UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF

Institut für Systemische Neurowissenschaften

Prof. Dr. med. Christian Büchel

Computational mechanisms of how prior expectations influence face perception

Dissertation

zur Erlangung des Doktorgrades Dr. rer. biol. hum.
an der Medizinischen Fakultät der Universität Hamburg

vorgelegt von:

Annika Garlichs
aus Celle

Hamburg 2025

Angenommen von der
Medizinischen Fakultät der Universität Hamburg am: 22.04.2025

Veröffentlicht mit Genehmigung der
Medizinischen Fakultät der Universität Hamburg.

Prüfungsausschuss, der/die Vorsitzende: Prof. Dr. Helen Blank

Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. Cristina Becchio

Inhaltsverzeichnis / Table of Contents

1. Introduction

Faces provide some of the richest information sources in social communication. For example, identifying familiar faces (e.g., family, friends) and inferring a person's demographics (e.g., age, gender, ethnicity) are essential for social interaction (for a review, see Bruce, 2002). Face identification is an automatic skill (Yan et al., 2017) that we develop as we become older and refers to the within-class differentiation of individuals sharing an overall configuration of facial features (i.e., eyes, nose, mouth; Carey et al., 1997; Mondloch et al., 2006). However, this automatism can be impeded depending on the context you encounter a person in. For example, you might be familiar with the situation that you see a colleague every day at work, but while being on vacation at the beach, you get a surprising text message: "Did you not see me just now?" In fact, you have just walked past your colleague, although you know their face and should be able to automatically recognize them. However, the unusual context and the low probability of meeting them have altered your perception, meaning that expectations derived from the context you were in influenced the processing of incoming sensory information.

This exemplifies the widely accepted phenomenon that perception is a combination of bottom-up as well as top-down information (Clark, 2013; Friston, 2005; Von Helmholtz & Nagel, 1909). In the research field of face perception, it has been shown that expectations accelerate identification of familiar faces, for example, if expectations are evoked by the prior presentation of the name of a person (Amado et al., 2018; Ambrus et al., 2019). Conversely, incongruent expectations can lead to an interference of face identification, measurable by a slower reaction time (RT; Todorova & Neville, 2020). Furthermore, the noisier or more ambiguous the sensory face information is, the stronger the influence of expectations on disambiguating the identity (Walther, Schweinberger, Kaiser, et al., 2013; Walther, Schweinberger, & Kovács, 2013).

A prominent framework connecting how expectations might influence perception is Predictive Coding (Friston, 2005, 2009). This framework postulates that the brain predicts upcoming sensory information by applying different priors (i.e., expectations) about what is

likely to be encountered during a certain situation. The ultimate goal of an organism is to reduce surprise by making predictions as accurately as possible. This is achieved by computing whether the incoming information deviates from the expected information (i.e., the prediction error, PE) and subsequently updating the internal predictive models for upcoming situations by integrating the PE. This process is likely performed based on Bayesian principles that consider the uncertainty (or inverse: the precision) of both the prediction and the sensory input for weighing their contribution to the posterior that will be used for future model predictions (Aitchison & Lengyel, 2017; Clark, 2013; Friston, 2005; Knill & Pouget, 2004). The Bayesian brain constantly learns the statistical regularities of the environment, making the inference about the probable causes of sensory observations as accurate as possible. In our example with the colleague, based on previous experiences, our brain would predict with a high certainty to meet the colleague in the office, resulting in a low (or effectively no) PE when we meet them there. However, when we unexpectedly meet them on vacation, we are surprised and might adjust our internal model by increasing the probability of meeting them in that context. This results in less surprise and higher recognition probability the next time we encounter them on vacation.

One aspect of the Predictive Coding framework is that it suggests humans to be active agents in their environment, constantly testing their prior beliefs by performing experiments to gather information whether their predictions are accurate (Friston et al., 2012). For this purpose, humans perform eye movements, i.e., saccades and fixations, to actively sample the environment for confirmation of the predictions, hence precluding surprises (PE). This has been supported by experiments showing that context information about the emotional state or ethnicity of a person influence how we sample ambiguous faces in line with expectations (Aviezer et al., 2008; Wang et al., 2015). Similar guidance of eye movements by context has been demonstrated in the scene-object congruency literature, showing that participants are faster at locating a cooking pot than a printer in a kitchen due to the typical occurrences of these objects in a scene (Võ & Henderson, 2011). However, there have been studies challenging this guidance of eye movements towards expected context information (Bonitz &

Gordon, 2008; Brockmole & Henderson, 2008). For example, Brockmole and Henderson (2008) report a faster fixation of semantically inconsistent compared to consistent objects in a scene (e.g., eggs compared to a book in a fridge). These studies on context effects on face and scene-object perception suggest that different visual information drive eye movement: on the one hand, context information influences the sampling of ambiguous faces and scenes towards expected information; on the other hand, there have been reports of unexpected information being fixated earlier than expected information, suggesting eye movements may be driven by PEs instead of confirmative sensory information.

To resolve this question whether expected information or PEs drive our eye movements, we conducted two pre-registered eye-tracking experiments for our first study. In our paradigm, participants learned to associate names with four faces. Each face had one distinct facial feature, e.g., a high forehead or a wide chin. In each trial, a name cue served as context information for which face would likely appear next, thereby allowing predictions. The upcoming face could either be the expected identity, an unexpected identity, or an ambiguous face morph between the expected and an unexpected face. In our first experiment, we limited sensory information by presenting the face for a short duration (100 ms), hence manipulating the precision of the sensory input. We allowed for predictive saccades towards the respective facial features by providing a time interval with a face outline for reference between the name cue and the face presentation. In the second experiment, the face was presented for a longer duration (4500 ms or until button press). This allowed us to investigate for the ambiguous face morphs whether participants preferentially sample expectation-confirming or -deviating face information. In these two experiment, we aimed at answering the following questions:

1) Do participants perform predictive saccades towards expected facial features (e.g., the forehead), in line with an active sampling of expected information?

2) Do participants show an earlier fixation of expected or unexpected facial features in face morphs, either in line with an active sampling of expected information or PEs?

3) Do participants show a preferred sampling of expected or unexpected facial features in face morphs in terms of number of fixations and dwell time?

4) Is there a link between eye movements and behaviour (e.g., a last-sampling bias (Zhu, 2022), making it more likely to choose the identity in a face morph that matches the facial feature which was fixated last)?

Our study revealed that participants predictively performed saccades towards locations of expected facial features. Furthermore, participants fixated expected facial features earlier than unexpected ones in face morphs, which is in line with an active sampling of the environment for expected information (Friston et al., 2012). Interestingly, the initial preferred sampling of expected features reversed over time into an increased sampling of unexpected information. This could indicate a shift towards the PE after evaluating the initial prediction, possibly for model updates. Furthermore, in trials with completely unexpected faces, the initial guidance of fixations by expectation was reduced: expected features were still fixated earlier than regions containing no information of the two faces contained in a face morph, but there was no difference in initially fixating the expected or unexpected facial feature anymore. This indicates that highly deviating bottom-up information can weaken guidance by expectation, equally leading to an early fixation of either expected or unexpected face information. We also found a direct connection between eye movements and perceived identity in a face morph, linking the observed assimilation effect (i.e., the perception of the expected face in a face morph) to the fixation of the expected facial feature at face onset (Experiment 1) and the last fixation during face presentation (Experiment 2). In conclusion, our first study's results are in line with the Predictive Coding framework by showing a preferred sampling as well as predictive saccades before face onset towards expected information. Also, we showed that highly expectation-deviating information can partially counteract the guidance by expectations. These findings highlight the supportive role of expectations for perception, while the detection of unexpected information (PE) is preserved and could be highly relevant for revising predictive models.

While our first study provided insights into how expectations directly influence which information we sample in our world, it remained unclear how expectations and sensory information are computationally combined on a neural level. In order to gain more insights into

the computational processes of the brain, we conducted a second study using functional magnetic resonance imaging (fMRI). On the neural level, the effect of expectations on face perception has been extensively studied using fMRI (de Gardelle et al., 2013; Egner et al., 2010; Grotheer & Kovács, 2015; Summerfield et al., 2008), electroencephalography (EEG; Summerfield et al., 2011), magnetoencephalography (MEG; Johnston et al., 2017), transcranial direct current stimulation (tDCS; Kongthong et al., 2013), and transcranial magnetic stimulation (TMS; Mattavelli et al., 2011). Most of these studies report a so-called expectation suppression effect, i.e., a reduced neural response to expected compared to unexpected faces. One prominent explanation for this effect stems from the Predictive Coding framework which suggests a hierarchical processing of information: while expectations about an upcoming face are forwarded from higher to lower (sub-)cortical areas which are closer to the direct processing of the visual input, these lower areas compute and send back the PE between expectation and sensory input to higher areas, which update the predictive models for upcoming scenarios (Friston, 2005, 2008; Mumford, 1992; Rao & Ballard, 1999). This framework would explain the observed expectation suppression effect as a reduced PE for expected faces, resulting in reduced neural activation in face-sensitive regions. However, previous literature on the neural mechanisms of the expectation suppression effect fell short in testing alternative models. One alternative explanation for the reduced activation for expected face is proposed by the Sharpening account (e.g., Alink & Blank, 2021; Kok et al., 2012; Lee & Mumford, 2003; for reviews, see Clark, 2013; de Lange et al., 2018; Walsh & McGovern, 2018). According to Sharpening, neurons tuned to expected face information get enhanced, suppressing irrelevant information of nearby neurons. This would lead to an overall reduced but less noisy activation for expected compared to unexpected faces. Therefore, both explanations, PE and Sharpening, would explain the expectation suppression effect based on two contrary computational mechanisms.

In our second pre-registered study, a multivariate fMRI study combined with deep convolutional neural networks (DCNN), participants learned to associate scene images with four faces. In each trial, a scene was followed either by the expected face, an unexpected face,

or a face morph containing the expected as well as an unexpected face. While the expected and unexpected faces allowed us to investigate the univariate expectation suppression effect, the face morphs enabled us to differentiate how expectations and sensory information are combined on the neural level, i.e., whether the expected identity (Sharpening) or the unexpected identity (PE) is represented more strongly. We focused on the analysis of face-sensitive regions along the ventral face-processing hierarchy, namely the occipital face area (OFA), the fusiform face area (FFA), and the anterior temporal lobe (aTL; Blank et al., 2015; Goesaert & Beeck, 2013; Guntupalli et al., 2017; Haxby et al., 2000). We performed analyses within these regions of interest (ROI) as well as whole-brain analyses to answer the following questions:

1) Can PE or Sharpening better explain how expectations and sensory information are computationally combined along the ventral face-processing hierarchy and in the whole brain?

2) Which account more likely explains the univariate expectation suppression effect in regions along the ventral face-processing hierarchy?

Our fMRI study revealed evidence for PE processing along the whole face-processing hierarchy, with additional Sharpening of expected information in early OFA and frontal regions. Simultaneously, the posterior part of the FFA (pFFA) as well as a more lateral area in inferior and medial temporal gyrus (ITG/MTG) showed an expectation suppression effect, suggesting PE processing as the underlying computational mechanism for the expectation suppression effect in these areas. Thereby, our results support the notion of a hierarchical processing proposed by the Predictive Coding framework, by showing PEs related to higher-level expectations in lower face-sensitive regions, similar to previous studies in the macaque brain (Nigam & Schwiedrzik, 2024; Schwiedrzik & Freiwald, 2017). In addition, we extended previous research by testing Sharpening as an alternative explanation for the univariate expectation suppression effect. We revealed a complex interplay between the representations of expected and unexpected face information, as shown by their co-occurrence in OFA and the whole brain results, and their importance for the representation of faces.

In conclusion, we investigated with our two studies how expectations induced by context information influence face processing on the behavioural and neural level by using eye-tracking and multivariate fMRI analyses. In our first study, we showed that expectations actively drive eye movements towards expected information, in line with an active sampling of our environment (Friston et al., 2012), which changes over time towards unexpected information. Furthermore, the initial guidance by expectations can be reduced if the sensory information deviates too strongly from the expectation. In our second study, we were able to provide evidence for the prominent view that the univariate expectation suppression effect could be linked to PE processing and revealed that face representation involves a combination of expected (Sharpening) and unexpected (PE) face information. These results support two aspects of the Predictive Coding framework: firstly, reinforcing the human as an active agent sampling expectation-confirming information in the world, and secondly showing the neural representation of PEs alongside sharpened expected face information, in line with the suggested hierarchical processing in the brain. These results offer new insights into the complex interplay of expectations and sensory information and how they both shape the perception of our world.

2. Context Effects on Behavioural Face Perception

Why is it easier to recognize a colleague's face in the office compared to in an unexpected context, for example, on vacation at the beach? In both contexts, the sensory information, i.e., the colleague's face, is identical. Still, the context in which we encounter them is different and seemingly influences how we process the incoming sensory information. A context can influence face perception by two different mechanisms: on the one hand, context can facilitate face processing to enable a faster and more correct identification, a process called priming. On the other hand, context can shift perception away from the expected input (i.e., from the expected face identity), a process called adaptation (for reviews, see Mueller et al., 2020; Snyder et al., 2015). In our experimental designs, we used priming paradigms by facilitating the processing of a target stimulus (i.e., a face) due to preceding information (i.e., a cue).

Facilitation is usually evaluated by a reduction in reaction time (RT) and higher recognition accuracy (= lower error rate) for expected compared to unexpected faces. One type of priming is repetition or identity priming in which the same face is repeatedly presented (Bruce & Valentine, 1985; Brunas et al., 1990; Ellis et al., 1990; Martin et al., 2010). A more subtle method is associative or semantic priming in which subsequent face processing is facilitated by inducing face expectations via semantically associated information, for example, by first presenting an associated face (e.g., Angela Merkel prior to Barack Obama; Bruce & Valentine, 1986; Schweinberger et al., 1995; Vitkovitch et al., 2006; Vladeanu et al., 2006; Young et al., 1994), a name of an associated person (Wiese & Schweinberger, 2008), an associated word (e.g., 'brother'; Todorova & Neville, 2020), a name (or the initials) of a person (Amado et al., 2018; Ambrus et al., 2019; Jemel et al., 2005; Schweinberger et al., 2001; Shehzad & McCarthy, 2019), or cross-modally via a corresponding voice (Blank et al., 2015; Bülthoff & Newell, 2017; Stevenage et al., 2014). EEG studies suggest that repetition priming affects early event-related potential (ERP) components, possibly reflecting the early pre-activation of a mental face representation (Schweinberger et al., 1995). Associative priming rather affects later ERP components and does not operate on the perceptual level, influencing a post-perceptual process by a facilitated access to semantic identity information (Bentin & Deouell, 2000; Schweinberger et al., 1995; for a review, see Schweinberger & Neumann, 2016). The typical facilitation effect can be reduced and reversed with prolonged cue presentation duration, leading to an inhibitory negative priming effect with longer RTs for expected compared to unexpected faces (Barbot & Kouider, 2012, Rieth & Huber, 2010; for a review, see Mueller et al., 2020). Priming paradigms are usually characterized by a short presentation duration of the cue and could lead to an assimilation effect, i.e., the perception of ambiguous face morphs as the expected identity. In contrast, in paradigms with prolonged cue presentation (usually the same face), the adaptation to the cue leads to the perception of the unexpected face in a face morph (contrastive effect) due to neural or representational habituation (Carbon & Ditye, 2011; Leopold et al., 2001; Rieth & Huber, 2010; for reviews, see Mueller et al., 2020; Snyder et al., 2015).

In our two studies, we used associative priming paradigms, a more naturalistic approach to our everyday life in which the context serves as an information source about people we are most likely to encounter (e.g., the office for our colleague). In our first study, we operationalized context information by presenting name cues to induce expectations about specific facial features of upcoming faces. In our second study, we employed scene images of indoor scenes to evoke expectations about upcoming faces. Thereby, we investigated two versions of associative priming to understand how expectations influence behaviour, eye movements, and neural responses to expected compared to unexpected faces.

3. Predictive Coding, Prediction Errors, and the Bayesian Brain

A widely accepted viewpoint is that perception is not solely a bottom-up by sensory information driven process, but a combination of bottom-up and top-down information (Aitchison & Lengyel, 2017; Clark, 2013; Friston, 2005; Von Helmholtz & Nagel, 1909). How our brain combines incoming sensory information with expectations, however, is not fully understood yet. The Predictive Coding framework suggests that instead of representing everything our sensory system encounters in the world, it is more efficient to only encode the information deviating from the expectation, i.e., the PE (Aitchison & Lengyel, 2017; Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). This approach is similarly used in data compression for image and video processing to reduce redundancy (e.g., in .jpg-formats of images): usually, the value of a pixel is highly correlated with neighbouring pixel values, while important features like boundaries between objects are marked by higher differences in pixel value (Clark, 2013; Mead, 1990; Posch et al., 2014; for reviews, see Huang & Rao, 2011; Shi & Sun, 2019). This implies that images can be efficiently encoded or compressed by only encoding the highly different pixels, or in other terms, the unexpected information or PE (Clark, 2013; Zhaoping, 2006). As events in the real world are similarly highly structured, a reduction to processing only unexpected information may constitute a beneficial mechanism for the brain to reduce processing effort.

Neurobiological evidence for PE processing comes from the human visual system: sensory information from the retina needs to pass several 'bottlenecks' to the primary visual cortex (V1), with limited cell numbers only allowing for a certain amount of information to be processed (for reviews, see Aitchison & Lengyel, 2017; Zhaoping, 2006). Furthermore, paradigms of expectation suppression, i.e., reduced neural activity for expected compared to unexpected information (e.g., de Gardelle et al., 2013; Egner et al., 2010; Johnston et al., 2017; Summerfield et al., 2008, 2011), have often been seen as evidence for PE processing. While this assumption can hold true to a certain extent, we will further discuss this for our second study in which we tested an alternative explanation for the reduced activation for expected face (Garlichs & Blank, 2024).

Sensory information as well as predictions inherently involve uncertainty. For example, a face can be partially occluded by a hat, leading to noisy and less reliable sensory input. Similarly, there are reoccurring situations and the brain develops expectations about how likely certain sensory input is under similar future circumstances (for example, your colleague is usually in home office on Mondays and Thursdays). However, nature and people do not stick to regularities perfectly, therefore, making predictions about upcoming sensory information not 100% correct (e.g., if the colleague unexpectedly comes into the office on a Monday). The brain tries to make predictions by inferring the most likely cause for a sensory sensation. One way how this inference could be achieved is by using hierarchical generative models. The brain has built these by learning statistical regularities in the environment about which causes are likely for the observed sensation (Clark, 2013; Friston, 2005; Von Helmholtz & Nagel, 1909). For example, a rustling of leaves could be either the wind or a dangerous predator depending on whether you are walking in a park or in a jungle (Aitchison & Lengyel, 2017). The underlying cause of the rustling can therefore be concluded with a high probability from the context information.

It has been proposed that the brain incorporates the inherent uncertainties in our predictions and sensory information within the Predictive Coding framework according to Bayesian principles, leading to the term of the Bayesian brain (Aitchison & Lengyel, 2017;

Friston, 2005, 2009; Knill & Pouget, 2004). While we make predictions based on generative models about possible causes of sensory events, following PEs can then be used to update these models to enhance future predictions. This might operate based on Bayesian principles by computing a posterior based on the prior probability of the expected sensory input and the actual sensory input, while considering both of their uncertainties (or inverse: their precision; Aitchison & Lengyel, 2017; Friston, 2005, 2009; Knill & Pouget, 2004). This posterior can then be used as the new prior for upcoming situations to achieve more accurate predictions. In this context, the term Hierarchical Predictive Coding refers to the principles of Predictive Coding via bidirectional connections across multiple layers of cortex (Clark, 2013). It is assumed that higher-level cortical regions send predictions via backward connections to lower cortical regions which in turn compute the PE and send it back via feedforward connections to enhance future predictions and eventually lead to the final percept (Clark, 2013; Friston, 2005, 2008; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999; Summerfield & Koechlin, 2008). Recent literature suggests that we even might be able to differentiate within a cortical region that deeper layers contain the expected information and superficial layers contain bottom-up (PE) information (Felleman & Van Essen, 1991; Friston, 2008; Thomas et al., 2024).

