. However, further research is warranted to fully understand the extent to which pre-activated knowledge contributes to the generation of prediction errors.

## 3.3 Expectation enhancement vs. expectation suppression

While pre-activation and the sharpening account offer plausible explanations for the heightened activation levels observed for expected stimuli compared to unexpected ones in *Study 2*, it is crucial to recognise that additional factors may have influenced the seemingly divergent outcomes of expectation modulation of *Study 2* and *Study 3*. Specifically, the differences arise from our findings of increased activation for expected events (expectation enhancement) in *Study 2*, contrasted with the reversed pattern (expectation suppression) observed in *Study 3*.

Firstly, in *Study 2*, we did not manipulate stimulus repetition and stimulus expectation orthogonally to each other. This means, that, during the task, alternating stimulus categories (i.e. face – place; place – face) were always more likely than repeating stimulus categories (i.e. face – face; place – place). Therefore, it is conceivable that low-level prediction errors resulting from alternating stimulus categories in *Study 2* gave rise to increased activation levels for the expected stimuli. Relatedly, the interaction of expectation with other top-down factors, such as attention, can exert a significant influence on the direction and magnitude of expectation effects in BOLD responses (Gordon et al., 2019; Rungratsameetaweemana & Serences, 2019; Summerfield & Egner, 2009). In *Study 2*, participants had to respond via button presses to the stimulus category that was presented to them. Therefore, alternating stimulus categories were meaningful to the task requiring participants to switch between the two buttons, possibly resulting in increased attentional engagement. This increased attentional load might have countered the expected suppression effect, leading to the observed activation increases (Jiang et al., 2013; Kok et al., 2012).

Another important factor that needs to be considered, lies in the temporal and spatial limitations of fMRI, which captures the collective activity of neuronal groups without discerning the nuanced timing of neuronal suppression and enhancement effects. For example, Segaert et al. (2013) suggested that repetition effects may be characterised by shifts from enhancement to suppression observed over time, as evidenced by multiple EEG and MEG studies (Marinkovic et al., 2003; Petit et al., 2006). This notion finds support in a newly developed two-process model proposed by Press and Yon (2019). The authors argue that initially, our perception is always biased towards what is most probable based on our expectations. This stage helps to quickly make sense of the most probable scenario. Subsequently, the perceptual system dynamically adjusts to ongoing stimuli leading to suppression of expected stimuli to maintain cognitive flexibility. As indicated before, the poor temporal resolution of fMRI might obscure the rapid dynamics of initial enhancements and subsequent suppressions of activity towards expected events. It is important to note that, despite its limitations in temporal resolution, fMRI remains indispensable in cognitive neuroscience due to its non-invasive nature, superior spatial specificity, and capacity to map large-scale brain networks.

## 3.4 Are positive and negative prediction errors processed in parallel?

After exploring the various factors influencing the formation of expectations, the next logical step was to examine the implications of expectation violations. *Study 2* and *Study 3* aimed to elucidate how different types of prediction error signals might be processed simultaneously through potentially separable circuits within the cortical hierarchy. In *Study 2* we found evidence for negative prediction errors in our univariate analysis where we compared hemodynamic changes for unexpected as compared to expected stimuli within the brain area non-preferentially processing that stimulus. Employing a PPI analysis, we examined whether activation in the right IFG was functionally correlated with positive and/or negative prediction error signals. We found IFG coupling with the PPA for unexpected places and the FFA for unexpected faces. Importantly, however, we did not find a positive association between activation of the right IFG and negative prediction error signals, implying that unpredicted presentations of stimuli drive coupling with the higher-level areas but omissions of expected stimuli do not. In *Study 3*, the results revealed evidence for co-occurring positive and negative prediction error signals during the univariate analysis for unexpected as compared to expected trials. Using Bayesian analyses we could further support our hypothesis that positive and negative prediction errors resulting from unexpected presentations and omissions of stimuli, respectively, would still be evident when comparing to a neutral category instead of the expected condition.

The simultaneous activation of two different types of prediction error signals suggests that positive and negative error computations within the brain are carried out by distinct processing circuits (Keller & Mrsic-Flogel, 2018; Rao & Ballard, 1999). According to Keller and Mrsic-Flogel (2018), different types of neurons are responsible for the processing of positive and negative prediction error signals at the cellular level. First, Type 1 neurons process positive prediction errors by computing the difference between the prediction and the actual sensory input (Sensory Input – Prediction). Furthermore, these positive prediction error neurons provide bottom-up excitation to the area that is associated with processing the unexpected stimulus, thus activating a new cluster of neurons. Second, Type 2 neurons compute negative prediction errors via the reversed computation (Prediction – Sensory Input), having an opposite effect as positive prediction error neurons on their targets. It is suggested that negative prediction error neurons engage bottom-up inhibition to the areas that are processing the omitted stimulus, thereby suppressing the current internal representation neurons.