4. Expectation Effects and Eye Movements

The idea of Predictive Coding led to several conclusions for human behaviour. Considering that the brain's goal is to reduce surprise (i.e., the PE) in the environment, it has been suggested that actions are used to change the sensory input accordingly (Friston, 2009). Concretely, this could be achieved by using eye movements to sample highly expected sensory information, constantly testing sensory predictions to reduce surprise (Friston et al., 2012). For example, imagine you are sitting in the cafeteria waiting for your colleague to arrive. Through the buzzling of different people talking, you are now hearing a very low voice (similar to your colleague's) behind you. Your brain will form the hypothesis that the voice was caused by your colleague based on your expectation of meeting him there. Firstly, this will reduce the general surprise of suddenly hearing a very low voice close to you in that context. Secondly,

you will turn your head and direct your gaze towards the anticipated head location of your colleague (judging from the voice source and your knowledge of his height). This gaze direction should minimize the uncertainty about your perceptual hypothesis. If it is your colleague, your hypothesis was correct and your saccade confirms your hypothesis with high certainty. This everyday example shows that humans do not only constantly predict their surroundings, but also actively use eye movements to test perceptual hypotheses to reduce possible surprise (Friston et al., 2012).

4.1 Expectation Effects on Eye Movements in Face Perception

Expectations thereby do not only influence behaviour in terms of RT, accuracy, and perceived identity, they can also influence how eyes are moved to gather information in the world. In general, when perceiving faces, humans typically first look at a point slightly below the eyes on the nose (Han et al., 2021). They usually show a left-bias, i.e., not fixating the centre between the eyes but shifting the gaze slightly towards the left eye (Chakravarthula et al., 2021; Hsiao & Cottrell, 2008; Or et al., 2015; Peterson & Eckstein, 2012). The location slightly below the eye region is a highly informative region for facial identity, allowing to simultaneously perceive the eyes with high resolution and also integrate information across facial features (Han et al., 2021). Furthermore, the eyes as well as the mouth region contain important information for emotion processing, with the eye region being fixated more for fearful, sad, or neutral faces, and the mouth region for positive emotions (Scheller et al., 2012; Schurgin et al., 2014). Importantly, context can influence this automatic face viewing behaviour: usually, viewing of angry faces involves a high focus on the eyes region while viewing of disgusted faces is shown by a symmetrical scanning pattern between the mouth and the eyes region (Wong et al., 2005). This can be reversed by presenting an angry face in a disgust context (i.e., a man holding a trash bag) versus a disgust face in an angry context (i.e., a man showing a fist; Aviezer et al., 2008). This demonstrates how higher-level context information can not only change what we perceive, but also *how* we perceive it in terms of eye movements.

Other context factors that can induce expectations and influence face sampling include familiarity and ethnicity. Familiarity can influence how we sample face information, as shown by a more distributed and less eye-focused scanning for personally familiar and famous compared to unfamiliar faces (Barton et al., 2006; Van Belle et al., 2010). This scanning pattern can be altered by the type of task applied: if participants are required to recall a person's name or judge their fame instead of evaluating their familiarity, fixation of the eyes region increases for familiar or famous faces compared to unfamiliar faces (Althoff & Cohen, 1999; Heisz & Shore, 2008). Another possible influence on face viewing behaviour involves ethnicity. Research has shown that viewing faces of other ethnicities includes increased sampling of the eyes region while the focus for one's own ethnicity shifts to other diagnostic features (e.g., the nose or mouth; Fu et al., 2012; Wang et al., 2015). Importantly, when participants were presented with ethnically ambiguous faces, morphed between Caucasian and Asian faces, viewing patterns could be reversed depending on the expectation they had about the ethnicity of the presented face: if Asian participants expected the presented face to be Caucasian, they focused on the eyes region, while shifting their gaze towards the nose region for expected Asian faces (Wang et al., 2015). Thereby, it has been shown that context information, such as emotional context, familiarity, and ethnicity, introduce high-level expectations about a face and can alter typical viewing behaviour for faces, mainly shifting the visual focus towards expected information in line with an active sampling of predicted sensory information (Friston et al., 2012).

4.2 Expectation Effects on Eye Movements in Object-Scene Recognition

Eye-tracking literature from the scene-object recognition domain suggests diverging influences of expectations on eye movements: on the one hand, there is evidence that humans locate expected items more quickly than unexpected items in a familiar scene. For example, experiences with situations and semantic and episodic knowledge about where objects are typically located within a scene can lead to an earlier and increased sampling of context-

expected (e.g., a pot in a kitchen) compared to context-unexpected objects (e.g., a printer in a kitchen; Bornstein et al., 2011; Castelhano & Henderson, 2007; Hwang et al., 2011; Spotorno & Tatler, 2017; Võ & Henderson, 2011; for reviews, see Wolfe, 2020; Wolfe et al., 2011). Further examples from the sports domain show how increasing expertise can lead to anticipatory saccades towards the expected location of a ball, e.g., towards the location where it will most likely hit the ground (Hayhoe et al., 2012; Land & McLeod, 2000; Mann et al., 2019; Vater & Mann, 2023). Also, on an everyday basis, humans perform predictive saccades towards the next steps while walking up and down the stairs (Zietz & Hollands, 2009). On the other hand, some studies show a preferred sampling of unexpected information, measured as earlier, more, and/or longer fixations of unexpected objects (Bonitz & Gordon, 2008; Coco et al., 2020; Cornelissen & Võ, 2017; Underwood et al., 2008; Võ & Henderson, 2009). For example, as opposed to the above-mentioned example of the earlier fixation of the pot in the kitchen (Võ & Henderson, 2011), there have been reports of an earlier fixation of semantically inconsistent objects (e.g., of eggs compared to a book in a fridge; Brockmole & Henderson, 2008). Similarly, Bonitz & Gordon (2008) showed that scene-inconsistent objects (e.g., a wrench compared to a fork on a dinner plate) were fixated more often and longer than scene-consistent objects. This highlights how fast our visual system can be at accessing semantic knowledge and accordingly guiding our eye movement towards unexpected information (Coco et al., 2020).

5. Study 1: Expectation Influence on Eye Movements Prior to and During Face Perception

Garlichs, A., Lustig, M., Gamer, M., & Blank, H. (2024). Expectations guide predictive eye movements and information sampling during face recognition. *iScience, 27*(10), 110920. https://doi.org/ 10.1016/j.isci.2024.110920

5.1 Introduction

There has been diverging evidence from the face perception literature and the object-scene congruency literature whether expected or unexpected information drive eye movements. On the one hand, there have been studies reporting that context information guides information sampling towards expected facial features or objects (e.g., Aviezer et al., 2008; Barton et al., 2006; Bornstein et al., 2011; Hwang et al., 2011; Wang et al., 2015). Correspondingly, it has been shown that participants perform predictive saccades towards expected locations of interest, for example, towards the expected ball location during sports (Hayhoe et al., 2012; Land & McLeod, 2000; Mann et al., 2019; Vater & Mann, 2023) or the next steps while walking the stairs (Zietz & Hollands, 2009). This would be in line with the notion of the Predictive Coding framework, describing humans as active agents sampling expected information in their environment to reduce surprise (Friston et al., 2012). On the other hand, there is literature showing a preferred sampling and fast detection of unexpected information, which would be in line with PEs driving eye movements (e.g., Coco et al., 2020; Cornelissen & Võ, 2017; Underwood et al., 2008). In our first study, we investigated in two pre-registered experiments how expectations about a face with a distinct facial feature influence eye movements during the anticipation and presentation of a face, either hinting towards an active sampling of expected features or an increased sampling of unexpected (PE) information.

5.2 Methods

We conducted two eye-tracking experiment (*N* = 34, each). In both experiments, participants learned to associate four names with four faces, each having one distinct facial feature (high forehead, wide chin, large ears, large nose; *see Figure 1A*). In the main experiments, the upcoming face image was cued by presenting the corresponding name (Ari, Bob, Cid, Dan) in one of the four monitor corners (see *Figure 1B, D-E*). Afterwards, either the expected face (match condition), an unexpected face (mismatch condition), or an ambiguous face morph between the expected and an unexpected face (partial condition) was presented.

In Experiment 1, we investigated whether context information, i.e., a name cue, would lead to anticipatory eye movements towards expected facial features. We tested this by presenting a prolonged face outline ('base face' outline without any special features) during an anticipation phase (inter-stimulus interval, ISI) before the face was shortly presented for 100 ms (see *Figure 1D*). The short presentation duration would require participants to make predictive saccades towards the expected feature during the face outline to be able to answer which face they perceived. In Experiment 2, the face was presented for a longer duration (4500 ms or until button press; see *Figure 1E*). We investigated whether context information would influence the order and the number of fixations and dwell time spent on facial features in ambiguous face morphs containing expectation-confirming and -deviating information.

We conducted analyses on three levels: firstly, on the behavioural level, we tested in both experiments whether there was a facilitation effect, i.e., a faster processing of expected compared to unexpected faces, as well as an assimilative or contrastive effect, i.e., the identification of ambiguous face morphs as the expected or unexpected face. Secondly, on the eye-tracking level, we investigated in Experiment 1 whether expectations led to predictive saccades towards the expected facial feature during the presentation of the face outline. In Experiment 2, we tested whether expectations influenced the order in which the different facial features in faces were sampled. Specifically in ambiguous face morphs, we tested whether the number of fixations and dwell time spent on expected and unexpected features differed. Thirdly, we tested behaviour and eye movements combined: in Experiment 1, in trials with 'clear' faces (match and mismatch conditions), we investigated whether fixations on the expected facial feature at face onset (which should occur after predictive saccades towards the expected feature during the ISI) would result in higher accuracies classifying a face as expected or unexpected compared to when it was not fixated. Further, in trials with face morphs (partial condition), we tested whether fixating the expected facial feature at face onset would increase the chance to classify a face morph as the expected identity. In Experiment 2, we investigated a last-sampling bias (Zhu, 2022), i.e., whether participants perceived the expected identity in a face morph more often if they fixated the corresponding facial feature last.
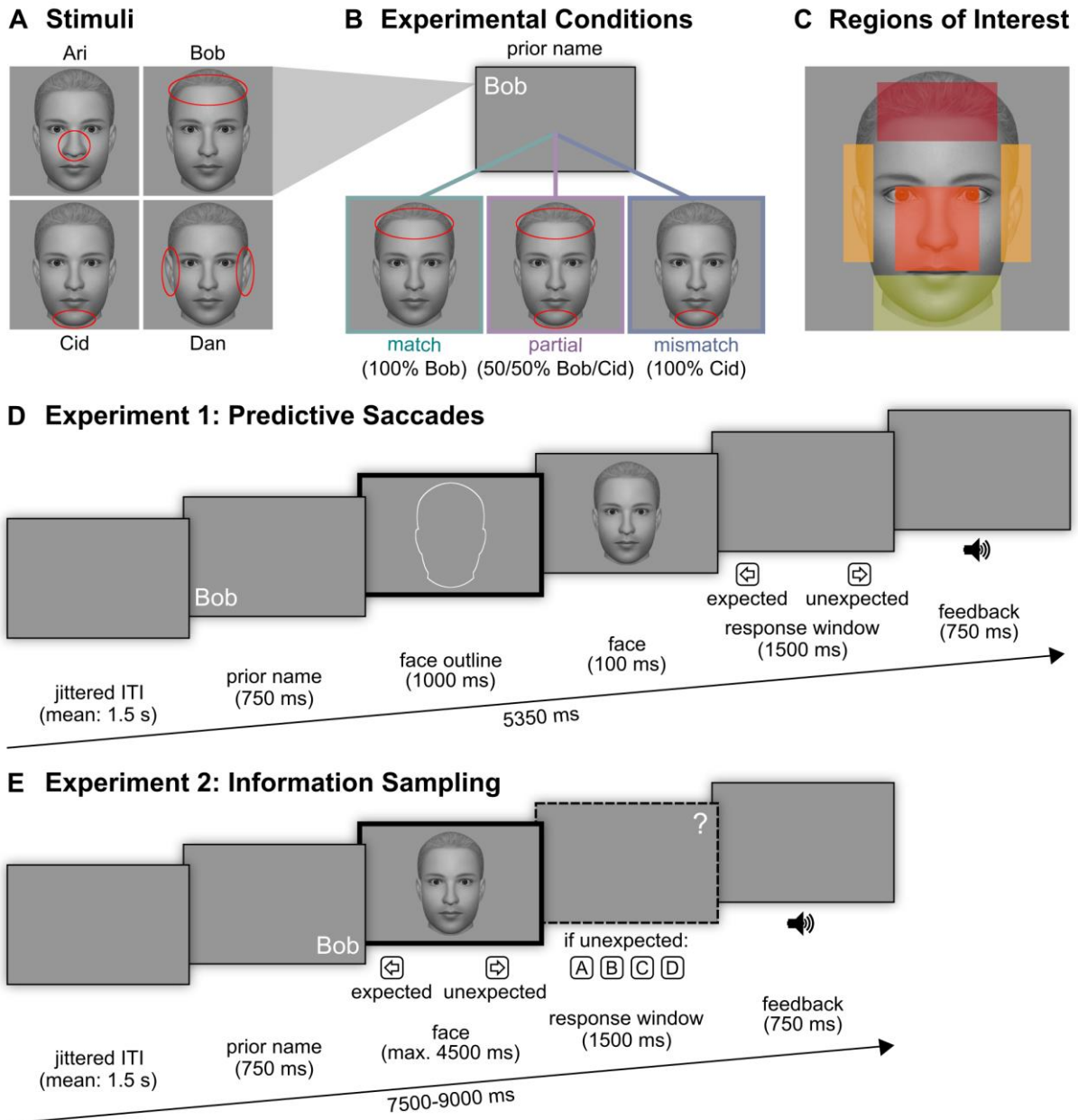
Figure 1. Stimuli, experimental conditions, regions of interest (ROI), and exemplary trials of both eye-tracking experiments. A) Stimuli: Each of the four faces had one distinct facial feature, highlighted by red circles for visualization. Faces were created using FaceMaker (Schwind et al., 2017). **B) Experimental conditions:** In each trial, the upcoming face was cued by presenting one of the four names (Ari, Bob, Cid, Bob) in one of the four monitor corners. Afterwards, the expected face (match condition), an unexpected face (mismatch condition), or a face morph between the expected and an unexpected face (partial condition) was presented. **C) ROIs:** For the eye-tracking analyses, the four distinct facial features were analysed using ROIs with the same area. **D) Exemplary trial of Experiment 1:** After a jittered inter-trial-interval (ITI), a name cue was presented in one of the four monitor corners. Afterwards, a face outline of the 'base face' without any of the four distinct facial features was presented. Next, a face was briefly shown (clear or morphed), followed by a response window in which the participant had to indicate whether the face had been 'expected' or 'unexpected' based on the preceding

name cue. Auditory feedback whether the response was correct, incorrect, or too slow was provided. For face morphs, both responses were counted as correct. The black frame indicates the time of window of interest for our eye-tracking analyses. **E) Exemplary trial of Experiment 2:** In contrast to Experiment 1, the face was presented for up to 4500 ms (or until button press). If participants responded 'unexpected', they were prompted with a question mark randomly allocated to one of the four monitor corners to indicate which of the four identities they saw in the presented face. Auditory feedback for too slow responses was provided. Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 1 of the original paper.

5.3 Results

Both experiments revealed a behavioural facilitation effect, i.e., a faster classification of expected faces compared to unexpected faces and face morphs (see *Figure 2A*). Furthermore, we found an assimilation effect, i.e., participants were more likely to perceive the expected identity in ambiguous face morphs (see *Figure 2B*).
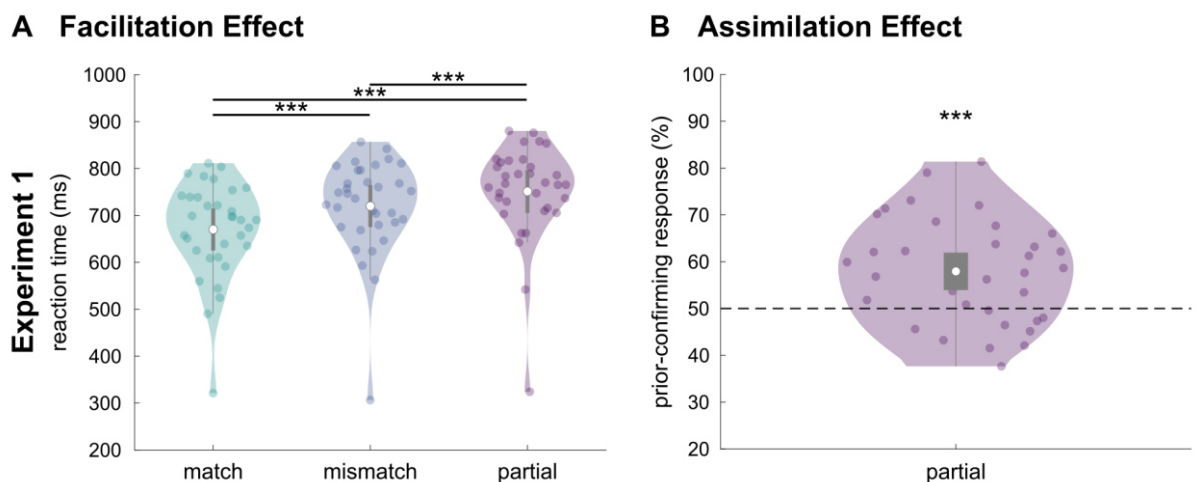
Figure 2. Facilitation and assimilation effect in Experiment 1. A) Facilitation effect: Expectations led to faster recognition of expected faces (match condition) compared to unexpected faces (mismatch condition) and face morphs (partial condition). Dots represent single participants, the white dot the mean, the grey box the 95% confidence interval, and the whiskers the $Q_{1/3}$ −/+ 1.5*interquartile range. Lines and asterisks symbolize $p < .001$. **B) Assimilation effect:** Participants perceived the expected identity in a face morph. The dashed line represents the chance level of perceiving the expected and unexpected face identity equally often (50%). Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure is based on Figure 2, excluding panels C-D containing similar results for Experiment 2, for conciseness of this dissertation.

Our eye-tracking analyses revealed that expectations led to predictive saccades towards the expected facial feature during the anticipation phase (ISI) with the 'face outline' (see *Figure 3*). Participants were more likely to perform saccades to an ROI if it had been expected compared to when it had not been expected.
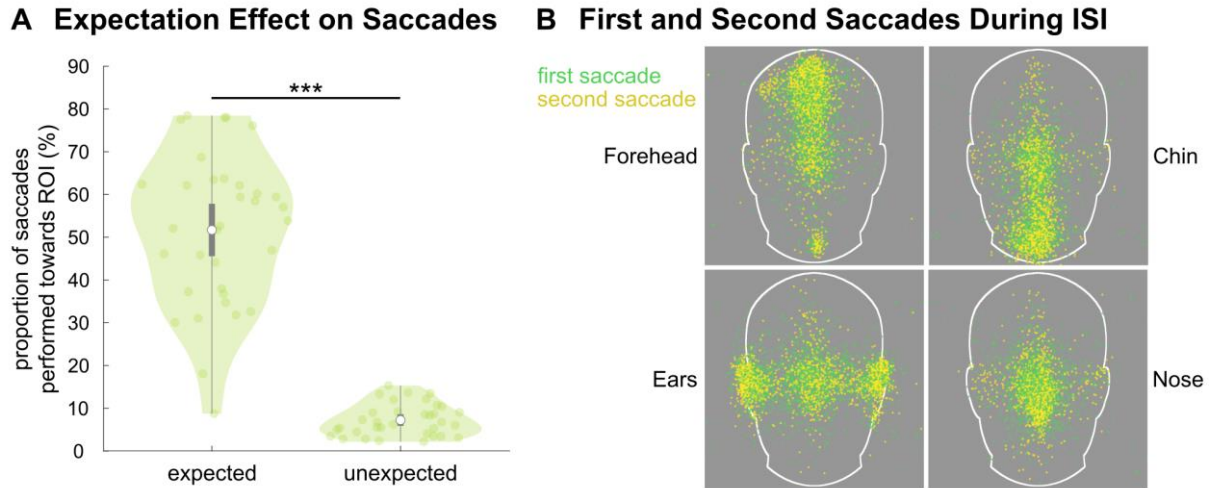

Figure 3. Predictive saccades towards expected facial features in Experiment 1. A) Predictive saccades: Expectations about an upcoming face and its distinct facial feature led to more saccades towards an ROI during the anticipatory time window (face outline) compared to when the respective face had not been expected. Dots represent single participants, the white dot the mean, the grey box the 95% confidence interval, and the whiskers the $Q_{1/3}$ −/+ 1.5*interquartile range. The asterisks symbolize $p < .001$. **B) Visualization of predictive saccades:** During the face outline, the end points of the first and second saccades are plotted across participants in green and yellow, respectively. Saccades are grouped by trials in which the distinctive features were expected (forehead, chin, ears, nose). Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 3 of the published paper.