This account can also explain ambiguities that came up with the explanation of negative prediction error signals and omission responses (Walsh, 2020, see also section 3.2). According to predictive coding models, predictions are simply subtracted from bottom-up sensory input.

This would lead to negative firing rates for omitted stimuli and the implication that omission responses mainly reflect the activity of representation neurons rather than prediction error neurons (Wacongne et al., 2012). According to Keller and Mrsic-Flogel (2018), however, negative prediction errors are computed in a reversed manner; by subtracting the sensory input from the prediction. Additionally, it has been suggested that omission responses are composed of both the activity in representation neurons depicting the prediction of the omitted stimulus and the prediction error response to the mismatch between the expected and actual sensory input (Walsh, 2020).

If error processing is indeed carried out by two separable prediction error circuits, it is conceivable that both positive and negative prediction errors can be processed in parallel. In terms of hierarchical processing, however, it might also be plausible that negative prediction errors are processed earlier in the processing stream, as they rely on pre-activated sensory expectations (Stange, 2023). Therefore, information about the omitted stimulus is already available before onset of the unexpected stimulus (or omission of the expected stimulus). Nonetheless, evidence about the time course of positive and negative prediction errors is scarce. Even though there are studies that investigated temporal aspects of negative prediction errors (omission responses) or positive prediction errors, these results are hardly comparable with each other due to several factors. Looking at EEG studies, for example, the time course of error responses depends on i) the location in which these responses were recorded, ii) the sensory modality that was investigated (e.g. visual or auditory), and iii) the experimental procedure that was employed. Moreover, the poor temporal resolution of fMRI studies that investigated different types of prediction errors make it hard to draw any specific conclusions on the time course of positive and negative prediction errors. Therefore, future research is needed to explore the temporal dynamics of co-occurring positive and negative prediction error signals in one experimental design using a method that offers sufficiently high temporal and spatial resolution.

## 3.5 The potential interplay of low-level and high-level prediction error signals

Theoretical frameworks concur that neurons involved in computing prediction errors and those responsible for generating internal models exist at all levels of the processing hierarchy (Epstein, 1993). However, consensus is lacking with regard to how predictions from independent models at different stages of the sensory hierarchy are integrated during the computation of prediction errors (Tabas & von Kriegstein, 2021, 2024).

In *Study 3,* we set out to examine whether positive and negative prediction error signals are processed independently of stimulus category repetitions and alternations. Using frequentist and Bayesian analysis, the behavioural and fMRI results clearly showed that, in our study, the effects resulting from violations of expectations were independent of stimulus repetition effects, supporting previous research investigating stimulus repetition and expectation orthogonal to each other (Grotheer & Kovács, 2016; Todorovic & de Lange, 2012). More specifically, we found increased responses to unexpected trials for both preferred and non-preferred stimuli irrespective of whether the unexpected stimulus was a repetition or alternation of the previous stimulus category. However, the mechanisms underlying expectation and repetition effects might still be comparable as the effect of alternating stimuli has been shown to be similar to the effect of a contextual prediction errors (Amado et al., 2016; Grotheer & Kovács, 2016). Alternating stimuli result in increased responses as compared to repeating stimuli just like unexpected stimulus presentations result in elevated responses when compared with expected stimuli. Therefore, it has been argued that both stimulus repetition and expectation effects are prediction errors that manifest on different time scales and hierarchy levels (Auksztulewicz & Friston, 2016; Garrido et al., 2009; Kiebel et al., 2008).

The idea that predictive mechanisms could emerge across multiple, hierarchically organised levels was initially introduced in a study by Wacongne et al., (2011). The results of their auditory novelty paradigm in which standard tones, deviant tones or tone omissions were presented, could show that at least two different levels of top-down predictions influenced the MEG signal of their participants. Other studies investigating functional connectivity suggest that repetition effects might reflect changes in intrinsic (within a brain region) and extrinsic (between brain regions) neural connectivity (Ewbank et al., 2013; Garrido et al., 2009; Kohn & Movshon, 2004). While changes within a region are proposed to reflect an increase in precision of the prediction, changes in connectivity between regions are thought to reflect increases in predictability of repetitions (Garrido et al., 2009). The notion that expectation effects probably mirror computations at higher levels is supported by Grotheer and Kovacs (2015) who found Brodmann Area (BA) 7 to play an important role in the generation of expectations. Moreover, the PPI analysis of *Study 2* revealed functional coupling between the right IFG and positive prediction error signals. Consequently, higher-level prediction errors likely influence neuronal responses in sensory cortices via top-down connections. Regarding lower-level prediction errors, signals are possibly processed more locally in sensory areas (De Gardelle et al., 2013).