In Experiment 2, in all trials (match, mismatch, partial), the expected feature was fixated early, as the first or second out of all four ROIs, reflecting an initial guidance of eye movements by expectations (see *Figure 4A*). In partial trials, this initial guidance by expectation was confirmed by showing earlier fixations of the expected feature compared to the unexpected and the other two features in a morph (see *Figure 4B*). Interestingly, the unexpected feature was also fixated earlier than the other two ROIs, hinting towards a guidance of fixations by top-down expectations as well as bottom-up salient deviations. In mismatch trials, initial guidance by expectations was reduced by unexpected sensory information: while the expected feature

was still fixated earlier than the other two ROIs, there was no difference in fixating the expected or unexpected ROI first, suggesting that the initial guidance of fixations by expectations can be counteracted if the bottom-up sensory information deviates strongly enough.
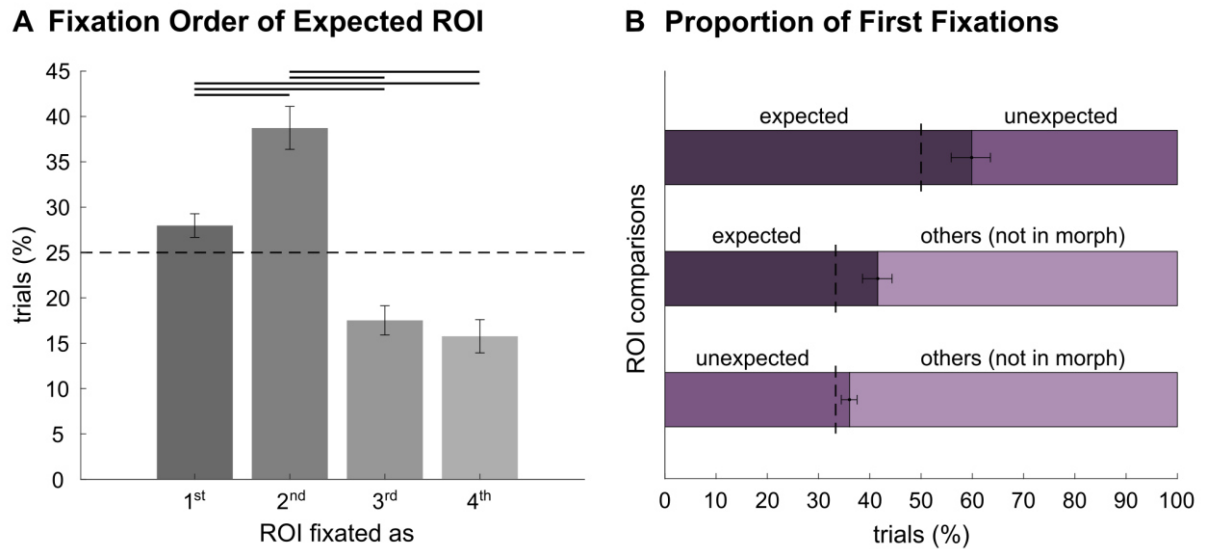
Figure 4. Fixation order towards the regions of interest (ROI) in Experiment 2. A) Expectation-induced fixation order: In all trials (match, mismatch, partial), participants were most likely to fixate the expected ROI first or second out of all four ROIs. Error bars represent 95% confidence intervals (CI), the dotted line the chance level ($\pi = 25\%$), and the solid lines $p < .001$. **B) Fixation order in face morphs:** Participants were more likely to fixate the expected ROI prior to the unexpected ROI as well as the other two ROIs. The unexpected ROI was also fixated earlier than the other two ROIs. Error bars represent 95% CIs, dotted lines represent chance levels ($\pi = 50\%$ and $\pi = 33.33\%$). Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 4 of the published paper.

Furthermore, we investigated sampling behaviour during the presentation of face morphs in Experiment 2. Our analyses of number of fixations and dwell on the expected and unexpected facial features revealed an increased sampling of expected information. Interestingly, further exploratory analyses revealed that this increased sampling was mainly driven by an initial sampling of the expected features in the first 1000 ms which reversed towards the unexpected features in later time windows (1500-2000 ms; see *Figure 5*).

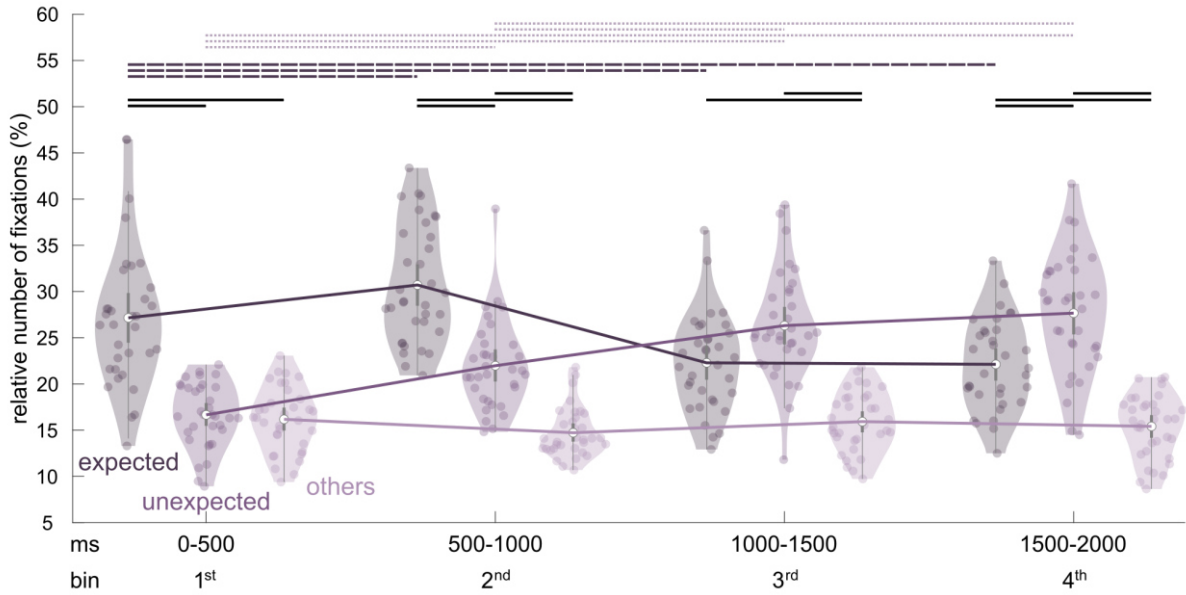**Figure 5. Relative number of fixations per time window in Experiment 2.** After an initial preferred sampling of the expected facial feature in face morphs in the first two time windows of 500 ms, the sampling switched towards the unexpected facial feature in the fourth time window (1500-2000 ms). Dots represent single participants, the white dot the mean, the grey box the 95% confidence interval, and the whiskers the $Q_{1/3}$ −/+ 1.5*interquartile range. The lines indicate $p < .05$ (solid: comparisons across all conditions within a time window; dashed: comparisons within the expected condition across time windows; dotted: comparisons within the unexpected condition across time windows). Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 (https://creativecommons.org/licenses/by/4.0/). This figure is based on Figure S1, excluding panel B containing similar results for relative dwell time per time window, for conciseness of this dissertation.

Lastly, we investigated whether there was a link between behavioural performance and eye movements. In Experiment 1, in match and mismatch trials, we found a higher accuracy in identifying a face as expected or unexpected if the expected facial feature was fixated at face onset (see *Figure 6A*). In partial trials, participants were more likely to perceive the expected identity in a face morph if they fixated on an ROI with the expected feature at face onset compared to if they did not fixate it, possibly connecting the assimilation effect to an accumulation of expected information (see *Figure 6B*). In Experiment 2, we found a last-sampling bias, i.e., fixating an ROI last made it more likely than chance to also perceive that identity in a face morph (see *Figure 6C*).
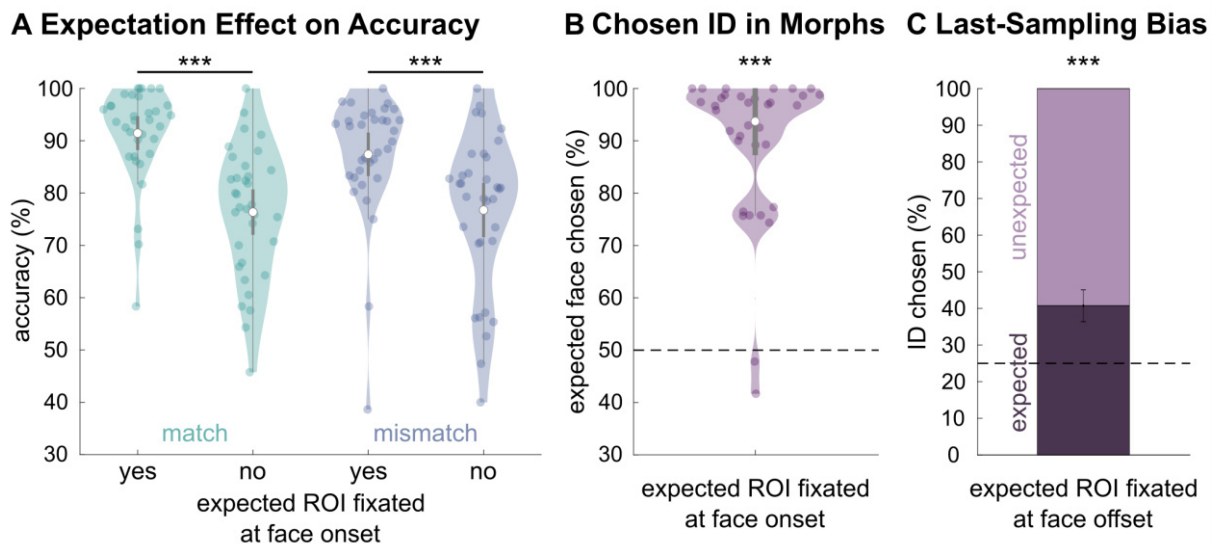
Figure 6. Results of combined behavioural and eye-tracking analyses of both experiments. A) Expectation effect on accuracy: In Experiment 1, participants were more accurate in correctly identifying a clear face (match and mismatch conditions) as expected or unexpected if they fixated the expected region of interest (ROI) at face onset compared to when they did not fixate it. Dots represent single participants, the white dot the mean, the grey box the 95% confidence interval (CI), and the whiskers the $Q_{1/3}$ −/+ 1.5*interquartile range. The asterisks symbolize $p < .001$. **B) Expectation effect on chosen identity in face morphs:** In Experiment 1, participants were more likely to perceive the expected identity in a face morph if they fixated the expected ROI at face onset compared to when they did not fixate it. Only trials in which either the expected or unexpected ROI were fixated at face onset were considered (dotted line: chance level of $\pi = .50$). The white dot represents the median. Corresponding percentages of how often the expected ROI was not fixated if the expected face was chosen can be inferred as follows: in 75% of the trials in which the expected face was chosen, the expected ROI was fixated; accordingly, in 25% of the trials in which the expected face was chosen, the ROI was not fixated. **C) Last-sampling bias:** In Experiment 2, participants were more likely than chance level to perceive the expected face in a morph if they fixated the corresponding ROI last. Trials in which one of the four ROIs was fixated last were considered ($\pi = .25$). The error bars represent the 95% CI. For **(C)**, a bar plot was chosen to illustrate the underlying subject-level proportional test as opposed to Wilcoxon tests in A and B. Our paper Garlichs et al. (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure is based on the data of the original paper, but was only created for this dissertation.

5.4 Discussion

In two experiments, we tested how expectations about an upcoming face alter anticipatory eye movements towards expected facial features as well as sampling behaviour during face presentation. Expectations were induced by name cues for four faces which each had one distinct facial feature. Sensory information was limited by a short presentation duration (Experiment 1) and ambiguous due to morphing between the expected and an unexpected identity (both experiments).

Participants showed the hypothesized facilitation effect, i.e., faster RTs for expected compared to unexpected faces, like it has been reported by other studies (e.g., Amado et al., 2018; Blank et al., 2015; Schweinberger et al., 1995; Todorova & Neville, 2020). They also showed an assimilation effect, i.e., the identification of face morphs as the expected identity. Furthermore, participants performed anticipatory eye movements towards the expected facial feature in the preceding time window in which a face outline was presented. This finding is in line with previous literature about predictive saccades towards expected locations of interest, in sports (Hayhoe et al., 2012; Land & McLeod, 2000; Mann et al., 2019; Vater & Mann, 2023) and during walking the stairs (Zietz & Hollands, 2009). This supports the idea of an active sampling of our environment, i.e., the investigation of expected locations of interest to test perceptual hypotheses (Friston et al., 2012).

In Experiment 2, we focused our analyses on the ambiguous face morph trials in which faces with the expected and an unexpected facial feature were presented. Participants showed an initial sampling of the expected facial feature, earlier than the unexpected and other two ROIs, as well as an earlier fixation of the unexpected ROI compared to the other two ROIs. These findings are in line with studies of context information guiding eye movements in face perception (Aviezer et al., 2008; Van Belle et al., 2010; Wang et al., 2015) as well as the initial sampling of scene-congruent objects (Spotorno & Tatler, 2017; Võ & Henderson, 2011). On the other hand, our mismatch trials, in which an unexpected face was shown, hinted towards an additional bottom-up factor driving eye movements. In these trials, the expected ROI was still fixated earlier than the other two ROIs, but not fixated earlier than the unexpected ROI

anymore. This shows that the guidance of expectations for eye movements can be reduced if sensory information differs strongly from the expectation.

In accordance with the initial sampling preference for expected facial features, our results showed a congruency effect, i.e., an increased fixation number and dwell time on expected compared to unexpected facial features (Bornstein et al., 2011), which is in disagreement with previous studies showing an incongruency effect (Coco et al., 2020; Cornelissen & Võ, 2017). More valuable insights about the interplay of bottom-up and top-down information were gained by our number of fixation and dwell time analyses over time, revealing information about the temporal domain. We found that the initial sampling of expected face information reversed towards an increased sampling of unexpected information the longer the face was presented. This finding is partially in line with Coco et al. (2020) who showed that the probability to look at inconsistent objects increased over time, hinting towards the additional high importance of unexpected information for sensory information processing.

Our results show that expectations drive anticipatory saccades and initial fixations which can be reduced by strong deviating sensory information or reversed towards an increased sampling of unexpected information over time in ambiguous face morphs. Our findings support the Predictive Coding idea of an active sampling of the environment, leading to an initial fixation of expected facial information. After information gathering at the expected location, bottom-up deviations drive eye movements towards the unexpected feature in a face morph, possibly reflecting a processing of the PE. Taken together, our results hint towards a more complex interplay of top-down and bottom-up information guiding face perception.

6. Context Effects on Face Processing in the Brain

While our first study gave insight into how expectations can guide anticipatory and sampling eye movements, it did not directly show how the brain represents and combines the sampled visual information with expectations. The behavioural results do not necessarily mean that the expected information is highly represented in the brain – rather, the Predictive Coding framework would suggest that for an efficient processing of the world mainly the unexpected information is relevant (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999). In our second study, we used fMRI to investigate whether and where expected or unexpected face information is represented in the brain and provide further insights into how expectations and sensory information are computationally combined.

6.1 The Ventral Face-Processing Hierarchy

Regarding brain regions that represent information about faces and facial identity, there have been descriptions of the core face-processing hierarchy encompassing regions in the inferior occipital gyrus (IOG), i.e., the OFA, in the lateral fusiform gyrus, i.e., the FFA, and in the superior temporal sulcus (STS; for a review, see Haxby et al., 2000). This core network has been functionally segregated into a ventral pathway including the FFA and a dorsal pathway including the STS, processing rather static facial information (such as identity) and changeable aspects (such as emotions), respectively (for reviews, see Bernstein & Yovel, 2015; Calder & Young, 2005; Grill-Spector et al., 2017; Haxby et al., 2000). Since we were interested in the processing of face identities, we focused on the ventral pathway (i.e., OFA and FFA) and another higher region along the ventral temporal face-processing hierarchy involved in identity processing, namely the anterior temporal lobe (aTL; Blank et al., 2015; Goesaert & Beeck, 2013; Guntupalli et al., 2017; Nestor et al., 2011).

Importantly, information about faces that is processed along the ventral face-processing hierarchy increases in complexity from lower to higher areas: OFA is the start point for both the ventral and dorsal pathways (Fox et al., 2009) and sensitive to low-level image properties such as image pixel value (Tsantani et al., 2021). It represents single features such as the

eyes, nose, and mouth (Pitcher et al., 2011), and head view but not identity, as shown by representational geometry grouping images by view but not identity (Guntupalli et al., 2017). At an intermediate processing stage, the FFA shows representations of face symmetry (Caldara et al., 2006; Caldara & Seghier, 2009), while head view and identity representations are both present, but entangled (Guntupalli et al., 2017). It further processes social higher-level information such as traits and gender (Tsantani et al., 2021). Lastly, a region in the aTL, which has been differently named the anterior inferotemporal cortex (aIT; Kriegeskorte et al., 2007), the anterior temporal face area (ATFA; Guntupalli et al., 2017), or the anterior temporal face patch (ATFP; Rajimehr et al., 2009; Tsao et al., 2008), is the highest ventral region involved in the processing of individual face information (Kriegeskorte et al., 2007), irrespective of view (Guntupalli et al., 2017; Yang et al., 2016).

Regarding lateralization, i.e., the question whether faces are rather processed and represented in the left or the right brain hemisphere, neuroimaging results have been mixed. Although face-processing is a bilateral process, there has been extensive research showing a right hemispheric dominance for faces (Ishai et al., 2005; Jonas et al., 2016; Kriegeskorte et al., 2007; Nestor et al., 2011; Rangarajan et al., 2014; Tsantani et al., 2019; Volfart et al., 2022; for a review, see Rossion & Lochy, 2021). Importantly, there have been studies challenging this view by showing bilateral face processing without a right-hemispheric preference (Lee et al., 2022; Meng et al., 2012; Thome et al., 2022). In addition, Wu et al. (2021) showed with high-definition transcranial direct current stimulation (HD-tDCS) that only stimulation of the left FFA decreased performance in a face view discrimination task for subjects with good baseline performance and increased performance for subjects with worse baseline performance. Considering that there has been evidence for face processing being a bilateral neural process with specific left-hemispheric relevance, despite its mainly right-hemispheric dominance, we investigated both hemispheres in our second study.