Another factor that supports the processing of repetition and expectation effects at different levels of the sensory hierarchy is the timing of both effects. Repetition effects have been shown to occur rapidly after stimulus onset usually starting between 40 and 60ms in the auditory domain (Todorovic and de Lange, 2012) and between 100 and 250ms in the visual domain (Henson et al., 2004; Schendan & Kutas, 2007; Summerfeld et al., 2011). Expectation effects, in contrast, can be detected somewhat later around 100-200ms after stimulus onset in the auditory domain and 300ms in the visual domain (Summerfield et al., 2011; Todorovic and de Lange, 2012).

Taken together, we found convincing evidence that stimulus repetition and stimulus expectation exert dissociable effects on their targets that can both be explained by hierarchical predictive processing mechanisms. Repetition effects might result from lower-level expectations that are consistently weighted towards the most recently encountered stimuli, expectation effects, as manipulated in our studies, are more influenced by the broader context, such as the probabilistic structure of the stimulus sequence, highlighting the importance of overall patterns rather than just immediate past occurrences (Auksztulewicz & Friston, 2016; Summerfield & De Lange, 2014).

In light of the current studies, one might also speculate that positive and negative prediction error signals are not restricted to contextual predictions and therefore higher levels of the sensory hierarchy. Instead, these signals might represent a general mechanism of error processing that exists at different levels of the hierarchy.

Concerning the repetition effects as investigated in our studies, however, we can also not exclude the possibility that these effects arise from neural adaptation leading to low-level changes in the responsiveness of neurons tuned to specific stimulus-features (Walsh et al, 2020). More research is needed to explore this option. One way of investigating whether repetition effects are due to expectations at lower-levels of the sensory hierarchy would be to detect representational and prediction error units at those levels that are processing stimulus alternations via prediction error neurons.

## 3.5 What's in a neutral condition?

In all three studies, we could show significant neuronal differences between expected and unexpected or random events. However, the main limitation of *Study 2* and other previous studies comparing expected with unexpected events is that this comparison cannot clarify whether the differences between both conditions are due to the increase of one condition or the decrease of the other condition (Amado et al., 2016; Kovacs and Vogels, 2014). Therefore, we introduced a "neutral" condition in *Study 3,* where an equal probability of repeating and alternating trials was contained and no clear expectations about the upcoming stimulus category could be made.

It is important to clarify what "neutral" with regard to our experimental design means. Our definition of the neutral condition followed previous suggestions by Amado et al. (2016) and Feuerriegel et al. (2021). We defined trials belonging to the neutral condition as trials that appeared with equal probability so that no strong or systematic expectation biases would be present that could favour one stimulus category over another. Moreover, we defined the "neutral" stimulus condition in relation to the two other conditions that were either expected (70% probability) or unexpected (30% probability).

In line with findings from the previous study by Amado et al., 2016, we revealed evidence for increases for unexpected when compared to neutral events, supporting the role of expectation violations in predictive coding processes (Friston, 2012). Nevertheless, this finding was only supported by our Bayesian analysis. The frequentist analysis revealed no significant difference between unexpected and neutral events. This underlines the need for more research and the question whether our "neutral" condition was perfectly suited as condition in which no expectations about the upcoming stimulus category could be made. Even though we tried to introduce a condition in which no clear expectations about the upcoming stimulus could be made, many aspects of the experiment were still much more predictable than in our naturalistic environment including for example the stimulus type, the presentation duration, the inter-stimulus-interval, or the trial structure (Feuerriegel et al., 2021). Moreover, although equal and unequal stimulus transitions were balanced across the whole block, during short episodes within a block, it was possible that one stimulus transition was slightly more likely than the other.

Another option of testing whether responses to unexpected omissions and presentations of stimuli represent positive and negative prediction error signals would be to incorporate a condition in which omissions are expected. As discussed previously, we assumed that the negative prediction error signal is composed of both pre-activation of internal representation neurons and error neurons. The comparison between expected and unexpected omissions would us, therefore, allow to differentiate between the top-down prediction (i.e. pre-activation of expected sensory input) and negative error processing, since the expected omission should not lead to error processing. Additionally, comparing unexpected and expected stimulus presentations should reveal positive error processing.