6.2 The Expectation Suppression Effect

There has been extensive research about the effect of expectations on the neural responses for faces. One effect related to face processing is the repetition suppression effect which describes the reduced neural activation for faces that are repeated after each other (fMRI: Henson et al., 2000; Henson et al., 2002; Hermann et al., 2017; Ishai et al., 2004; EEG: Cao et al., 2015; Eimer et al., 2010; Kovács et al., 2006; MEG: Ishai et al., 2006; single-cell recordings in macaques: Li et al., 1993; Miller et al., 1991; for reviews, see Grill-Spector et al., 2006; Henson, 2016). This effect could be interpreted to be in line with the Predictive Coding framework (Friston, 2005), arguing that the high predictability of the upcoming face leads to a reduced PE and, hence, a reduced activation for repeated faces. However, alternative explanations such as neural fatigue, i.e., neurons being less responsive due to constant stimulation, have been discussed (for reviews, see Auksztulewicz & Friston, 2016; Grill-Spector et al., 2006; Grotheer & Kovács, 2016; Summerfield & de Lange, 2014). In order to be able to differentiate whether the reduced activation for repeated faces is really due to expectations influencing sensory processing, additional experimental manipulations are required.

Paradigms involving the direct manipulation of expectations about an upcoming face report a so-called expectation suppression effect, which means the reduced neural activation for expected compared to unexpected faces. Similarly to the repetition suppression effect, it has been measured across methodologies (fMRI: de Gardelle et al., 2013; Egner et al., 2010; Grotheer & Kovács, 2015; den Ouden et al., 2010; Pajani et al., 2017; Summerfield et al., 2008; EEG: Summerfield et al., 2011; MEG: Johnston et al., 2017). The expectation suppression effect has been observed in OFA (Grotheer & Kovács, 2015) as well as FFA (Egner et al., 2010; Grotheer & Kovács, 2015; Larsson & Smith, 2012; den Ouden et al., 2010; Pajani et al., 2017; Summerfield et al., 2008). These studies showed, for example, that repeated face stimuli had a more pronounced repetition suppression than alternating face stimuli in a context in which repetitions were more likely (e.g., 80%) compared to a context in which they were rare (e.g., 20%; de Gardelle et al., 2013; Pajani et al., 2017; Summerfield et al., 2008). It has been

further shown that repetition suppression and expectation suppression effects are additive mechanisms that lead to an even more pronounced reduced activation for expected, repeated faces in OFA and FFA (Grotheer & Kovács, 2015), which had previously also been found using simple tones in the auditory domain (Todorovic & de Lange, 2012).

However, the underlying computational mechanism how expectations are combined with sensory information leading to this expectation suppression effect is still unclear. The hierarchical Predictive Coding framework suggests that expectations are sent by higher-level regions to lower-level regions along the hierarchy, which enables them to test predictions by comparing the expected to the actual sensory input by computing the PE. This PE is in turn sent back to higher regions via feedforward connections for internal model updates to improve predictions for the future (Friston, 2008; for a review, see Summerfield & de Lange, 2014). In line with this framework is that in the macaque brain, which has a similar face-processing hierarchy as humans, the lowest level (middle lateral (ML) face patch, analogous to the human OFA) contained PE information about face identity which should be limited to a higher region (Nigam & Schwiedrzik, 2024; Schwiedrzik & Freiwald, 2017). This would fit into the suggested interpretation that OFA and FFA represent PEs, i.e., the difference of face expectations and presented faces (Egner et al., 2010; den Ouden et al., 2010; Summerfield et al., 2008).

Therefore, it is intuitive to conclude that expectation suppression effects in OFA and FFA could reflect PE processing. However, previous studies did not sufficiently test alternative models that are also in line with Predictive Coding, but assume a different computational mechanism of how expectations and sensory information are combined. For example, Egner et al. (2010) showed that a Predictive Coding model involving the computation of PEs could explain the expectation suppression effect in FFA better than a contrasting feature detection model which was solely input-driven, but did not consider expectations in their alternative model. Stefanics et al. (2019) used a computational model simulating an ideal Bayesian observer to show that in FFA, both precision-weighted PEs regarding colour as well as predictions about facial emotions were present, but did also not test differing models about how expectations and sensory information could be computationally combined. This led us to

our second study in which we contrasted the prominent PE account to an alternative model that also takes expectations into account, namely a Sharpening model (Blank et al., 2018; Blank & Davis, 2016).

7. Sharpening

One of the most prominent Predictive Coding explanations for the expectation suppression effect is that expected faces elicit a reduced activation compared to unexpected faces in face-sensitive region due to a lower PE (Friston, 2005). Alternatively, the reduced univariate activation could be explained by a Sharpening of expected information (Aitchison & Lengyel, 2017; Blank et al., 2018; Blank & Davis, 2016; González-García & He, 2021; Grill-Spector et al., 2006; Kok et al., 2012; Lee & Mumford, 2003; for reviews, see Clark, 2013; de Lange et al., 2018; Walsh & McGovern, 2018). Sharpening describes the process of single neurons tuned or specialized for the expected sensory input becoming more active while neurons specialized for unexpected sensory input become suppressed. Sharpening of expected information along the ventral face-processing hierarchy has previously been shown by Blank et al. (2023) who revealed two effects of expectations on face representations: firstly, face representations during prior presentation scaled with the strength of expectation (low, middle, high) in higher level area aTL. This indicates that the precision of the expectation influenced pre-activated face representations. Secondly, presented faces similarly scaled with expectation levels in lower region OFA. This is interesting because it indicates in line with Predictive Coding that expectations can influence representations of actual sensory incoming information in lower sensory areas. In a different study investigating visual representations, prior expectations sharpened neural representations of Mooney images, i.e., ambiguous black-and-white images, after exposure to their 'clear' version, starting from V1 across the ventral stream to the fusiform gyrus (González-García & He, 2021).

Although these findings showed that Sharpening is a process involved in the representation of visual information along the ventral face-processing hierarchy, its involvement in the univariate expectation suppression effect remains unclear. Univariate fMRI analyses contrast

the neural activation of two experimental conditions (e.g., 'unexpected faces vs. expected faces') in voxels encompassing hundreds of thousands to millions of single neurons, reflecting an accumulated response of millions of neurons (Ip & Bridge, 2021; Smith et al., 2001). The expectation suppression effect could either be explained by a high activation of neurons encoding the unexpected information (PE), leading to reduced activation for expected faces. Or it could be explained by a high activation of specialized neurons for the expected input, leading to an overall reduced but less noisy activation for expected faces. To overcome this limitation of univariate analyses, we used Representational Similarity Analysis (RSA; Kriegeskorte, 2008; Nili et al., 2014) to differentiate the two computational mechanisms (PE and Sharpening) on the multivariate level (Alink & Blank, 2021; Blank et al., 2018; Blank & Davis, 2016; Ufer & Blank, 2023; see *Figure 7*).

In RSA (Kriegeskorte et al., 2008; Nili et al., 2014), different hypothesis representational dissimilarity matrices (i.e., hypothesis RDMs) are formulated. These hypothesis RDMs pose different assumptions about the correlational structure of experimental conditions. In our second study (Garlichs & Blank, 2024), we showed experimental trials in which the face of Bob was expected due to the preceding presentation of his associated scene image, but a face morph between Bob and Cid was presented. According to PE calculations, the representation of the face morph should reflect the unexpected face part in the face morph (Cid). In our PE hypothesis RDM, this led to an assumed high similarity of the face morph to the 'pure' face representation of Cid measured in neutral trials (in which all faces were equally probable, resulting in representations independent of face-specific expectations). Correspondingly, we assumed a high dissimilarity of the face morph to the representation of Bob. In contrast, our hypothesis RDM based on Sharpening computations suggested, after the scene indicative for Bob, a high similarity of the face morph between Bob and Cid to the expected face (Bob) and a high dissimilarity to the representation of the unexpected face (Cid). These hypothesis RDMs could then be correlated with the neural RDM, i.e., the correlational structure of the neural activation for the experimental conditions (Kriegeskorte et al., 2008; Nili et al., 2014). Direct comparison of the correlation values for the different hypothesis RDMs with the neural RDM

allowed to decide which hypothesis model reflects the neural representational similarity structure better. By using multivariate RSA, we thereby overcame the limitation of univariate contrasts and were able to differentiate both computational mechanisms, PE and Sharpening, on a neural level (Blank et al., 2023; Blank & Davis, 2016; Garlichs & Blank, 2024).
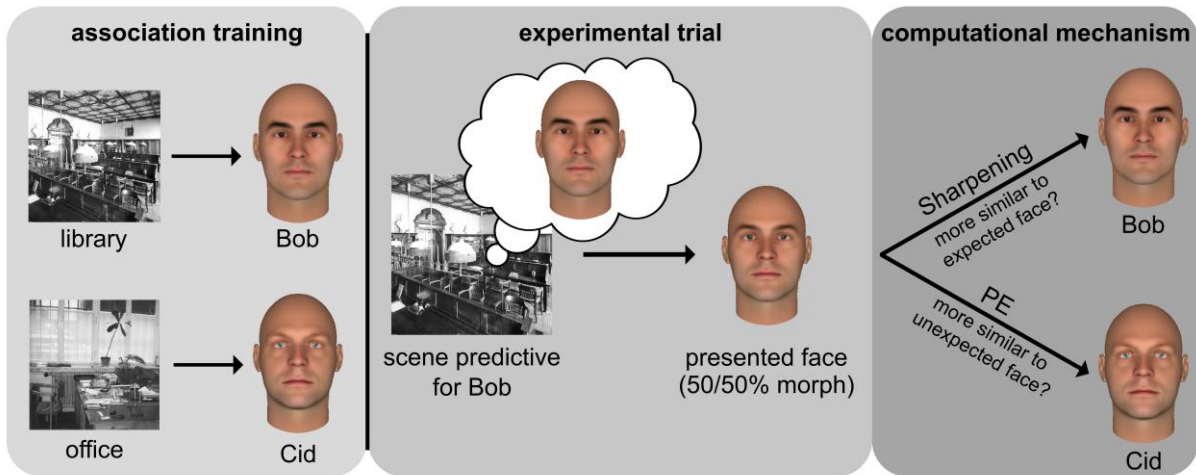
Figure 7. Visualization of how prior expectations could be combined with sensory information during face perception in Garlichs & Blank (2024). In a training outside of the scanner, participants learned to associate four indoor scenes with four faces. Each scene was predictive of one upcoming face. During the main experiment in the scanner, expectations about the upcoming face in each trial were evoked by a scene image. Afterwards, ambiguous face morphs containing the expected as well as an unexpected face were shown. Using multivariate Representational Similarity Analysis (Kriegeskorte et al., 2008; Nili et al., 2014), we investigated whether the representation for the face morph was more similar to the unexpected face, indicative for a neural PE processing, or more similar to the expected face, indicative of a Sharpening of expected face information. The face images were created with FaceGen Modeller Core 3.22. Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure was created based on Figure 1c and reduced to the face morph trials to emphasize our research question. For visualization, a training image for Cid was added. The shown scene images are in public domain and available at https://commons.wikimedia.org, but not part of the original study material due to copyright reasons. For the original stimulus set, please refer to Garlichs & Blank (2024) and the corresponding Open Science Framework repository (https://osf.io/765jx/).

8. Deep Convolutional Neural Networks (DCNN)

In our second study, we combined multivariate RSA (Kriegeskorte et al., 2008; Nili et al., 2014) with DCNNs to fine-tune our hypothesis RDMs for PE and Sharpening processing. Over the past decades, there has been growing research using DCNNs as artificial neural networks to answer questions about neural processing (for reviews, see Aloysius & Geetha, 2017; Farsal et al., 2018; Gu et al., 2018; LeCun et al., 2015; Minar & Naher, 2018; O'Toole et al., 2018). It started by imitating simple neurons with artificial perceptrons (Farsal et al., 2018; Rosenblatt, 1958) and evolved by taking the organization of the visual cortex of cats into consideration (Hubel & Wiesel, 1962, 1965). Different layers of simple and complex cells were combined in neural networks called Neocognitrons to copy the human ability of pattern recognition, e.g., of handwritten Arabic letters (Fukushima, 1980; Fukushima et al., 1983).

In neural networks, the term 'convolutional' refers to the first layer in a network which is comparable to a receptive field: it performs a feature extraction on the input by running a filter or kernel over the input image, combining multiple pixel values (e.g., by multiplication), and passing the bundled information to the next layer (Databricks, 2019; O'Toole et al., 2018). Following pooling layers reduce the computational complexity by decreasing the spatial dimension of the activation maps without information loss (Aloysius & Geetha, 2017). Fully-connected layers that have connections to all neurons finally lead to a classification layer that usually identifies the input image as one of the learned objects.

For a long time, the limitation of convolutional networks has been the lack of large training data sets and computing power (Gu et al., 2018). In 2012, the 'deep' CNN AlexNet was developed, an enhanced version of LeNet-5 (LeCun et al., 1989, 1998), winning the annual ImageNet Large Scale Visual Recognition Challenge, showing a low error rate in classifying 1.3 million images into 1000 different classes (Krizhevsky et al., 2012). The term 'deep' refers to the increasing complexity of neural networks (i.e., the number of layers and transforming input information via multiple non-linear transformations; Minar & Naher, 2018). The increase in depth helps in getting better feature representations and classification performance at the cost of higher complexity and a risk of overfitting (Gu et al., 2018).

8.1 Face-Trained DCNNs and Face Processing

While neural networks initially started to target problems like object localization and recognition (e.g., of hand-written digits of U.S. postal codes; LeCun et al., 1989, 1998), advances to use convolutional networks also for face localization in images followed (Vaillant et al., 1994). For example, Lawrence et al. (1997) trained neural networks on 400 images of 40 individuals for face identification. The high number of different images per individual, showing different expression and poses, show first attempts at image-variant face identification, closer to face encounters in our daily lives.

The ultimate goal of training DCNNs on face images is to achieve automatic face recognition in images and videos with a preferably high accuracy. One of those networks, which we used in our second study, is VGG-Face which was trained on 2.6 million images of 2.622 celebrities (Parkhi et al., 2015). Compared to other state-of-the-art face recognition DCNNs, such as DeepFace (Facebook; Taigman et al., 2014) and FaceNet (Google; Schroff et al., 2015), it achieves comparable face recognition accuracy of over 98% on the Labeled Face in the Wild dataset by having a simpler network architecture and using way less data (Parkhi et al., 2015).

DCNNs trained on face images show high similarities to human face processing, making them an attractive methodology of choice to answer *why* and *how* questions of neural face processing (Dobs et al., 2023; Kanwisher, Khosla, et al., 2023; for a review, see van Dyck & Gruber, 2023). For example, they exhibit typical psychological phenomena like the face inversion effect, i.e., a disproportional decline in recognition/neural activation for inversed compared to upright faces (Dobs et al., 2023; Yin, 1969; Yovel & Kanwisher, 2005; for a review, see van Dyck & Gruber, 2023). DCNNs can also be used for age estimation, e.g., the VGG-Face network showed a higher accuracy (59.90%) compared to GoogLeNet (45.07%) in correctly estimating images into one out of eight age ranges, even under challenging sensory circumstances with images differing in resolution quality, blurring amount, lighting, pose, and facial expression (Qawaqneh et al., 2017). Similarly, modified DCNNs trained on face images can be used for highly accurate emotion recognition with 92.10% accuracy on average

(Jaquetti et al., 2022). Interestingly, if ambiguous sensory information concerns the facial identity more strongly, e.g.., by evasion disguise (change of one's appearance to look unlike one's self) or impersonation disguise (change of one's appearance to look like someone else), DCNNs show a drop in performance similar to humans (Noyes et al., 2021; Noyes & Jenkins, 2019). This further highlights the parallel processing mechanisms between face-trained DCNNs and human neural face processing.

While these similarities and relatively high accuracies already show the highly promising nature of drawing parallels between DCNNs and neural face processing, using DCNNs also has several advantages over classical (non-)invasive neurocognitive methodologies. Firstly, some knowledge and insights about the face-processing hierarchy and specificity across brain regions have only been possible due to lesions in face-sensitive regions like the fusiform gyrus, leading to acquired prosopagnosia, i.e., the inability to recognize faces (Barton et al., 2002; Roberts et al., 2015; for a review, see Rossion, 2009), or developmental prosopagnosia with a deficit in face recognition from birth (Li et al., 2020; Wegrzyn et al., 2019; Zhao et al., 2018). Alternatively, temporal defunctioning of brain regions (e.g., using tDCS) needed to be performed (Kho et al., 2023; Wu et al., 2021). In contrast, DCNNs have the major advantage of providing an in silico way to investigate the face processing hierarchy without long-lasting deprivations, extensive exposure, or creation of lesions (for a review, see van Dyck & Gruber, 2023).

Secondly, while traditional neuroscientific studies are highly controlled by contrasting, for example, faces with house images, allowing high interpretability and clear taxonomies of neural selectivity, they are also limited by the human ability to provide interpretable labels for the measured neural activity (for a review, see Doerig et al., 2023). What our brain does and how it represents information is not necessarily restricted by how humans can interpret the neural signal, but it is likely more complex, which in turn needs new models to find more complex non-interpretable features. Although people call DCNNs a 'black box', they are way more accessible and controllable than the brain and more of a modelled 'transparent box' that is easier to study (Doerig et al., 2023; van Dyck & Gruber, 2023).

Thirdly, similarities between the neural face-processing hierarchy and DCNN architectures have been drawn, highlighting their potential power to investigate unanswered questions about face processing (for a review, see van Dyck & Gruber, 2023). Early layers of VGG-Face, for example, the second convolutional layer, capture low-level image features and have been used to exclude the influence of low-level image properties on face representations measured with RSA (Dobs et al., 2019). Middle layers of VGG-Face, like pool4 and pool5, are rather stable for low-level changes in luminance, background, and grey-scaling, but prone to identity-violating manipulations (Grossman et al., 2019). Their image transformations from layers pool4 and pool5 correlated with the neural representational space extracted from OFA and FFA, respectively, making VGG-Face a suitable DCNN for usage in our second study. Furthermore, the last fully-connected layer is robust to identity-preserving face manipulations, e.g., across viewpoints (Grossman et al., 2019). It also differentially correlates higher with age information of familiar compared to unfamiliar faces, but higher with identify information for unfamiliar compared to familiar faces (Dobs et al., 2019).

8.2 Object-Trained DCNNs and Face Processing

While it is intuitive to assume that DCNNs specifically and solely trained on faces should be 'experts' in face recognition and representation, humans are not only confronted with faces throughout their lives, but also with objects. The question arises whether DCNNs trained on face images are 'special' in comparison to networks trained, additionally or solely, on objects. This directly touches upon the neural question of how 'special' face processing in corresponding neural areas is: early neurocognitive research has suggested that the FFA is specifically activated for faces (Sergent et al., 1992; Haxby et al., 1994; Puce et al., 1995; Kanwisher et al., 1997; for a review, see Iidaka, 2014). However, the expertise hypothesis suggests that the FFA is rather process-specific for differentiating exemplars of highly overlearned categories, for example, birds (Gauthier et al., 2000), cars (Gauthier et al., 2000; McCugin et al., 2012) or Greebles – novel objects resembling aliens, with subjects learning to

individualize them on the 'familiy' and 'individual' level (Gauthier et al., 1999; for a meta-analysis, see Burns et al., 2019).