# 4. Critical Reflections and Future Directions

The three studies outlined in this thesis were designed to emphasise important components of the hierarchical predictive coding framework. We particularly focused on the temporal aspects of the origin of expectations and we addressed the open question whether positive and negative prediction errors co-exist. While several limitations of the presented studies have been highlighted in the discussion section already, the most important open questions will be discussed here giving rise to potential future studies where these limitations can be accounted for. In *Study 2* and *3*, we employed an experimental paradigm in which two distinct stimulus categories, namely face and place stimuli, triggered positive and negative prediction error processing. We made use of the advantage that both faces and places are processed within two different functional brain areas. Therefore, using the high spatial resolution of fMRI, we could reveal the activation of both areas resulting from two different types of prediction errors at the same time. Nevertheless, the relatively poor temporal resolution of fMRI might have resulted in a blurring of the timing of positive and negative prediction errors. This drawback highlights the need for future studies that explore the temporal dynamics of the different error signals. One possibility would be to combine EEG and fMRI so shed light on the question whether negative prediction errors are processed earlier than negative prediction errors as speculated before.

positive

A more general question that needs to be answered in the future is how different prediction and prediction error signals from various sources are eventually combined during perceptual inference. On the one hand, one potential scenario would be that all regions send their predictions to a common blackboard within, for example, the primary visual cortex for visual information or the primary auditory cortex for auditory information. On the other hand, the integration site may also vary depending on the level of detail provided by the prediction (de Lange et al., 2018). Other questions that remain open for debate concern the specific computational role of expectations and prediction errors. As discussed before, expected sensory input might be suppressed to save resources or, in turn, improve sensory representations to increase processing efficiency. It is also possible that a combination of both is used but if so, it is still a key gap in our knowledge how these goals might be aligned during neural processing.

# 5. Conclusion

In this thesis, we were the first to demonstrate that positive and negative prediction errors are not two phenomena that manifest in isolation but instead do occur simultaneously, a relationship that had not been investigated before. We could show that positive prediction errors seem to represent signals which drive the revision of the current internal model as response to the presentation of an unexpected visual stimulus. Negative prediction errors, in turn, result from the omission of an expected stimulus and are supposedly composed of a mixture of pre-activated sensory information about the expected stimulus and the error response resulting from the omission. We argue that both positive and negative prediction error signals are processed by two distinct neuronal error circuits in which one type of neurons processes positive prediction errors and the other type processes negative prediction error. This latter error signal could also be related to the findings from our first study in which we found evidence for sensory pre-activation as response to learned contingencies between presented stimuli. Importantly, we could show that this pre-activation started as early as the predictive cue and lasted into the period after the presentation of the expected image highlighting the facilitatory effect of expectation formation as hypothesised by predictive coding (Friston, 2009). Our results support the idea that the brain generates predictions about incoming sensory data based on prior knowledge and uses these predictions to effectively match them with actual sensory input. This process reduced the computational load in the brain, minimising processing time and highlighting any prediction errors, which are crucial for updating and refining the brain's current internal model about the external world leading to better predictions and adaptive behaviour.

While we have made significant progress in understanding how the brain employs predictive coding for perception, the intricates of the mechanisms involved are of such complexity that it may take more than the Earth's lifetime to thoroughly explore and comprehend every detail of them.

# References

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of Neuroscience, 30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

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. https://doi.org/10.1016/j.neuropsychologia.2016.02.003

Auksztulewicz, R., & Friston, K. (2016). Repetition suppression and its contextual determinants in predictive coding. *Cortex, 80*, 125–140. https://doi.org/10.1016/j.cortex.2015.11.024

Bar, M. (2009). The proactive brain: Memory for predictions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1235–1243. https://doi.org/10.1098/rstb.2008.0310

Barlow, H. B. (1961). Possible principles underlying the transformation of sensory messages. *Sensory communication*, *1*(01), 217-233.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. [Neuron. 2012] - PubMed - NCBI. *Neuron2*, *76*(4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038.Canonical

Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience, 29*(26), 8447–8451. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

Brodski-Guerniero, A., Paasch, G. F., Wollstadt, P., Özdemir, I., Lizier, J. T., & Wibral, M. (2017). Information-theoretic evidence for predictive coding in the face-processing system. *Journal of Neuroscience*, *37*(34), 8273–8283. https://doi.org/10.1523/jneurosci.0614-17.2017

Cao, L., Thut, G., & Gross, J. (2017). The role of brain oscillations in predicting self-generated sounds. *NeuroImage*, *147*(November), 895–903. https://doi.org/10.1016/j.neuroimage.2016.11.001

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2017). Busting Out: Predictive Brains, Embodied Minds, and the Puzzle of the Evidentiary Veil. *Nous*, *51*(4), 727–753. https://doi.org/10.1111/nous.12140

Craik, K. J. W. (1943). Physiology of colour vision. *Nature*, *151*(3843), 727-728.