Research on object-trained networks has provided mixed results concerning the specificity of face processing. On the one hand, as expected, networks trained for object recognition showed poor performance for face recognition, while networks trained on face recognition showed poor performance for object recognition (Dobs et al., 2022; Kanwisher, Gupta, et al., 2023). If a network was trained on both faces and objects, however, the network was able to perform both tasks with high accuracy due to a spontaneous segregation into processing systems for faces and objects in mid-level network stages, similar to a probable neural segregation in the brain (Dobs et al., 2022). On the other hand, there have been studies showing that networks trained on object recognition can differentiate objects from faces. While this could be limited to an ability to classify images into objects and non-objects, they are also prone to typical psychological phenomenons such as the face inversion effect and face pareidolia (Zhao et al., 2024), i.e., recognizing facial features in objects even though there is no real face (for example, in a power socket; e.g., Liu et al., 2014; Wardle et al., 2020). This shows that even though the network has been trained on object images, its representations became tuned to upright face configurations, showing a natural evolvement of face sensitivity (Zhao et al., 2024). In contrast, other studies report the absence of a face inversion effect in object-trained networks, arguing that face-trained networks learn something specific about faces beyond a general differentiability of complex visual stimuli (Dobs et al., 2023). This highlights the still unresolved ambiguity in this relatively new research field, similar to the discussion about the specificity of face-sensitive areas. Face recognition may require more domain-general processing mechanisms, while face identification is a highly domain-specific mechanism (for a review, see van Dyck & Gruber, 2023).

Most important for our research have been findings concerning the representational similarity of object-trained networks with neural representations along the ventral face-processing hierarchy. Although there have been significant correlations between image transformations derived from low- and mid-level layers pool4 and pool5 of the face-trained

network VGG-Face and OFA and FFA, respectively, similar correlations have been found using VGG-16 layer activations (Grossman et al., 2019). Additionally, Ratan Murty et al. (2021) showed in a large-scale comparison of face- and object-trained networks that multiple of the latter had a higher predictability of neural responses to faces compared to VGG-Face, with ResNet50 showing the highest predictability. Therefore, considering the mixed results regarding the face sensitivity of object-trained DCNNs and some clear indices for their ability to achieve equal performance to or surpass face-trained DCNNs (Grossman et al., 2019; Ratan Murty et al., 2021), we additionally investigated VGG-16 (Simonyan & Zisserman, 2015), a network with the same architecture as VGG-Face but trained on the ImageNet dataset (Deng et al., 2009), as well as ResNet50 (He et al., 2016) in our second study.

9. Study 2: Expectation Influence on Face Representations in the Brain

Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. *Nature Communications*, *15*(1), 3407.

9.1 Introduction

In our second study, we aimed at answering the research question how expectations and sensory information are computationally combined on a neural level. Our first eye-tracking study (Garlichs et al., 2024) revealed an initial sampling of expected facial features in line with an active sampling of predicted information (Friston et al., 2012). However, this behaviour measured by eye movements does not reveal how top-down expectations and bottom-up information are combined and represented on a neural level.

We specifically investigated the ventral face-processing hierarchy including areas OFA, FFA, and aTL, as these provide clearly defined face-sensitive regions increasing in their complexity of represented face information (Blank et al., 2015; Goesaert & Beeck, 2013; Guntupalli et al., 2017; Haxby et al., 2000; Nestor et al., 2011). Previous research had shown an expectation suppression effect in these regions, i.e., a reduced activation for expected compared to unexpected faces (e.g., Egner et al., 2010; den Ouden et al., 2010; Summerfield

et al., 2008). This finding could be interpreted as evidence for the computation of PEs in line with the Predictive Coding framework (Friston, 2005), i.e., a reduced PE for expected faces. In the macaque face-processing hierarchy, there has been evidence for the computation of PEs in a lower area similar to OFA reflecting higher-level identity information (Nigam & Schwiedrzik, 2024; Schwiedrzik & Freiwald, 2017). This information should be limited to higher face-sensitive regions, arguing in favour of the hierarchical information processing structure proposed by the Predictive Coding framework. However, an alternative explanation for the expectation suppression effect could be Sharpening, the tuning of specialized neurons towards the expected face information. This would result in a reduced overall, but less noisy and therefore sharper signal for expected faces (Aitchison & Lengyel, 2017; Blank et al., 2018; Blank & Davis, 2016; González-García & He, 2021; Grill-Spector et al., 2006; Kok et al., 2012; de Lange et al., 2018; Lee & Mumford, 2003). As these two mechanisms, PE and Sharpening, are not differentiable on the univariate analysis level as they both result in a main effect of supressed signal for expected inputs, we conducted multivariate analyses (Alink & Blank, 2021; Blank et al., 2018; Blank & Davis, 2016; Ufer & Blank, 2023). We combined RSA (Kriegeskorte et al., 2008; Nili et al., 2014) with DCNNs to investigate how expectations and sensory information are computationally combined in the ventral face-processing hierarchy and in the whole brain.

9.2 Methods

Participants (*N* = 43) learned to associate images of indoor scenes with four male face images (see *Figure 8a*). In each trial, participants first saw a scene image followed by a face image (see *Figure 8c*). Their task was to indicate which of the four learned identities they (mostly) saw in a face. Crucially, in addition to showing the expected face (match condition) or an unexpected face (mismatch condition), we used face morphs between two face images (partial condition), always containing the expected as well as an unexpected identity. In a prior face morph calibration, we determined each participant's perceptual 50/50% threshold for each of the six morph combinations, i.e., the morph level at which they were equally likely to perceive

either identity in a face morph. Importantly, our approach of using face morphs allowed us to have the same sensory information but with different expectations (e.g., the face morph between Bob and Cid, either shown after the scene predictive for Bob or for Cid). Thus, any differences in neural activation for the face morphs could be related to top-down expectations induced by the scene cues.
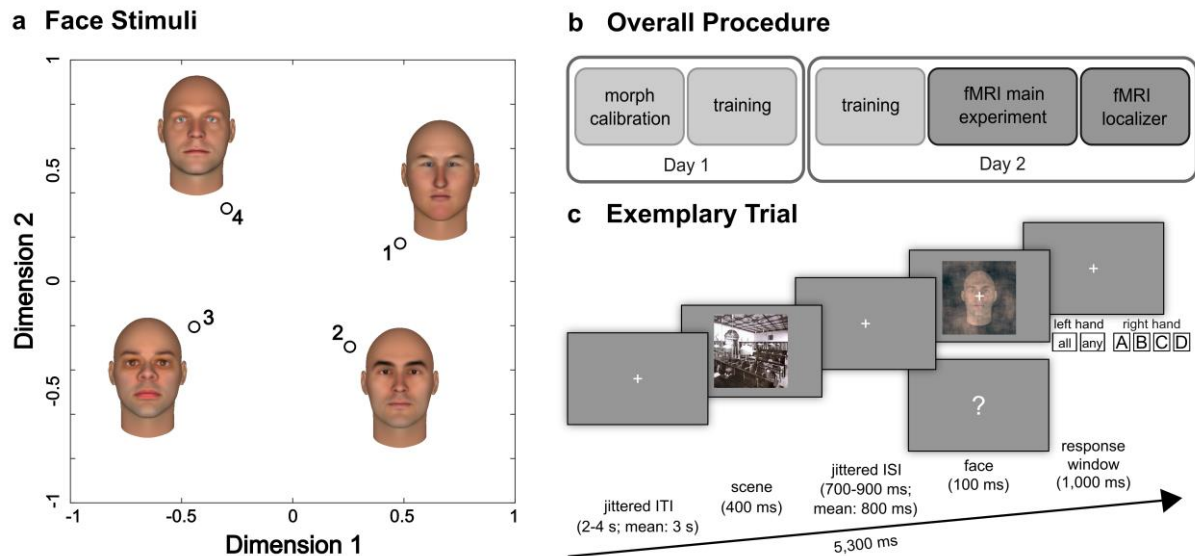

Figure 8. Face stimuli, procedure, and exemplary trial of the experiment. a) Multidimensional scaling of the four face images: The face stimuli were created using FaceGen Modeller Core 3.22. To ensure that the four faces were equally distinct from each other, a classical MDS based on layer pool4 activations of the neural network VGG-Face (Parkhi et al., 2015) was calculated. Their distance was measured using '1 – Pearson correlation' and the dissimilarities rescaled to numbers between 0 and 1. Their dissimilarity structure is presented based on two dimensions and shows that the faces were equally dissimilar to each other. **b) Overall procedure:** The experiment took place on two consecutive days. On the first day, participants completed a morph calibration experiment to determine their individual perceptual threshold for each face morph combination at which they classify each morph in 50% of the cases as the expected and in 50% of the cases as the unexpected identity. They also completed a training to learn the scene-face associations and the task. On the second day, a shortened training was performed, followed by the main experiment in the fMRI scanner. Afterwards, a functional localizer to define our regions of interest (occipital face area, OFA; fusiform face area, FFA; anterior temporal lobe, aTL) was run. **c) Exemplary trial:** In each trial, a scene was presented. If a face was presented afterwards, the task was to indicate with the right hand which face they (mostly) saw. In case of a question mark, the task was to indicate which face they anticipated based on the preceding scene. If the fifth, neutral scene was presented, all faces had an equal probability of appearing. Responses for faces or question marks after the neutral scene had to be performed with the left hand. Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure was created based on Supplementary Figure

6 and Figure 1 of the original paper. The shown scene image in c) is in public domain and available at https://commons.wikimedia.org, but not part of the original study material due to copyright reasons. For the original stimulus set, please refer to the original paper and the corresponding Open Science Framework repository (https://osf.io/765jx/).

We used multivariate RSA (Kriegeskorte et al., 2008; Nili et al., 2014) which involved defining hypothesis RDMs for PE processing and Sharpening and correlating them with the neural RDM based on T-images of different experimental conditions (Blank et al., 2023; Blank & Davis, 2016). In our study, we investigated whether the neural representation for a face morph was more similar to the expected identity, which would be in line with a Sharpening account, or more similar to the unexpected identity, indicative for PE processing (see *Figure 9*). For this, we compared the neural activation patterns for face morphs with the 'pure' face representations for the identities measured in 'neutral' trials in which the neural representations were based on the sensory input independent of expectations. For example, our PE hypothesis RDM predicted that the representation of the face morph between Bob and Cid, presented after the scene predictive for Bob, should be more similar to the Cid representation (see *Figure 9b*). In contrast, our Sharpening hypothesis RDM predicted that the representation would be more similar to the Cid representation due to a Sharpening of expected face information.

Moreover, we combined our hypothesis RDMs with activations derived from the DCNN VGG-Face (Parkhi et al., 2015; see *Figure 9*). Grossman et al. (2019) had shown, using single-cell recordings in patients, that the neural representational dissimilarity space for face images significantly correlates with the representational space of layer activations for face images. Specifically, the representational space derived from lower-level layer pool4 activations corresponded to the representational activation space measured in OFA, while higher-level pool5 activations corresponded to region FFA. As pool5 is the highest pooling layer in VGG-Face, we also used pool5 activations for the computation of the hypothesis RDMs for highest area aTL. As it has been shown that representational spaces derived from object-trained DCNNs can outperform or are as equally good at capturing the neural representational space for face images (Grossman et al., 2019; Ratan Murty et al., 2021), we also performed our

multivariate RSA using hypothesis RDMs based on VGG-16 (Simonyan & Zisserman, 2015) and ResNet50 (He et al., 2016) activations. Furthermore, because humans can differ in the amount they use prior information, we weighted our hypothesis RDMs with individual prior weights derived from behavioural measurements how often they perceived the expected or unexpected identity in a face morph (Blank et al., 2018; Lee & Geng, 2017; Levine & Schwarzbach, 2021).
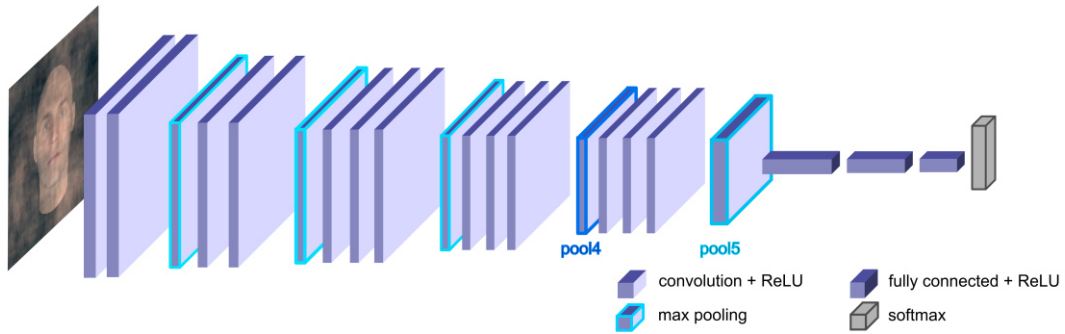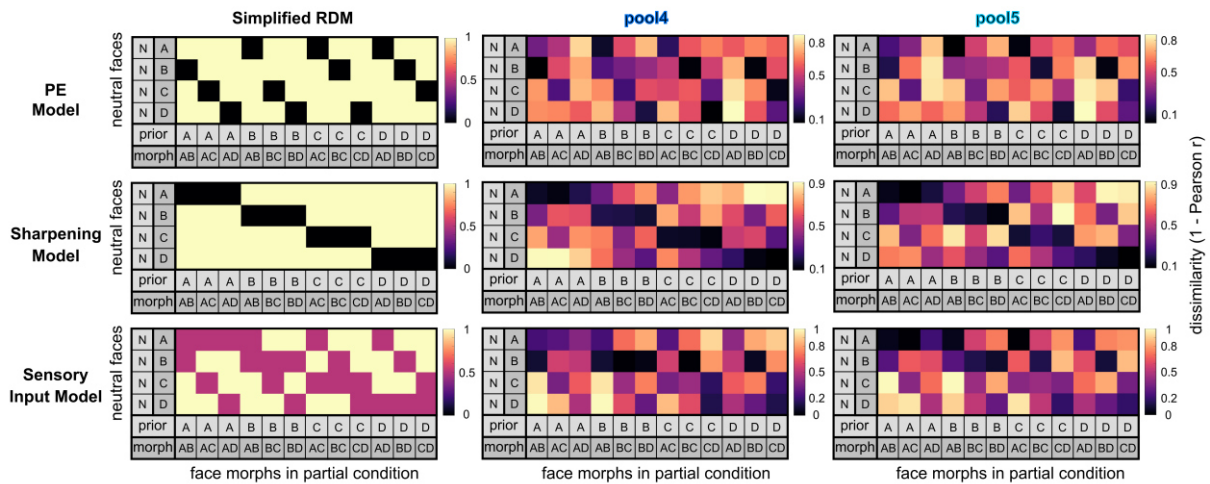
Figure 9. Architecture of the deep convolutional neural network VGG-Face and our hypothesis RDMs used for the Representational Dissimilarity Analysis (RSA). a) Network architecture of VGG-Face (Parkhi et al., 2015): For our hypothesis RDMs, we exported vector activations from layers pool4 and pool5. The shown face image was created with FaceGen Modeller Core 3.22. **b) Hypothesis RDMs:** In our RSA (Kriegeskorte et al., 2008; Nili et al., 2014), we tested three hypothesis models: a Prediction Error (PE) model, a Sharpening model, and a pure Sensory Input model in which expectations did not influence neural representations. Hypothesis RDMs based on pool4 and pool5 activations were used for OFA and FFA/aTL, respectively. The PE and Sharpening models were additionally weighted by individual behaviour to consider differences in prior usage. Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 2 of the published paper.

9.3 Results

On the behavioural level, we observed an assimilation effect, i.e., participants were more likely to perceive the expected than the unexpected identity in a face morph (see *Figure 10a*). Furthermore, expectations led to a facilitation effect, i.e., participants were faster in identifying an expected face compared to an unexpected face or a face morph (see *Figure 10b*). Additionally, we split up the face morph trials by response, i.e., whether participants perceived the expected or the unexpected face in a face morph. RT analyses revealed that participants were faster in giving prior-confirming responses compared to contrastive responses (see *Figure 10c*), although the prior-confirming responses were still slower compared to match trials in which the face completely matched their expectation.
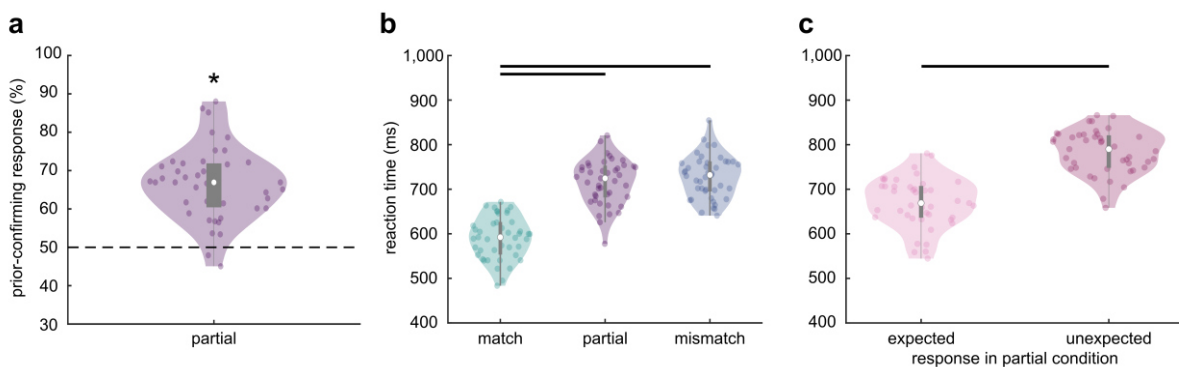
Figure 10. Behavioural effects of expectations during face perception. a) Assimilation effect: Participants were more likely to perceive the expected identity in a face morph. Single participants are represented by dots, the white dot symbolizes the median, the grey rectangle represents the interquartile range (IQR; Q_1, Q_3), and the whiskers represent $Q_1 - 1.5*IQR$ and $Q_3 + 1.5*IQR$. The asterisk indicates $p < .001$. **b) Facilitation effect:** Participants identified expected faces (match condition) faster than completely unexpected faces (mismatch condition) and face morphs (partial condition). **c) Face morphs split up by response:** Participants classified face morphs faster as the expected identity compared to when they perceived them as the unexpected identity. Black lines in (**b**) and (**c**) indicate $p < .001$. Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 3 of the published paper.

We used our functional localizer contrast 'faces > scenes' to define our ROIs (OFA, FFA, aTL; see *Figure 11a*). Previous research has suggested a segregation of FFA into an anterior and posterior part (Grill-Spector et al., 2017; Pinsk et al., 2009; Weiner & Grill-Spector, 2010). Because our activation cluster only showed an overlay with the latter (Zhen et al., 2015), we report results for the posterior FFA (pFFA).

On the univariate fMRI analyses level, we observed reduced activation for expected comparted to unexpected faces (contrast 'mismatch > match') in the ventral face-processing hierarchy in the left pFFA as well as in a cluster more lateral than our pFFA cluster, located in the ITG and MTG; see *Figure 11b*). We extracted that ITG/MTG cluster for further multivariate analyses to investigate whether the observed expectation suppression effect might result from PE or Sharpening computations. Furthermore, whole-brain analyses revealed expectation suppression effects in the bilateral anterior insula, superior parietal lobule, left thalamus, and right caudate, areas classically involved in surprise (Blank et al., 2023; Fouragnan et al., 2018) and error processing (Ham et al., 2013).

On the multivariate fMRI level, we aimed at differentiating whether the observed univariate reduced activation for expected compared to unexpected faces might be due to a reduced PE processing or a Sharpening of expected information. We found evidence for PE in areas involved in the univariate expectation suppression effect, namely pFFA and ITG, as well as in the lower and higher areas of the hierarchy, OFA and aTL (see *Figure 11c-f*). Interestingly, evidence for PE was found for both face- and object-trained DCNNs. However, VGG-Face only revealed PE processing in OFA, pFFA, and MTG, while VGG-16 showed PE along the whole face-processing hierarchy and outperformed VGG-Face in middle to higher areas pFFA and aTL. Furthermore, there were indications for additional Sharpening in OFA with VGG-Face, signifying a co-occurrence of both computational mechanisms, i.e., PE processing and Sharpening, in this lower area.
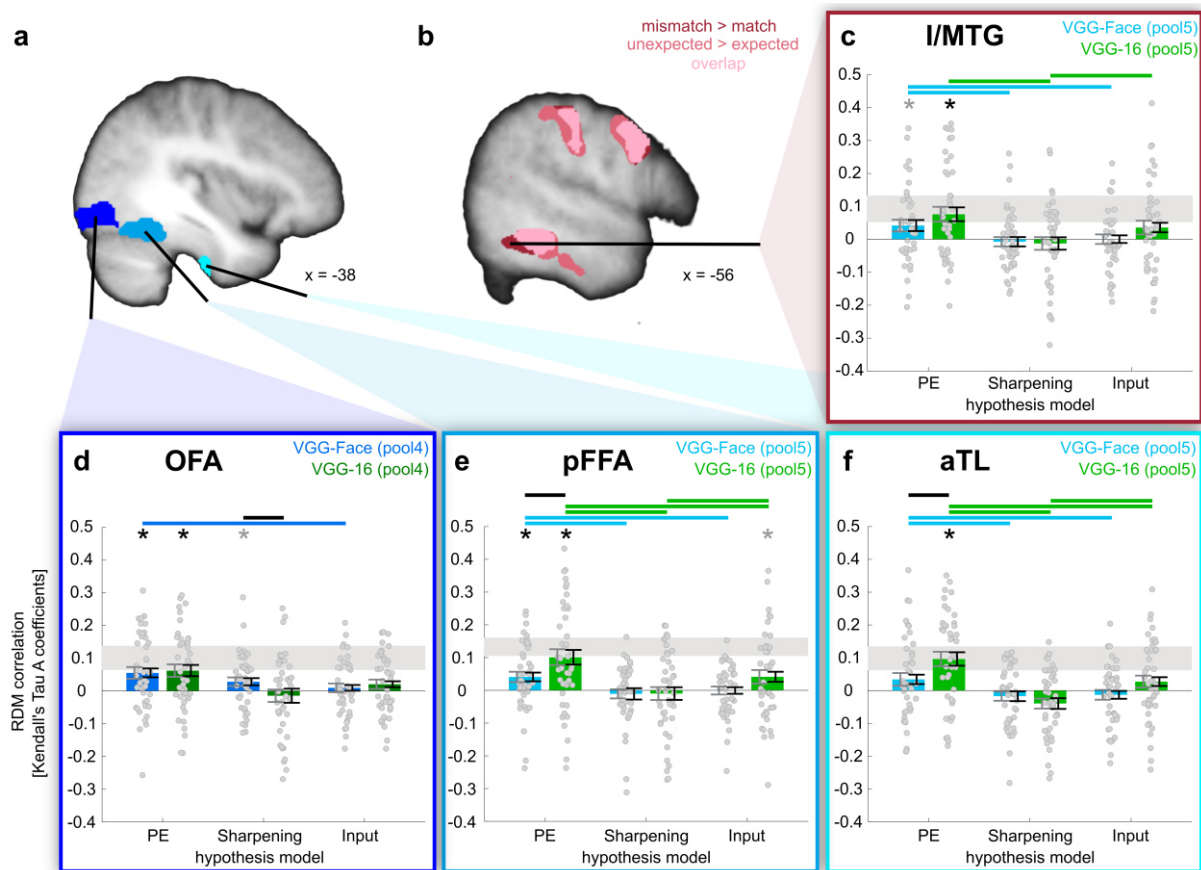
Figure 11. Regions of interest (ROI), univariate expectation suppression effect, and multivariate Representational Dissimilarity Analysis (RSA) results of our ROI analyses. a) Face-selective ROIs: We defined our ROIs (occipital face area, OFA; posterior fusiform face area, pFFA; anterior temporal lobe, aTL) based on the functional localizer contrast 'faces > scenes'. **b) Expectation suppression effect:** Univariate analyses revealed a network of temporal, parietal, and frontal regions which showed higher activations for 'mismatch' versus 'match' faces (p(unc.) < .001). Similar activations were found when contrasting trials in which participants perceived the 'unexpected' compared to the 'expected' identity in a face morph. In addition to our OFA, pFFA, and aTL ROIs, we extracted a cluster in the inferior/middle temporal gyrus (I/MTG), more lateral to our pFFA ROI, for further multivariate analyses. **c-f) Bilateral multivariate RSA results:** We correlated our hypothesis models (Prediction Error, PE; Sharpening; Sensory Input) with our neural representational dissimilarity matrices (RDM; Kriegeskorte, 2008; Nili et al., 2014). Single dots represent participants, blue and green bars the mean correlations of the hypothesis models based on VGG-Face and VGG-16, respectively. Grey error bars indicate the between-subject standard error of the mean (SEM), black error bars the within-subject SEM (Morey, 2008). The grey shaded areas represent the upper and lower noise ceiling, i.e., how well any model could perform given the noise in the data (Nili et al., 2014). Significances: asterisks (black: p(Bonf.) < .05; grey: p(unc.) < .05); lines: p(FDR) < .05 (blue: model comparisons within VGG-Face; green: model comparisons within VGG-16; black: model comparisons VGG-Face vs. VGG-16). Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure is based on Figure 4 of the original paper, excluding panels g-j that contain the unilateral RSA results, for conciseness of this dissertation.

Whole-brain searchlight analyses revealed PE processing in line with the ROI analyses, with additional PE in temporo-parietal as well as frontal regions (see *Figure 12*; similar to Muukkonen et al., 2020; Summerfield et al., 2006; Visconti di Oleggio Castello et al., 2017). Only Sharpening searchlight analyses based on the face-trained DCNN VGG-Face revealed sharpened face representations in a frontal cluster as well as in the anterior hippocampus (see *Figure 12c*). When contrasting PE and Sharpening searchlight analyses based on individual difference correlation maps, there were higher correlations of the PE compared to the Sharpening hypothesis RDM (pool4) with the neural RDM in the right angular gyrus (AnG), middle frontal gyrus (MFG), and putamen (see *Figure 12b*).
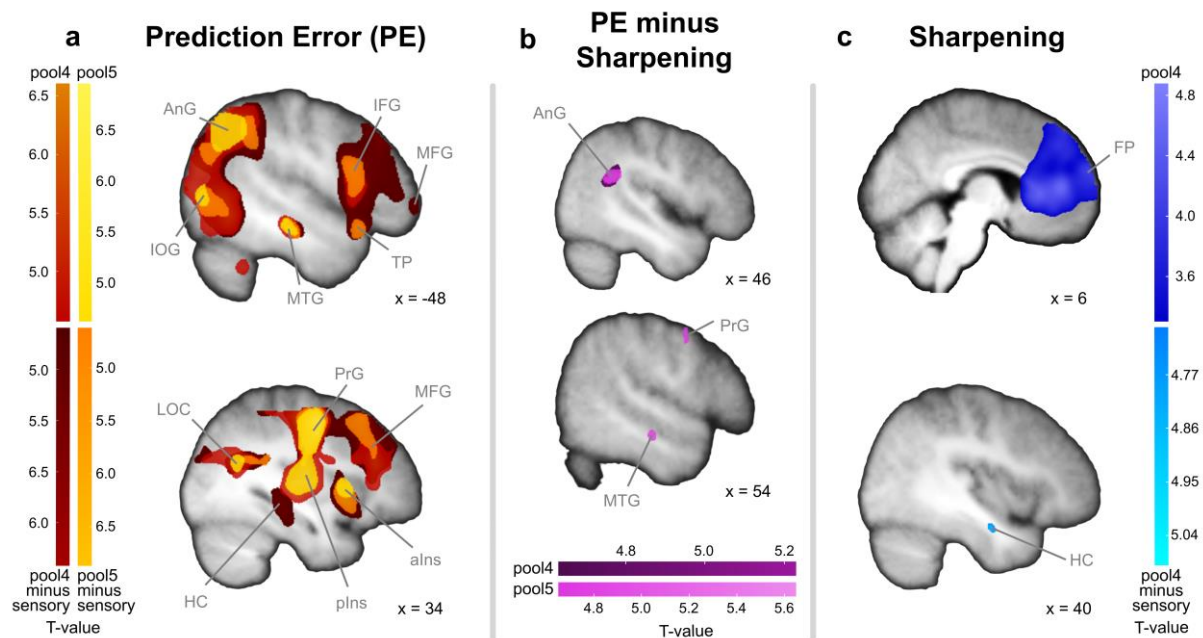
Figure 12. Multivariate searchlight analyses results for our hypothesis models based on VGG-Face (Parkhi et al., 2015). a) Prediction Error (PE) searchlight analyses: We found correlations in the angular gyrus (AnG), inferior occipital gyrus (IOG), inferior frontal gyrus (IFG), middle frontal gyrus (MFG), temporal pole (TP), middle temporal gyrus (MTG), inferior occipital gyrus (IOG), lateral occipital cortex (LOC), precentral gyrus (PrG), anterior insula (aIns), posterior insula (pIns), and hippocampus (HC). **b) PE minus Sharpening searchlight analyses:** We found higher correlations for the PE compared to the Sharpening hypothesis model in right AnG, bilateral SMG, PrG, and MTG. **c) Sharpening searchlight analyses:** We found correlations in the frontal pole (FP) and hippocampus. All maps in (**a**)–(**c**) are shown at p(FWE) < .05 (peak-corrected), except for Sharpening pool4 in (**c**), shown at p(FWE) < .05 (cluster-corrected), with a cluster-inducing threshold of p(unc.) < .001. Our paper Garlichs & Blank (2024) was published under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/). This figure equals Figure 5 of the published paper.

9.4 Discussion

In our second study, we investigated the neural computational mechanisms of how expectations are combined with sensory information along the ventral face-processing hierarchy by using multivariate RSA (Kriegeskorte et al., 2008; Nili et al., 2014) combined with DCNNs. We induced expectations by presenting scene images associated with one of four faces or presented a neutral scene after which each face was equally likely to appear. Similar paradigms usually showed an expectation suppression effect, i.e., a reduced activation for expected compared to unexpected faces (e.g., Egner et al., 2010; den Ouden et al., 2010; Summerfield et al., 2008). Our aim was to disentangle whether this known univariate expectation suppression effect could be either explained by a reduced PE processing for expected faces, as suggested by the Predictive Coding framework (Friston, 2005), or a sharpened representation of expected face information (e.g., Aitchison & Lengyel, 2017; Blank et al., 2018; Blank & Davis, 2016).

We replicated behavioural findings of a facilitation and assimilation effect, i.e., an accelerated identification of expected compared to unexpected faces, as well as a perceptual shift for ambiguous face morphs towards the expected identity (e.g., Amado et al., 2018; Blank et al., 2015; Schweinberger et al., 1995; Todorova & Neville, 2020). Furthermore, we replicated the univariate expectation suppression effect in parts of the ventral face-processing hierarchy, specifically in pFFA and in a more lateral area in the ITG/MTG. We showed in both regions that the neural representations of face morphs after scene-induced expectations were more similar to the unexpected face identity, arguing in favour of the PE calculation as the computational mechanism behind the univariate effect. This finding corroborates previous research that showed PE processing in FFA, but did not investigate sharpened representations as an alternative computational model (Apps & Tsakiris, 2013; Egner et al., 2010; Zaragoza-Jimenez et al., 2023). Further searchlight analyses revealed a distributed network of PE processing in parietal, temporal, occipital, and frontal regions that have been previously linked to surprise processing and face recognition (e.g., Blank et al., 2014, 2023; Fouragnan et al., 2018; Ham et al., 2013; Lee & Geng, 2017).

In addition, we showed a co-existence of unexpected and sharpened expected face information in OFA, suggesting a simultaneous computation of the PE as well as the confirmation of expected information. This should only be possible if expectations of higher-level regions got transmitted in a hierarchical fashion to lower-level area OFA, as suggested by the Predictive Coding framework (Friston, 2008; Nigam & Schwiedrzik, 2024; Schwiedrzik & Freiwald, 2017). We further found sharpened representations in the frontal pole, in line with previous research showing predictive face information in frontal regions (Summerfield et al., 2006). Taken together, these results suggest a complex interplay of both mechanisms enabling the neural representation and combination of expectations as well as sensory information.

Furthermore, we used the face-trained DCNN VGG-Face (Parkhi et al., 2015) and the object-trained DCNNs VGG-16 (Simonyan & Zisserman, 2015) and ResNet50 (He et al., 2016) to fine-tune our hypothesis RDMs. These networks revealed overlapping as well as diverging representations of face information: while all networks showed PE processing along the face-processing hierarchy, sharpened face representations in OFA were only evident by the DCNN VGG-Face. Furthermore, when directly comparing VGG-Face and VGG-16 – which have the same network architecture but were trained on different datasets – the evidence for PE processing along the face-processing hierarchy (OFA, pFFA, aTL) for VGG-Face was limited to OFA and pFFA, while VGG-16 showed correlations in all regions and outperformed VGG-Face in pFFA and aTL. These diverging findings highlight two important factors: on the one hand, in line with previous studies (Grossman et al., 2019; Ratan Murty et al., 2021), taking object-trained DCNNs to investigate neural face processing can reveal face-related representations that are (less) picked up by face-trained DCNNs. This could relate to the natural evolvement of object-trained networks towards upright face configurations (Zhao et al., 2024) or, in our study, might also be related to higher-level contextual information induced by the scene information that are rather picked up by object-trained networks. On the other hand, the finding of Sharpening in OFA only with VGG-Face suggests a specificity of face-trained networks (Dobs et al., 2023), for example, they might be able to pick up more fine-grained facial features, revealing sharpened face representations in OFA and frontal regions.

Importantly, our findings highlight the advantages of using both types of DCNNs for face recognition and sensitize for careful model selection as these might lead to different interpretations of results. Further research is necessary to disentangle in more detail which information about faces is captured by face-trained compared to object-trained networks (e.g., Dobs et al., 2022; Kanwisher, Gupta, et al., 2023). It would be interesting to see whether other datasets would also reveal a Sharpening of expected face information only with face-trained DCNNs and a more pronounced PE processing with object-trained DCNNs, giving more insights into their facial feature specialization.

10. Discussion

This cumulative dissertation is based on two studies which investigated context-induced expectation effects on the perception and processing of face images. We discovered a complex interplay of bottom-up and top-down processes, evident as an initial guidance and preferred sampling of expected facial features which reversed into an increased sampling of unexpected information over time, as well as a predominant PE processing along the ventral face-processing hierarchy which was complemented by a Sharpening of expected information in lower-level face-specific area OFA and frontal areas.

Our two studies addressed different hypotheses suggested by the overarching framework of Predictive Coding and Bayesian principles (Aitchison & Lengyel, 2017; Clark, 2013; Friston, 2005, 2009; Knill & Pouget, 2004). Our first eye-tracking study was based on the assumption that our brain is constantly trying to reduce the surprise of incoming sensory information (Friston, 2005, 2009; Friston et al., 2012). This can be achieved not only by having accurate predictions about upcoming sensory events, but also actively sampling our environment at locations where we predict the most probable cause of a sensory event with the highest probability (Friston et al., 2012). If the sensory input matches our prediction, this leads to a minimization of the PE and a confirmation and reduction of the uncertainty of our prediction. In line with this, our eye-tracking study showed a preferred sampling of expected face information and predictive saccades to assumed locations of information prior to the actual face. Our study

revealed that this shifts towards an increased sampling of unexpected face information (PE) that might be additionally used to the expected information for model/prediction updates (Aitchison & Lengyel, 2017; Clark, 2013; Friston, 2005, 2008; Summerfield & Koechlin, 2008). Furthermore, our mismatch trials highlighted that, if sensory information strongly deviates from our expectation, it can attract eye movements similarly to expected sensory input (Anderson et al., 2016; Elazary & Itti, 2008). In a more naturalistic setting, it might be relevant for survival to keep our sensory system open to large deviations from our predictions, as danger (e.g., an escaped tiger from a zoo) is not expected, but highly important to notice. One future study could investigate the 'sweet spot' of when predictive power is outbalanced by sensory saliency. An important variable that comes into play here is precision, i.e., the inverse of variance, which can vary for the sensory input as well as for the prediction. Precision of the sensory input depends on its ambiguity and noise level and can be operationalized, for example, by blurring the presented face (e.g., Bruce & Valentine, 1986; Dunovan & Wheeler, 2018; Garlichs & Blank, 2024), providing different durations of face presentation (Leopold et al., 2005; Rieth & Huber, 2010), or varying the percentage of two identities in a face morph, making them clear or more ambiguous (e.g., Blank et al., 2015; de Gardelle et al., 2011; Gao & Wang, 2020; Lee & Geng, 2017; Walther, Schweinberger, & Kovács, 2013). The precision of predictions can be operationalized, for example, by manipulating the transition probabilities between priors and upcoming stimuli (Becker et al., 2024; Blank et al., 2023). The Predictive Coding framework suggests that PEs with high precision will have a greater effect on model updating and perceptual hypotheses by modulating the synaptic gain of PE units (Feldman & Friston, 2010; Friston, 2009, 2010; for reviews, see den Ouden et al., 2012; Hohwy, 2012; Walsh et al., 2020). In a follow-up study, it would be interesting to manipulate the precision of the prior (e.g., transition probabilities) as well as the sensory information (e.g., by different morph levels and/or additional noise on the face images) to pinpoint the interplay when the prediction dominates our perception and when the sensory input deviates too much from our expectation so that our prediction gets discarded (for related reviews for assimilative and contrastive effects, see Mueller et al., 2020; Snyder et al., 2015).

Importantly, although our first study revealed how expectations can guide eye movements (Friston et al., 2012), its methodology did not allow for a differentiation of how expectations and sensory information are neurally combined. One further aspect of Predictive Coding theories is the hierarchical transmission of predictions from higher-level areas to lower-level areas, which in turn compute a PE that is sent back to higher regions for model updating (Friston, 2005, 2008; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999; Summerfield & Koechlin, 2008). With our second study, we revealed the computation of face information PEs along the ventral face-processing hierarchy in line with previous research in the macaque brain (Nigam & Schwiedrzik, 2024; Schwiedrzik & Freiwald, 2017). We extended previous research by testing an alternative model called Sharpening (e.g., Aitchison & Lengyel, 2017; Blank et al., 2018; Blank & Davis, 2016; González-García & He, 2021; Kok et al., 2012; Lee & Mumford, 2003) and found that the univariate expectation suppression effect could rather be explained by PE processing, in line with previous interpretations (den Ouden et al., 2010; Egner et al., 2010; Summerfield et al., 2008). Simultaneously, we found an increased representation of the expected facial features in OFA and frontal regions. This finding expands the Predictive Coding framework by suggesting a simultaneous Sharpening of expected information. This could be interpreted as the Sharpening of prior information in lower face regions, or as the posterior in higher processing areas, i.e., the updated model representation for upcoming scenarios. While our experimental design and temporal resolution did not allow for a distinction between both types of sharpened expected face information (prior and posterior), a follow-up study using time-sensitive methodology (e.g., EEG) or an event-related fMRI design with a longer time period between prior and face onset (Blank et al., 2023) could provide further insights. A prior should be represented after cue onset but *before* face onset, possibly as a pre-activation of a mental face representation (Schweinberger et al., 1995). A posterior should be represented *after* face onset, possibly with a certain delay. Furthermore, in a replication study of our fMRI experiment, a higher spatial resolution (e.g., 7T) could allow to differentiate whether, as suggested by previous research, PEs ('error units') are rather represented in superficial layers while predictions ('representational units') are rather

represented in deeper cortical layers, which would be especially interesting in OFA in which we found the co-existence of PE and sharpened face information (Felleman & Van Essen, 1991; Friston, 2008; Thomas et al., 2024; Walsh et al., 2020).

When comparing the results of both of our studies, on the first glance, we found seemingly contradictory evidence showing a preferred sampling of expected information in the first study and a predominantly PE processing in the second study. One might intuitively wonder whether we did not already found evidence for a Sharpening of expected information in our first study. This discrepancy between eye movements and neural representations can appear puzzling but is not contradictory in terms of the Predictive Coding framework. Both results show a usage of expectations during the perception of faces, with the first targeting the active seeking for hypothesis confirmation using eye movements (Friston et al., 2012), and the second being related to the hierarchical nature of information processing between higher- and lower-level regions and how expectations and sensory input are computationally combined (Aitchison & Lengyel, 2017; Clark, 2013; Friston, 2005, 2009; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999; Summerfield & Koechlin, 2008). It is important to note that Sharpening describes a process on the neural level, i.e., the enhancement of neurons' representations specifically tuned to a stimulus (Kok et al., 2012). While the fixation of expected facial features could theoretically be accompanied by a sharpened neural representation in face-related areas, these two measurements cannot be directly linked to each other given our methodology and experimental setup in the first study.