De Gardelle, V., Waszczuk, M., Egner, T., & Summerfield, C. (2013). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cerebral Cortex*, *23*(9), 2235–2244. https://doi.org/10.1093/cercor/bhs211

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Den Ouden, H. E. M., Daunizeau, J., Roiser, J., Friston, K. J., & Stephan, K. E. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, *30*(9), 3210–3219. https://doi.org/10.1523/JNEUROSCI.4458-09.2010

Den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex*, *19*(5), 1175–1185. https://doi.org/10.1093/cercor/bhn161

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434. https://doi.org/10.1016/j.neuron.2012.01.010

Doherty, J. R., Rao, A., Mesulam, M. M., & Nobre, A. C. (2005). Synergistic effect of combined temporal and spatial expectations on visual attention. *Journal of Neuroscience, 25*(36), 8259–8266. https://doi.org/10.1523/JNEUROSCI.1821-05.2005

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

Epstein, W. (1993). The representational framework in perceptual theory. *Perception & Psychophysics*, *53*(6), 704–709. https://doi.org/10.3758/BF03211747

Ewbank, M. P., Henson, R. N., Rowe, J. B., Stoyanova, R. S., & Calder, A. J. (2013). Different neural mechanisms within occipitotemporal cortex underlie repetition suppression across same and different-size faces. *Cerebral Cortex*, *23*(5), 1073–1084. https://doi.org/10.1093/cercor/bhs070

Feuerriegel, D., Yook, J., Quek, G. L., Hogendoorn, H., & Bode, S. (2021). Visual mismatch responses index surprise signalling but not expectation suppression. *Cortex*, *134,* 16-29.

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

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Galloway, L. (2004). View from the top. *T and D*, *58*(4), 38-43+4. https://doi.org/10.2307/j.ctv21r3j2h.13

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., & Friston, K. J. (2009). Repetition suppression and plasticity in the human brain. *NeuroImage*, *48*(1), 269–279. https://doi.org/10.1016/j.neuroimage.2009.06.034

Gordon, N., Tsuchiya, N., Koenig-Robert, R., & Hohwy, J. (2019). Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. *PLoS Biology, 17*(4), 1–29. https://doi.org/10.1371/journal.pbio.3000233

Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *290*(1038), 181-197.

Grotheer, M., & Kovács, G. (2014). Repetition probability effects depend on prior experiences. *Journal of Neuroscience, 34*(19), 6640–6646. https://doi.org/10.1523/JNEUROSCI.5326-13.2014

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

Grotheer, M., & Kovács, G. (2016). Can predictive coding explain repetition suppression? *Cortex*, *80*, 113–124. https://doi.org/10.1016/j.cortex.2015.11.027

Henson, R. N., Rylands, A., Ross, E., Vuilleumeir, P., & Rugg, M. D. (2004). The effect of repetition lag on electrophysiological and haemodynamic correlates of visual object priming. *NeuroImage*, *21*(4), 1674–1689. https://doi.org/10.1016/j.neuroimage.2003.12.020

Herholz, S. C., Lappe, C., & Pantev, C. (2009). Looking for a pattern: An MEG study on the abstract mismatch negativity in musicians and nonmusicians. *BMC Neuroscience, 10*, 1–10. https://doi.org/10.1186/1471-2202-10-42

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*(APR), 1–14. https://doi.org/10.3389/fpsyg.2012.00096

Huang, Y., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593. https://doi.org/10.1002/wcs.142

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, *195*(1), 215–243. https://doi.org/10.1113/jphysiol.1968.sp008455

Huotilainen, M., Winkler, I., Alho, K., Escera, C., Virtanen, J., Ilmoniemi, R. J., Jääskeläinen, I. P., Pekkonen, E., & Näätänen, R. (1998). Combined mapping of human auditory EEG and MEG responses. *Electroencephalography and Clinical Neurophysiology - Evoked Potentials*, *108*(4), 370–379. https://doi.org/10.1016/S0168-5597(98)00017-3

Jaramillo, S., & Zador, A. M. (2011). The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nature Neuroscience*, *14*(2), 246–253. https://doi.org/10.1038/nn.2688

Jiang, J., Summerfield, C., & Egner, T. (2013). Attention sharpens the distinction between expected and unexpected percepts in the visual brain. *Journal of Neuroscience*, *33*(47), 18438–18447. https://doi.org/10.1523/JNEUROSCI.3308-13.2013

Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., ... & Young, A. W. (2017). Temporal and spatial localization of prediction-error signals in the visual brain. *Biological psychology*, *125*, 45-57.