Similarly, some might find the linkage of behaviour to neural representations partially counterintuitive. As we have shown in both studies, we found an assimilation effect for face morphs, in the second study with a dominance of PE representations throughout the ventral face-processing hierarchy. Correspondingly to the notion that sampling of expected face information does not have to indicate sharpened face representations, the representation of PEs in face morphs did not result in expectation-discarding behaviour. The behavioural assimilation effect could reflect a response bias towards the expected identity (Wickens, 2001) or indicate that the PE was not large enough to dismiss the prior of the expected identity. As

discussed in Garlichs & Blank (2024), participants seemed to have noticed the unexpected sensory information in face morphs, shown by longer RTs for 'expected' responses for face morphs compared to faster RTs in match trials with clear expected faces. It has been shown that using morph levels containing different amounts of the expected identity can influence whether contrastive effects arise and that these effects correlate with ERP components (Walther, Schweinberger, Kaiser, et al., 2013). In a future study, it would be interesting to investigate whether face morphs deviating less or more from the expected face are accompanied by weaker or stronger PE representations and to further examine their possible linkage to behaviour.

Further insights into the effect of expectations on perception and perceptual decisions could be gained by re-analysing our two studies' behavioural, eye-tracking, and neural data using Drift Diffusion Models (DDM; Ratcliff, 1985; Ratcliff et al., 2016; Ratcliff & Hacker, 1981; Ratcliff & McKoon, 2008; Wagenmakers, 2009). These models were originally designed to investigate questions about two-choice perceptual matching tasks. They use RTs and accuracies to model perceptual decisions as an evidence accumulation for either option until a certain threshold is reached and a choice for one option is made. One key aspect of DDMs is that they allow to differentiate which cognitive process expectations influence: for example, a starting point bias towards one perceptual option can be thought of as a mental pre-activation of the expected input which can lead to the respective choice threshold being reached more quickly (de Lange et al., 2013; Mulder et al., 2012; for a review, see Feuerriegel, Blom, et al., 2021). Another possibility is that expectations can modify the drift rate, i.e., facilitate evidence accumulation for one option, thereby reaching the respective decision threshold faster (Diaz et al., 2024; Todorova & Neville, 2020; Yon et al., 2021), or it can be a combination of both biases (Dunovan et al., 2014; van Ravenzwaaij et al., 2012). A further possibility is that expectations lead to a shift in the perceptual threshold, e.g., reducing the accumulated evidence needed for the expected option to reach the respective decision threshold (De Loof et al., 2016; Domenech & Dreher, 2010). In our data, it would be interesting to analyse different aspects using DDMs: firstly, in our second study, the RTs in our partial trials split up by

response ('expected' or 'unexpected') could be analysed. This difference could be due a starting point, drift rate, threshold, or combined bias, and an additional analysis could reveal the underlying cognitive effect of face expectations on the perceptual decision. Additionally, it might be interesting to analyse our neural activations using DDMs by adding covariates (e.g., for the starting point) into the individual General Linear Models (GLM; Mulder et al., 2012). This analysis could reveal whether changes in starting points might be related to frontoparietal regions (Mulder et al., 2012), possibly answering the question whether our assimilation effect and the Sharpening in frontal regions might be related to starting point biases in the form of sharpened expected face information. It could also show whether we can find pre-activations in lower-level regions such as FFA and OFA, in line with the hierarchical viewpoint of Predictive Coding (Friston, 2005, 2008; Lee & Mumford, 2003; Mumford, 1992; Rao & Ballard, 1999; Summerfield & Koechlin, 2008). Lastly, in our eye-tracking study, we could investigate whether in addition to RTs and accuracies, fixations to either perceptual option (i.e., either identity in a face morph) contribute to perceptual decisions by using attentional DDMs (aDDM; Krajbich et al., 2010, 2012; Krajbich & Rangel, 2011; Tavares et al., 2017). An aDDM postulates that evidence accumulation is additionally dependent on where a person is looking, meaning that longer fixations would result in more evidence accumulation for a decision option (Krajbich et al., 2012). Using an aDDM could directly link our first study's finding of a congruency effect, i.e., a longer and more frequent fixation of the expected identity in a face morph, to the observed assimilation effect, similar to studies showing that participants more likely choose an option that is displayed longer (Armel et al., 2008; for a review, see Ting & Gluth, 2024).

Finally, I would like to address possible practical implications of our study findings beyond contributions to the basic (neuro-)psychological research field. Possible neuropsychological groups related to our research are people with prosopagnosia, people with autism spectrum disorder (ASD), or people with schizophrenia. While people with congenital and acquired prosopagnosia are able to recognize single features within a face, their configural processing ability is impaired (Barton et al., 2002; Ramon & Rossion, 2010). Although it was long thought to be an untreatable impairment, trainings focusing on the relations among facial features

improved behavioural face recognition performance in experiments as well as in real life in people with developmental prosopagnosia (DeGutis et al., 2014; DeGutis et al., 2007). In a future study, it might be interesting to track eye movements of healthy participants for a specific set of faces, identifying locations of major fixation that allow the correct identification of these faces. Afterwards, people with prosopagnosia could be trained to focus on these diagnostic features (e.g., by showing red circles around the features as in the training for our first study). In a test phase, eye-tracking could reveal whether participants with this alternate training to fixate diagnostic features would show a higher recognition performance increase than participants who completed the classical configural training by DeGutis et al. (2014), similar to previous trainings showing improvements after focus on facial features (Brunsdon et al., 2006; Schmalzl et al., 2008; for a review, see DeGutis et al., 2014). Real-life generalizability of the face recognition improvements could be measured by comparing face fixations of trained people with prosopagnosia and healthy people while wearing portable eye-tracking mounts (Varela et al., 2023; Vehlen et al., 2021). Further research using fMRI could be used to investigate whether this training would also lead to changes on the neural level. It could be tested whether face representations after facial feature training become more identity-specific and might allow for better identity decoding in people with prosopagnosia. Combining this with the experimental design of our second study, it would be also interesting to see whether this training affects neural representations related to expectation. If the training already leads to a clearer mental representation of the expected face after a cue, this could also result in more identity-specific PE and sharpened representations for face morphs. Similar training and subsequent eye-tracking and fMRI analyses could be performed with people with ASD that typically show less feature-specific scanning behaviour (Wilson et al., 2012), especially of the eyes region (Wang et al., 2015).

Furthermore, it has been argued that in terms of Predictive Coding, which suggests an interplay between top-down and bottom-up information, people with ASD put greater weight on sensory information and consider contextual information less (Friston et al., 2013; Pellicano & Burr, 2012; for reviews, see Palmer et al., 2017; Van de Cruys et al., 2014). It has also been

proposed that people with ASD overweigh the perceptual PE; this could lead to an overly high influence of small PE on model updates, although small irregularities are natural to the variance in our world (Van de Cruys et al., 2014). Similarly, in interpersonal Predictive Coding designs in which the action performed by one person can be anticipated by the action of another person (e.g., waving hands for greeting), people with high-functioning ASD showed reduced usage of the social motion information provided by the other person for the interpretation of an action compared to a control group, further highlighting the different usage of context information when interpreting sensory information (von der Lühe et al., 2016). In people with schizophrenia, it has been suggested that there are deficiencies in deviance detection, with a reduced attention to relevant PEs. Instead, there could be an increased processing of (objectively) neutral or unimportant information, leading to incorrect model updating due to misleading PEs, which can result in delusions and hallucinations (for reviews, see Fletcher & Frith, 2009; Kirihara et al., 2020). This unbalanced PE computation can result from imprecise predictions, leading to imprecise PEs, and even further wrongly updating of predictive models (for a review, see Liddle & Liddle, 2022). In the context of our fMRI study, it would be interesting to investigate in people with ASD and people with schizophrenia whether face morphs would lead to a reduced or enhanced PE processing in the brain, either due to a reduced usage of context information or an overweighing of sensory deviations. By using an adjusted fMRI paradigm with different precisions of prior and sensory information, it would also be interesting to directly compare these groups with a control group to further disentangle the contribution of both information sources (expectations and sensory input) on neural representations and their effect on PE and sharpened representations.

In conclusion, we have shown in two studies using different methodologies and experimental designs how expectations about an upcoming face shape the sampling of face information and how our brain computationally combines expectations with sensory information, differentiating between PE processing and Sharpening. We have provided evidence for characteristics of the Predictive Coding framework, namely an active sampling of expected information as well as a hierarchical organization involving PE computations along

the ventral face-processing hierarchy. Additionally, we found a Sharpening of expected face information in lower region OFA and frontal regions, arguing for a complex interplay of both representations – of expected and unexpected information – during face perception. We extended previous research by supporting the prominent view that the univariate expectation suppression effect could be related to PE processing and used state-of-the-art DCNNs to fine-tune our hypothesis models for neural face representations. This cumulative thesis provides the basis for much further research regarding how our eye-tracking findings can be used to improve neuropsychological phenomena related to face recognition deficits. It further offers insights into Predictive Coding principles in neural face processing which could be directly compared to possibly altered mechanisms in clinical samples.

11.   Zusammenfassung

In dieser Dissertation untersuchten wir Erwartungseffekte in der Gesichterwahrnehmung im Kontext von Predictive Coding. Diese Theorie nimmt an, dass wir unsere Welt durch interne Modelle basierend auf Erfahrungen vorhersagen. Akkurate Vorhersagen reduzieren Überraschung, d.h. die Abweichung zwischen Erwartung und sensorischer Information (Prediction Error, PE). Es wird angenommen, dass Menschen aktive handelnde Personen in ihrer Umwelt sind. Das bedeutet, dass sie Handlungen (z.B. Augenbewegungen) ausführen, um erwartete sensorische Information zu untersuchen und den PE zu reduzieren. Unsere erste Eye-Tracking-Studie bekräftigte dies durch eine bevorzugte und frühe Fixation von erwarteter Gesichtsinformation und das Ausführen prädiktiver Sakkaden zu erwarteter Gesichtsinformation. Über die Zeit wurde vermehrt unerwartete Gesichtsinformation betrachtet. Außerdem wurde erwartete Gesichtsinformation weniger initial fixiert, wenn die sensorische Information stark von der Erwartung abwich.

Die Predictive Coding-Theorie nimmt zudem eine hierarchische Informationsverarbeitung an, so dass höhere kortikale Areale Vorhersagen an niedrigere kortikale Areale senden, die den PE berechnen und diese für Modellupdates an höhere Areale zurücksenden. Bisher wurde angenommen, dass der neuronale ‚expectation suppression effect', d.h. die reduzierte Aktivierung für erwartete im Vergleich zu unerwarteten Gesichtern, durch die Berechnung von höheren PEs für unerwartete Gesichter erklärt werden könnte. Es wurden jedoch wenige alternative Modelle getestet. In unserer zweiten fMRT-Studie testeten wir mit multivariaten Analysen und Deep Convolutional Neural Networks, ob der Effekt besser durch Sharpening erklärt werden könnte, einer Aktivierung von Neuronen spezialisiert auf die erwartete Gesichterinformation. Sharpening würde den expectation suppression effect durch eine insgesamt reduzierte, aber weniger ‚noisy' Aktivierung für erwartete im Vergleich zu unerwarteten Gesichtern begründen. Wir fanden Evidenz für PEs entlang der gesamten ventralen Gesichterverarbeitungshierarchie sowie gleichzeitig geschärfte Gesichter-Informationen im niedrigeren Areal ‚occipital face area' und frontalen Regionen. Zusätzlich fanden wir PEs in Arealen, die den univariaten expectation suppression effect zeigten, was

vorherige Annahmen unterstützt, dass sie der zugrundliegende Prozess für den Effekt sein könnten.

Insgesamt unterstützen unsere Ergebnisse verschiedene Aspekte von Predictive Coding und betonen die Relevanz von sowohl erwartungskonformen als auch -abweichenden Informationen. Sie heben die Rolle von Erwartungen in unserem alltäglichen Leben hervor und lassen vermuten, dass Wahrnehmung sowohl auf Verhaltens- als auch auf neuronaler Ebene auf einem komplexen Zusammenspiel beider Informationsquellen beruht.

12. Zusammenfassung (Englisch)

In this dissertation, we investigated expectation effects on face perception in the context of Predictive Coding. This framework suggests that the brain represents the world by predicting its contents using internal models based on experiences. Accurate predictions reduce the surprise, i.e., the deviation between expectation and sensory information (Prediction Error, PE). Humans are suggested to be active agents in their environment, meaning that they perform actions (e.g., eye movements) towards anticipated locations of expected sensory information to reduce PEs. In our first eye-tracking study, we found compelling evidence in favour of an active sampling of our environment, shown by a preferred and earlier fixation of expected compared to unexpected face information and expectations predictively guiding saccades. We also showed a reversal into an increased sampling of unexpected face information over time as well as a reduction of the initial fixation of the expected facial feature if sensory information differed strongly from expectations.

Secondly, the Predictive Coding framework suggests a hierarchical information processing, with higher cortical areas sending predictions to lower cortical areas, which compute PEs and send them back to higher areas for model updating. Previous literature suggested that the neural expectation suppression effect, i.e., the reduced activation for expected compared to unexpected faces, could be explained by a higher PE for unexpected faces. However, there was a lack in alternative model testing. We tested with multivariate analyses combined with Deep Convolutional Networks whether this effect could be better explained by Sharpening, an activation of neurons tuned to the expected face information. Sharpening would explain the expectation suppression effect as an overall reduced, but less noisy activation for expected compared to unexpected faces. We found evidence for PE processing along the whole ventral face-processing hierarchy as well as a co-occurrence of sharpened face representations in lower area occipital face area and frontal regions. Furthermore, we found PE processing in areas showing the univariate expectation suppression effect, supporting previous assumptions about it being the underlying computational mechanism for that effect.

Taken together, our findings support different aspects of the Predictive Coding framework, enhancing the importance of both expectation-confirming and -deviating information. They highlight the importance of expectations for our everyday life and suggest that perception on the behavioural and neural level is most likely based on a complex interplay of both information sources.

13.   Abkürzungsverzeichnis / List of Abbreviations

aDDM   attentional Drift Diffusion Model

ASD   autism spectrum disorder

aIT   anterior inferotemporal cortex

AnG   angular gyrus

ATFA   anterior temporal face area

aTL   anterior temporal lobe

CI   confidence interval

DCNN   Deep Convolutional Neural Network

DDM   Drift Diffusion Model

EEG   electroencephalography

ERP   event-related potential

FFA   fusiform face area

fMRI   functional magnetic resonance imaging

GLM   General Linear Model

HD-tDCS   high-definition transcranial direct current stimulation

IOG   inferior occipital gyrus

IQR   interquartile range

ISI   inter-stimulus interval

ITG   inferior temporal gyrus

MEG   magnetoencephalography

MFG   middle frontal gyrus

MTG   middle temporal gyrus

OFA   occipital face area

PE   prediction error

pFFA   posterior fusiform face area

RDM   representational dissimilarity analysis

ROI   region of interest

RT   reaction time

SEM   standard error of the mean

STS   superior temporal sulcus

tDCS   transcranial direct current stimulation

TMS   transcranial magnetic stimulation

V1   primary visual cortex

14. Tools & References

Tools

scienceOS: https://www.scienceos.ai/

- Defining/understanding the concepts of associative and repetition priming and getting an overview of their differences ("What is associative priming?")
- Literature search:
    - "Do humans perform predictive saccades during face perception?"
    - "How do humans look at face morphs?"
    - "What do humans look at in face morphs?"

References

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *Current Opinion in Neurobiology*, *46*, 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Alink, A., & Blank, H. (2021). Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage*, *231*, 117824. https://doi.org/10.1016/j.neuroimage.2021.117824

Aloysius, N., & Geetha, M. (2017). A review on deep convolutional neural networks. *2017 International Conference on Communication and Signal Processing (ICCSP)*, 0588–0592. https://doi.org/10.1109/ICCSP.2017.8286426

Althoff, R. R., & Cohen, N. J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 997–1010. https://doi.org/10.1037/0278-7393.25.4.997

Amado, C., Kovács, P., Mayer, R., Ambrus, G. G., Trapp, S., & Kovács, G. (2018). Neuroimaging results suggest the role of prediction in cross-domain priming. *Scientific Reports*, *8*(1), Art. 1. https://doi.org/10.1038/s41598-018-28696-0

Ambrus, G. G., Amado, C., Krohn, L., & Kovács, G. (2019). TMS of the occipital face area modulates cross-domain identity priming. *Brain Structure and Function*, *224*(1), 149–157. https://doi.org/10.1007/s00429-018-1768-0

Anderson, N. C., Donk, M., & Meeter, M. (2016). The influence of a scene preview on eye movement behavior in natural scenes. *Psychonomic Bulletin & Review*, *23*(6), 1794–1801. https://doi.org/10.3758/s13423-016-1035-4

Apps, M. A. J., & Tsakiris, M. (2013). Predictive codes of familiarity and context during the perceptual learning of facial identities. *Nature Communications*, *4*(1), Art. 1. https://doi.org/10.1038/ncomms3698

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396–403. https://doi.org/10.1017/S1930297500000413

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

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, Disgusted, or Afraid?: Studies on the Malleability of Emotion Perception. *Psychological Science*, *19*(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Barton, J. J. S., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*(1), 71–78. https://doi.org/10.1212/WNL.58.1.71

Barton, J. J. S., Radcliffe, N., Cherkasova, M. V., Edelman, J., & Intriligator, J. M. (2006). Information Processing during Face Recognition: The Effects of Familiarity, Inversion, and Morphing on Scanning Fixations. *Perception*, *35*(8), 1089–1105. https://doi.org/10.1068/p5547

Becker, J., Viertler, M., Korn, C. W., & Blank, H. (2024). The pupil dilation response as an indicator of visual cue uncertainty and auditory outcome surprise. *European Journal of Neuroscience*, *59*(10), 2686–2701. https://doi.org/10.1111/ejn.16306

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, *17*(1), 35–55. https://doi.org/10.1080/026432900380472

Bernstein, M., & Yovel, G. (2015). Two neural pathways of face processing: A critical evaluation of current models. *Neuroscience & Biobehavioral Reviews*, *55*, 536–546. https://doi.org/10.1016/j.neubiorev.2015.06.010

Blank, H., Alink, A., & Büchel, C. (2023). Multivariate functional neuroimaging analyses reveal that strength-dependent face expectations are represented in higher-level face-identity areas. *Communications Biology*, *6*(1), 1–10. https://doi.org/10.1038/s42003-023-04508-8

Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Blank, H., Kiebel, S. J., & von Kriegstein, K. (2015). How the human brain exchanges information across sensory modalities to recognize other people: Information Across Sensory Modalities. *Human Brain Mapping*, *36*(1), 324–339. https://doi.org/10.1002/hbm.22631

Blank, H., Spangenberg, M., & Davis, M. H. (2018). Neural Prediction Errors Distinguish Perception and Misperception of Speech. *Journal of Neuroscience*, *38*(27), 6076–6089. https://doi.org/10.1523/JNEUROSCI.3258-17.2018

Blank, H., Wieland, N., & von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. *Neuroscience & Biobehavioral Reviews*, *47*, 717–734. https://doi.org/10.1016/j.neubiorev.2014.10.022

Bonitz, V. S., & Gordon, R. D. (2008). Attention to smoking-related and incongruous objects during scene viewing. *Acta Psychologica*, *129*(2), 255–263. https://doi.org/10.1016/j.actpsy.2008.08.006

Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011). Perception of object–context

relations: Eye-movement analyses in infants and adults. *Developmental Psychology*,

47(2), 364–375. https://doi.org/10.1037/a0021059

Brockmole, J. R., & Henderson, J. M. (2008). Prioritizing new objects for eye fixation in real-

world scenes: Effects of object–scene consistency. *Visual Cognition*, *16*(2–3), 375–

390. https://doi.org/10.1080/13506280701453623

Bruce, V. (2002). Face perception. In *Psychology at the Turn of the Millennium, Volume 1*

(1st ed., pp. 190–205). Psychology Press.