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive Processing: A Canonical Cortical Computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology*, *4*(11). https://doi.org/10.1371/journal.pcbi.1000209

Kohn, A., & Movshon, J. A. (2004). Adaptation changes the direction tuning of macaque MT neurons. *Nature Neuroscience*, *7*(7), 764–772. https://doi.org/10.1038/nn1267

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

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Kovács, G., & Vogels, R. (2014). When does repetition suppression depend on repetition probability? *Frontiers in Human Neuroscience*, *8*(SEP), 1–3. https://doi.org/10.3389/fnhum.2014.00685

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, *20*(7), 1434. https://doi.org/10.1364/josaa.20.001434

Limanowski, J., Litvak, V., & Friston, K. (2020). Cortical beta oscillations reflect the contextual gating of visual action feedback. *NeuroImage*, *222*(June), 117267.

https://doi.org/10.1016/j.neuroimage.2020.117267

Marr, D., & Aina, L. V. (1982). R ep resen tatio n and recognition of the m ovem ents of shapes. *Proc. R. Soc. Loud. B, 214*(November 1980), 501–524. https://royalsocietypublishing.org/

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal Dynamics of Modality-Specific and Supramodal Word Processing The importance of the left anterior temporal cortex for. *Neuron, 38*, 487–497. http://www.neuron.

Mayer, A., Schwiedrzik, C. M., Wibral, M., Singer, W., & Melloni, L. (2016). Expecting to see a letter: Alpha oscillations as carriers of top-down sensory predictions. *Cerebral Cortex*, *26*(7), 3146–3160. https://doi.org/10.1093/cercor/bhv146

Meyer, T., & Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(48), 19401–19406. https://doi.org/10.1073/pnas.1112895108

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125–12130. https://doi.org/10.1073/pnas.1101042108

Petit, J. P., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2006). On the time course of letter perception: A masked priming ERP investigation. *Psychonomic Bulletin and Review, 13*(4), 674–681. https://doi.org/10.3758/BF03193980

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 1–15. https://doi.org/10.1167/15.8.13

Poggio, T., & Riesenhuber, M. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–1025. http://www.nature.com/doifinder/10.1038/14819%5Cnpapers3://publication/doi/10.1038/14819

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Press, C., & Yon, D. (2019). Perceptual Prediction: Rapidly Making Sense of a Noisy World. *Current Biology*, *29*(15), R751–R753. https://doi.org/10.1016/j.cub.2019.06.054

Ran, G., Zhang, Q., Chen, X., & Pan, Y. (2014). The effects of prediction on the perception for own-race and other-race faces. *PLoS One*, *9*(11), e114011.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87. https://doi.org/10.1038/4580

Richter, D., Kietzmann, T. C., & Lange, F. P. de. (2023). High-level prediction errors in low-level visual cortex. *BioRxiv*, 2023.08.21.554095. https://www.biorxiv.org/content/10.1101/2023.08.21.554095v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.08.21.554095v1.abstract

Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, *29*, 181–186. https://doi.org/10.1016/j.copsyc.2019.03.014

SanMiguel, I., Saupe, K., & Schröger, E. (2013). I know what is missing here: Electrophysiological prediction error signals elicited by omissions of predicted “what” but not “when.” *Frontiers in Human Neuroscience*, *7*(JUL), 1–10. https://doi.org/10.3389/fnhum.2013.00407

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*(5), 734–749. https://doi.org/10.1162/jocn.2007.19.5.734

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.

Seriès, P., & Seitz, A. R. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*(OCT), 1–14. https://doi.org/10.3389/fnhum.2013.00668

Spaak, E., Fonken, Y., Jensen, O., & De Lange, F. P. (2016). The Neural Mechanisms of Prediction in Visual Search. *Cerebral Cortex*, *26*(11), 4327–4336. https://doi.org/10.1093/cercor/bhv210

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Frontiers in Computational Neuroscience*, *2*(OCT), 1–8. https://doi.org/10.3389/neuro.10.004.2008

Spratling, M. W. (2017). A review of predictive coding algorithms. *Brain and Cognition*, *112*, 92–97. https://doi.org/10.1016/j.bandc.2015.11.003

Stange, L. (2023). *Neural Correlates of Top-down and Bottom-up Communication in Sensory Networks*. *July*.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104. https://doi.org/10.1037/xge0000109

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural Networks*, *18*(3), 225–230. https://doi.org/10.1016/j.neunet.2004.12.004

Summerfeld, C., Wyart, V., Johnen, V. M., & de Gardelle, V. (2011). Human scalp electroencephalography reveals that repetiti suppression varies with expectation. *Frontiers in Human Neuroscience*, *5*(JULY), 1–13. https://doi.org/10.3389/fnhum.2011.00067

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Tabas, A., & von Kriegstein, K. (2021). Adjudicating Between Local and Global Architectures of Predictive Processing in the Subcortical Auditory Pathway. *Frontiers in Neural Circuits*, *15*(March), 1–14. https://doi.org/10.3389/fncir.2021.644743

Tabas, A., & von Kriegstein, K. (2024). Multiple concurrent predictions inform prediction error in the human auditory pathway. *Journal of Neuroscience*, *44*(1). https://doi.org/10.1523/JNEUROSCI.2219-22.2023

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/JNEUROSCI.2227-12.2012

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: An MEG study. *Journal of Neuroscience, 31*(25), 9118–9123. https://doi.org/10.1523/JNEUROSCI.1425-11.2011

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

Van Pelt, S., Heil, L., Kwisthout, J., Ondobaka, S., van Rooij, I., & Bekkering, H. (2016). Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Social Cognitive and Affective Neuroscience*, *11*(6), 973–980. https://doi.org/10.1093/scan/nsw017

Wacongne, C., Changeux, J. P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *Journal of Neuroscience*, *32*(11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604. https://doi.org/10.1038/nn858

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3593–3598. https://doi.org/10.1073/pnas.1120118109

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

Von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Voss.

# Abbreviations

| | |
|---|---|
| BA | Brodmann area |
| BOLD | blood-oxygenation-level dependent |
| EEG | electroencephalography |
| ERP | event-related potential |
| FFA | fusiform face area |
| fMRI | functional magnetic resonance imaging |
| IFG | inferior frontal gyrus |
| MEG | magnetoencephalography |
| PPA | parahippocampal place area |
| PPI | psychophysiological interaction |
| ROI | region of interest |
| V1 | primary visual cortex |

# List of Figures

# Curriculum Vitae

**Lena Maria Leeners**
28.03.1991
Münster, Germany

## Academic Career

| | |
|---|---|
| 10/2018 - present | **Doctoral Researcher**<br>University of Münster, Germany |
| 01/2017 – 09/2018 | **Doctoral Researcher**<br>Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany |
| 08/2014 – 08/2016 | **MSc – Cognitive and Clinical Neuroscience**<br>Maastricht University, The Netherlands |
| 11/2015 – 06/2016 | ***Research Intern – Master Thesis***<br>*Max Planck Institute of Psychiatry, Munich, Germany* |
| 09/2013 – 08/2014 | **MSc – Health & Social Psychology**<br>Maastricht University, The Netherlands |
| 01/2014 – 08/2014 | ***Research Intern – Master Thesis***<br>*Department of Experimental Psychology, University of Oxford, UK* |
| 09/2010 – 07/2013 | **BSc – Psychology**<br>Maastricht University, The Netherlands |
| 01/2013 – 03/2013 | ***Research Intern – Bachelor Thesis***<br>*Christophorusklinik, Münster, Germany* |
| 07/2012 – 12/2012 | ***Exchange Semester, BSc – Psychology***<br>*Stellenbosch University, South Africa* |

## Teaching Experience

| | |
|---|---|
| Winter Semester 2023/2024, 2020/2021 | **Biological Psychology Seminar – BSc Psychology**<br>University of Münster |

| | |
|---|---|
| Summer Semester 2019, 2020 | **Biological Psychology Seminar – BSc Psychology**<br>University of Münster |

## Supervised Theses

| | |
|---|---|
| Summer Semester 2021 | Annette Golla, Bachelor of Science (Psychology) |

## Publications

***Leeners, L. M.***, Pomp, J., Roehe, M. A., Zahedi, A., & Schubotz, R. I. (2024). Co-occurring positive and negative prediction error signals are independent of stimulus repetition effects. *Cerebral Cortex*. In revision.

Roehe, M. A., Kluger, D. S., Schroeder, S. C., ***Schliephake, L. M.,*** Boelte, J., Jacobsen, T., & Schubotz, R. I. (2021). Early alpha/beta oscillations reflect the formation of face-related expectations in the brain. *Plos One, 16*(7), e0255116.

***Schliephake, L. M.***, Trempler, I., Roehe, M. A., Heins, N., & Schubotz, R. I. (2021). Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA. *NeuroImage, 236*, 118028

Albantakis, L., Brandi, M. L., Zillekens, I. C., Henco, L., Weindel, L., Thaler, H., ***Schliephake, L.***, Timmermans, B. & Schilbach, L. (2020). Alexithymic and autistic traits: Relevance for comorbid depression and social phobia in adults with and without autism spectrum disorder. *Autism, 24*(8), 2046-2056

Zillekens, I. C., ***Schliephake, L.***, Brandi, M. L., & Schilbach, L. (2019). A look at actions: direct gaze modulates functional connectivity of the right TPJ with an action control network. *Social Cognitive and Affective Neuroscience, 14*(9), 977-986

Kadosh, K. C., Haller, S. P., ***Schliephake, L.***, Duta, M., Scerif, G., & Lau, J. Y. (2018). Subclinically Anxious Adolescents Do Not Display Attention Biases When Processing Emotional Faces–An Eye-Tracking Study. *Frontiers in Psychology, 9*, 1584

## Conference Presentations

***Schliephake, L.***, Brandi, M.-L., Schilbach, L. (2017). The influence of direct gaze on the mechanisms of action control. *IMPRS NeuroCom Summer School*, London, UK.

***Schliephake, L.***, Brandi, M.-L., & Schilbach, L. (2016). Modulation of brain activity in the inferior frontal gyrus during the generation of manual actions in a social context. *Cognitive and Affective Neuroscience (SANS) conference*, New York City, USA.

# Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate: Lena Maria Leeners

Title of dissertation: Neurobehavioural evidence for co-occurring positive and negative prediction errors during contextual model updating

## 1. Documentation of **open-science activities**

Manuscript 1

| | Yes | No | If yes, please specify the source |
|---|---|---|---|
| | | | |
| Pre-registration | | X | |
| Publication of data | X | | Open Science Framework (https://osf.io/vxqrh/) |
| Publication of analysis scripts | | X | |
| Publication of materials | | X | |
| Open access publication | X | | https://doi.org/10.1371/journal.pone.0255116 |

Manuscript 2

| | Yes | No | If yes, please specify the source |
|---|---|---|---|
| | | | |
| Pre-registration | | X | |
| Publication of data | X | | Open Science Framework (https://osf.io/vxqrh/) |
| Publication of analysis scripts | | X | |
| Publication of materials | | X | |
| Open access publication | X | | https://doi.org/10.1016/j.neuroimage.2021.118028 |

Manuscript 3

| | Yes | No | If yes, please specify the source |
|---|---|---|---|
| | | | |
| Pre-registration | | X | |
| Publication of data | | X | |
| Publication of analysis scripts | | X | |
| Publication of materials | | X | |
| Open access publication | | X | |

## 2. Declaration on **the consideration of ethical aspects**

| **Study number** | **Source (manuscript / chapter of dissertation):** e.g., study 1 in paper 2, study 1 described in chapter 4 | **Was the study reviewed by an ethics commission?** | |
|---|---|---|---|
| | | **yes** | **No** |
| 1 | Manuscript 1 | X | |
| 2 | Manuscript 2 | X | |
| 3 | Manuscript 3 | X | |

## **3. Declaration** of one's **personal contribution** to the submitted academic manuscripts by two or more authors

**Manuscript 1**

| | |
|---|---|
| Title | Early alpha/beta oscillations reflect the formation of face-relatedexpectations in the brain |
| Author(s) | Marlen Roehe, Daniel Kluger, Svea Schroeder, Lena M. Schliephake, Jens Boelte, Thomas Jacobsen & Ricarda I. Schubotz |

| | | | |
|---|---|---|---|
| Publication status: | not yet submitted | | (please mark with X) |
| | submitted | | |
| | in review | | |
| | in revision | | |
| | accepted | | |
| | published | X | |

| | |
|---|---|
| Journal | PLoS ONE |
| Year of publication | 2021 |

Description of your own contribution in the case of joint authorship:

- partly responsible for data collection
- partly responsible for processing, analysing and interpreting data
- party responsible for drafting and revising the manuscript

**Manuscript 2**

| | | | |
|---|---|---|---|
| Title | Positive and negative prediction error signals to violated expectations of face and place stimuli distinctively activate FFA and PPA | | |
| Author(s) | Lena M. Schliephake, Ima Trempler, Marlen A. Roehe, Nina Heins & Ricarda I. Schubotz | | |
| Publication status: | not yet submitted | | (please mark with X) |
| | submitted | | |
| | in review | | |
| | in revision | | |
| | accepted | | |
| | published | X | |
| Journal | NeuroImage | | |
| Year of publication | 2021 | | |

Description of your own contribution in the case of joint authorship:

- partly responsible for the study’s conception and design
- mainly responsible for data collection
- mainly responsible for processing, analysing and interpreting data
- mainly responsible for drafting and revising the manuscript
- marked as corresponding author

**Manuscript 3**

| | | | |
|---|---|---|---|
| Title | Co-occurring positive and negative prediction error signals are independent of stimulus repetition effects | | |
| Author(s) | Lena Maria Leeners, Jennifer Pomp, Marlen A. Roehe, Anoushiravan Zahedi & Ricarda I. Schubotz | | |
| Publication status: | not yet submitted | ☐ | (please mark with X) |
| | submitted | ☐ | |
| | in review | ☐ | |
| | in revision | ☒ | |
| | accepted | ☐ | |
| | published | ☐ | |
| Journal | Cerebral Cortex | | |
| Year of publication | | | |

Description of your own contribution in the case of joint authorship:

- partly responsible for the study’s conception and design
- mainly responsible for data collection
- mainly responsible for processing, analysing and interpreting data
- mainly responsible for drafting and revising the manuscript
- marked as corresponding author

Münster, 26.05.2024

Place, Date

Signature of the doctoral candidate