Bruce, V., & Valentine, T. (1985). Identity priming in the recognition of familiar faces. *British

Journal of Psychology*, *76*(3), 373–383. https://doi.org/10.1111/j.2044-

8295.1985.tb01960.x

Bruce, V., & Valentine, T. (1986). Semantic Priming of Familiar Faces. *The Quarterly Journal

of Experimental Psychology Section A*, *38*(1), 125–150.

https://doi.org/10.1080/14640748608401588

Brunas, J., Young, A. W., & Ellis, A. W. (1990). Repetition priming from incomplete faces:

Evidence for part to whole completion. *British Journal of Psychology*, *81*(1), 43–56.

https://doi.org/10.1111/j.2044-8295.1990.tb02344.x

Brunsdon, R., Coltheart, M., Nickels, L., & Joy, P. (2006). Developmental prosopagnosia: A

case analysis and treatment study. *Cognitive Neuropsychology*, *23*(6), 822–840.

https://doi.org/10.1080/02643290500441841

Bülthoff, I., & Newell, F. N. (2017). Crossmodal priming of unfamiliar faces supports early

interactions between voices and faces in person perception. *Visual Cognition*, *25*(4–

6), 611–628. https://doi.org/10.1080/13506285.2017.1290729

Caldara, R., & Seghier, M. L. (2009). The Fusiform Face Area responds automatically to

statistical regularities optimal for face categorization. *Human Brain Mapping*, *30*(5),

1615–1625. https://doi.org/10.1002/hbm.20626

Caldara, R., Seghier, M. L., Rossion, B., Lazeyras, F., Michel, C., & Hauert, C.-A. (2006).

The fusiform face area is tuned for curvilinear patterns with more high-contrasted

elements in the upper part. *NeuroImage*, *31*(1), 313–319.

https://doi.org/10.1016/j.neuroimage.2005.12.011

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and

facial expression. *Nature Reviews Neuroscience*, *6*(8), 641–651.

https://doi.org/10.1038/nrn1724

Cao, X., Ma, X., & Qi, C. (2015). N170 adaptation effect for repeated faces and words.

Neuroscience, *294*, 21–28. https://doi.org/10.1016/j.neuroscience.2015.03.009

Carbon, C.-C., & Ditye, T. (2011). Sustained effects of adaptation on the perception of

familiar faces. *Journal of Experimental Psychology: Human Perception and

Performance*, *37*(3), 615–625. https://doi.org/10.1037/a0019949

Carey, S., De Schonen, S., Ellis, H. D., Bruce, V., Cowey, A., Ellis, A. W., & Perrett, D. I.

(1997). Becoming a face expert. *Philosophical Transactions of the Royal Society of

London. Series B: Biological Sciences*, *335*(1273), 95–103.

https://doi.org/10.1098/rstb.1992.0012

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye

movement guidance in visual search. *Journal of Experimental Psychology: Human

Perception and Performance*, *33*(4), 753–763. https://doi.org/10.1037/0096-

1523.33.4.753

Chakravarthula, P. N., Tsank, Y., & Eckstein, M. P. (2021). Eye movement strategies in face

ethnicity categorization vs. Face identification tasks. *Vision Research*, *186*, 59–70.

https://doi.org/10.1016/j.visres.2021.05.007

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

Coco, M. I., Nuthmann, A., & Dimigen, O. (2020). Fixation-related Brain Potentials during

Semantic Integration of Object–Scene Information. *Journal of Cognitive

Neuroscience*, *32*(4), 571–589. https://doi.org/10.1162/jocn_a_01504

Cornelissen, T. H. W., & Võ, M. L.-H. (2017). Stuck on semantics: Processing of irrelevant

    object-scene inconsistencies modulates ongoing gaze behavior. *Attention,*

    Perception, & Psychophysics, *79*(1), 154–168. https://doi.org/10.3758/s13414-016-

    1203-7

Databricks. (2019, December 2). *What is a Convolutional Layer?*

    https://www.databricks.com/glossary/convolutional-layer

de Gardelle, V., Charles, L., & Kouider, S. (2011). Perceptual awareness and categorical

    representation of faces: Evidence from masked priming. *Consciousness and*

    Cognition, *20*(4), 1272–1281. https://doi.org/10.1016/j.concog.2011.02.001

de Gardelle, V., Stokes, M., Johnen, V. M., Wyart, V., & Summerfield, C. (2013). Overlapping

    multivoxel patterns for two levels of visual expectation. *Frontiers in Human*

    Neuroscience, *7*, 158. https://doi.org/10.3389/fnhum.2013.00158

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

de Lange, F. P., Rahnev, D. A., Donner, T. H., & Lau, H. (2013). Prestimulus Oscillatory

    Activity over Motor Cortex Reflects Perceptual Expectations. *Journal of*

    Neuroscience, *33*(4), 1400–1410. https://doi.org/10.1523/JNEUROSCI.1094-12.2013

De Loof, E., Van Opstal, F., & Verguts, T. (2016). Predictive information speeds up visual

    awareness in an individuation task by modulating threshold setting, not processing

    efficiency. *Vision Research*, *121*, 104–112.

    https://doi.org/10.1016/j.visres.2016.03.002

DeGutis, J., Cohan, S., & Nakayama, K. (2014). Holistic face training enhances face

    processing in developmental prosopagnosia. *Brain*, *137*(6), 1781–1798.

    https://doi.org/10.1093/brain/awu062

DeGutis, J. M., Bentin, S., Robertson, L. C., & D'Esposito, M. (2007). Functional Plasticity in

    Ventral Temporal Cortex following Cognitive Rehabilitation of a Congenital

Prosopagnosic. *Journal of Cognitive Neuroscience*, *19*(11), 1790–1802.

https://doi.org/10.1162/jocn.2007.19.11.1790

DeGutis, J. M., Chiu, C., Grosso, M. E., & Cohan, S. (2014). Face processing improvements

in prosopagnosia: Successes and failures over the last 50 years. *Frontiers in Human

Neuroscience*, *8*, 561. https://doi.org/10.3389/fnhum.2014.00561

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., Kok, P., & de Lange, F. P. (2012). How Prediction Errors Shape

Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*, 548.

https://doi.org/10.3389/fpsyg.2012.00548

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale

hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern

Recognition*, 248–255. https://doi.org/10.1109/CVPR.2009.5206848

Diaz, J. A., Pisauro, M. A., Delis, I., & Philiastides, M. G. (2024). Prior probability biases

perceptual choices by modulating the accumulation rate, rather than the baseline, of

decision evidence. *Imaging Neuroscience*, *2*, 1–19.

https://doi.org/10.1162/imag_a_00338

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over

time. *Nature Communications*, *10*(1), 1258. https://doi.org/10.1038/s41467-019-

09239-1

Dobs, K., Martinez, J., Kell, A. J. E., & Kanwisher, N. (2022). Brain-like functional

specialization emerges spontaneously in deep neural networks. *Science Advances*,

8(11), eabl8913. https://doi.org/10.1126/sciadv.abl8913

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023). Behavioral signatures of face

perception emerge in deep neural networks optimized for face recognition.

Proceedings of the National Academy of Sciences of the United States of America,

120(32), e2220642120. https://doi.org/10.1073/pnas.2220642120

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K.

    P., Konkle, T., van Gerven, M. A. J., Kriegeskorte, N., & Kietzmann, T. C. (2023). The

    neuroconnectionist research programme. *Nature Reviews Neuroscience*, *24*(7), 431–

    450. https://doi.org/10.1038/s41583-023-00705-w

Domenech, P., & Dreher, J.-C. (2010). Decision Threshold Modulation in the Human Brain.

    Journal of Neuroscience, *30*(43), 14305–14317.

    https://doi.org/10.1523/JNEUROSCI.2371-10.2010

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature

    predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*, 210–221.

    https://doi.org/10.1016/j.neuropsychologia.2014.06.024

Dunovan, K., & Wheeler, M. E. (2018). Computational and neural signatures of pre and post-

    sensory expectation bias in inferior temporal cortex. *Scientific Reports*, *8*(1), 13256.

    https://doi.org/10.1038/s41598-018-31678-x

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

Eimer, M., Kiss, M., & Nicholas, S. (2010). Response Profile of the Face-Sensitive N170

    Component: A Rapid Adaptation Study. *Cerebral Cortex*, *20*(10), 2442–2452.

    https://doi.org/10.1093/cercor/bhp312

Elazary, L., & Itti, L. (2008). Interesting objects are visually salient. *Journal of Vision*, *8*(3):3,

    1–15. https://doi.org/10.1167/8.3.3

Ellis, A. W., Young, A. W., & Flude, B. M. (1990). Repetition priming and face processing:

    Priming occurs within the system that responds to the identity of a face. *The Quarterly*

    Journal of Experimental Psychology Section A, *42*(3), 495–512.

    https://doi.org/10.1080/14640749008401234

Farsal, W., Anter, S., & Ramdani, M. (2018). Deep Learning: An Overview. *Proceedings of*

    the 12th International Conference on Intelligent Systems: Theories and Applications,

    38, 1–6. https://doi.org/10.1145/3289402.3289538

Feldman, H., & Friston, K. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, *1*(1), 1–47. https://doi.org/10.1093/cercor/1.1.1-a

Feuerriegel, D., Blom, T., & Hogendoorn, H. (2021). Predictive activation of sensory representations as a source of evidence in perceptual decision-making. *Cortex*, *136*, 140–146. https://doi.org/10.1016/j.cortex.2020.12.008

Fletcher, P. C., & Frith, C. D. (2009). Perceiving is believing: A Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, *10*(1), 48–58. https://doi.org/10.1038/nrn2536

Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping*, *39*(7), 2887–2906. https://doi.org/10.1002/hbm.24047

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. S. (2009). The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *NeuroImage*, *44*(2), 569–580. https://doi.org/10.1016/j.neuroimage.2008.09.011

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. (2008). Hierarchical Models in the Brain. *PLOS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

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

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as Hypotheses:

Saccades as Experiments. *Frontiers in Psychology*, *3*, 151.

https://doi.org/10.3389/fpsyg.2012.00151

Friston, K. J., Lawson, R., & Frith, C. D. (2013). On hyperpriors and hypopriors: Comment on

Pellicano and Burr. *Trends in Cognitive Sciences*, *17*(1), 1.

https://doi.org/10.1016/j.tics.2012.11.003

Fu, G., Hu, C. S., Wang, Q., Quinn, P. C., & Lee, K. (2012). Adults Scan Own- and Other-

Race Faces Differently. *PLOS ONE*, *7*(6), e37688.

https://doi.org/10.1371/journal.pone.0037688

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a

mechanism of pattern recognition unaffected by shift in position. *Biological

Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/BF00344251

Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a

mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and

Cybernetics*, *SMC-13*(5), 826–834. https://doi.org/10.1109/TSMC.1983.6313076

Gao, Y., & Wang, X. (2020). A proportionally suppressed and prolonged LPP acts as a

neurophysiological correlate of face identity aftereffect. *Brain Research*, *1746*,

146969. https://doi.org/10.1016/j.brainres.2020.146969

Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected

information across the face-processing hierarchy. *Nature Communications*, *15*(1),

3407. https://doi.org/10.1038/s41467-024-47749-9

Garlichs, A., Lustig, M., Gamer, M., & Blank, H. (2024). Expectations guide predictive eye

movements and information sampling during face recognition. *iScience*, *27*(10),

110920. https://doi.org/10.1016/j.isci.2024.110920

Goesaert, E., & Beeck, H. P. O. de. (2013). Representations of Facial Identity Information in

the Ventral Visual Stream Investigated with Multivoxel Pattern Analyses. *Journal of

Neuroscience*, *33*(19), 8549–8558. https://doi.org/10.1523/JNEUROSCI.1829-

12.2013

González-García, C., & He, B. J. (2021). A Gradient of Sharpening Effects by Perceptual

    Prior across the Human Cortical Hierarchy. *Journal of Neuroscience*, *41*(1), 167–178.

    https://doi.org/10.1523/JNEUROSCI.2023-20.2020

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

Grill-Spector, K., Weiner, K. S., Kay, K., & Gomez, J. (2017). The Functional Neuroanatomy

    of Human Face Perception. *Annual Review of Vision Science*, *3*(1), 167–196.

    https://doi.org/10.1146/annurev-vision-102016-061214

Grossman, S., Gaziv, G., Yeagle, E. M., Harel, M., Mégevand, P., Groppe, D. M., Khuvis, S.,

    Herrero, J. L., Irani, M., Mehta, A. D., & Malach, R. (2019). Convergent evolution of

    face spaces across human face-selective neuronal groups and deep convolutional

    networks. *Nature Communications*, *10*(1), 4934. https://doi.org/10.1038/s41467-019-

    12623-6

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

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, G.,

    Cai, J., & Chen, T. (2018). Recent advances in convolutional neural networks. *Pattern*

    Recognition, *77*, 354–377. https://doi.org/10.1016/j.patcog.2017.10.013

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the Representation

    of Identity from Head View Along the Human Face Processing Pathway. *Cerebral*

    Cortex, *27*(1), 46–53. https://doi.org/10.1093/cercor/bhw344

Ham, T. E., de Boissezon, X., Leff, A., Beckmann, C., Hughes, E., Kinnunen, K. M., Leech,

    R., & Sharp, D. J. (2013). Distinct Frontal Networks Are Involved in Adapting to

Internally and Externally Signaled Errors. *Cerebral Cortex*, *23*(3), 703–713.
https://doi.org/10.1093/cercor/bhs056

Han, N. X., Chakravarthula, P. N., & Eckstein, M. P. (2021). Peripheral facial features guiding
eye movements and reducing fixational variability. *Journal of Vision*, *21*(8):7, 1–30.
https://doi.org/10.1167/jov.21.8.7

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system
for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.
https://doi.org/10.1016/S1364-6613(00)01482-0

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in
natural vision. *Experimental Brain Research*, *217*(1), 125–136.
https://doi.org/10.1007/s00221-011-2979-2

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition.
Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,
770–778. https://doi.org/10.1109/CVPR.2016.90

Heisz, J. J., & Shore, D. I. (2008). More efficient scanning for familiar faces. *Journal of
Vision*, *8*(1):9, 1–10. https://doi.org/10.1167/8.1.9

Henson, R. N. (2016). Repetition suppression to faces in the fusiform face area: A personal
and dynamic journey. *Cortex*, *80*, 174–184.
https://doi.org/10.1016/j.cortex.2015.09.012

Henson, R. N. A., Shallice, T., Gorno-Tempini, M. L., & Dolan, R. J. (2002). Face Repetition
Effects in Implicit and Explicit Memory Tests as Measured by fMRI. *Cerebral Cortex*,
12(2), 178–186. https://doi.org/10.1093/cercor/12.2.178

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging Evidence for Dissociable Forms
of Repetition Priming. *Science*, *287*(5456), 1269–1272.
https://doi.org/10.1126/science.287.5456.1269

Hermann, P., Grotheer, M., Kovács, G., & Vidnyánszky, Z. (2017). The relationship between
repetition suppression and face perception. *Brain Imaging and Behavior*, *11*(4),
1018–1028. https://doi.org/10.1007/s11682-016-9575-9

Hohwy, J. (2012). Attention and Conscious Perception in the Hypothesis Testing Brain. *Frontiers in Psychology*, *3*, 96. https://doi.org/10.3389/fpsyg.2012.00096

Hsiao, J. H., & Cottrell, G. (2008). Two Fixations Suffice in Face Recognition. *Psychological Science*, *19*(10), 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x

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

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, *160*(1), 106–154. https://doi.org/10.1113/jphysiol.1962.sp006837

Hubel, D. H., & Wiesel, T. N. (1965). Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, *28*(2), 229–289. https://doi.org/10.1152/jn.1965.28.2.229

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research*, *51*(10), 1192–1205. https://doi.org/10.1016/j.visres.2011.03.010

Ishai, A., Bikle, P. C., & Ungerleider, L. G. (2006). Temporal dynamics of face repetition suppression. *Brain Research Bulletin*, *70*(4), 289–295. https://doi.org/10.1016/j.brainresbull.2006.06.002

Ishai, A., Pessoa, L., Bikle, P. C., & Ungerleider, L. G. (2004). Repetition suppression of faces is modulated by emotion. *Proceedings of the National Academy of Sciences*, *101*(26), 9827–9832. https://doi.org/10.1073/pnas.0403559101

Ishai, A., Schmidt, C. F., & Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Research Bulletin*, *67*(1), 87–93. https://doi.org/10.1016/j.brainresbull.2005.05.027

Jaquetti, P. F., Pilla, V., Borba, G. B., & Gamba, H. R. (2022). VGG FACE Fine-Tuning for Classification of Facial Expression Images of Emotion. In T. F. Bastos-Filho, E. M. de Oliveira Caldeira, & A. Frizera-Neto (Eds.), *XXVII Brazilian Congress on Biomedical*

Engineering (pp. 1539–1546). Springer International Publishing.

https://doi.org/10.1007/978-3-030-70601-2_226

Jemel, B., Pisani, M., Rousselle, L., Crommelinck, M., & Bruyer, R. (2005). Exploring the functional architecture of person recognition system with event-related potentials in a within- and cross-domain self-priming of faces. *Neuropsychologia*, *43*(14), 2024–2040. https://doi.org/10.1016/j.neuropsychologia.2005.03.016

Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., Kaufman, J., & Young, A. W. (2017). Temporal and spatial localization of prediction-error signals in the visual brain. *Biological Psychology*, *125*, 45–57. https://doi.org/10.1016/j.biopsycho.2017.02.004

Jonas, J., Jacques, C., Liu-Shuang, J., Brissart, H., Colnat-Coulbois, S., Maillard, L., & Rossion, B. (2016). A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proceedings of the National Academy of Sciences*, *113*(28), E4088–E4097. https://doi.org/10.1073/pnas.1522033113

Kanwisher, N., Gupta, P., & Dobs, K. (2023). CNNs reveal the computational implausibility of the expertise hypothesis. *iScience*, *26*(2), 105976. https://doi.org/10.1016/j.isci.2023.105976

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask 'why' questions of minds and brains. *Trends in Neurosciences*, *46*(3), 240–254. https://doi.org/10.1016/j.tins.2022.12.008

Kho, S. K., Keeble, D. R. T., Wong, H. K., & Estudillo, A. J. (2023). Investigating the role of the fusiform face area and occipital face area using multifocal transcranial direct current stimulation. *Neuropsychologia*, *189*, 108663. https://doi.org/10.1016/j.neuropsychologia.2023.108663

Kirihara, K., Tada, M., Koshiyama, D., Fujioka, M., Usui, K., Araki, T., & Kasai, K. (2020). A Predictive Coding Perspective on Mismatch Negativity Impairment in Schizophrenia. *Frontiers in Psychiatry*, *11*, 660. https://doi.org/10.3389/fpsyt.2020.00660

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

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

Kongthong, N., Minami, T., & Nakauchi, S. (2013). Semantic processing in subliminal face stimuli: An EEG and tDCS study. *Neuroscience Letters*, *544*, 141–146. https://doi.org/10.1016/j.neulet.2013.04.002

Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological Correlates of Visual Adaptation to Faces and Body Parts in Humans. *Cerebral Cortex*, *16*(5), 742–753. https://doi.org/10.1093/cercor/bhj020

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The Attentional Drift-Diffusion Model Extends to Simple Purchasing Decisions. *Frontiers in Psychology*, *3*, 193. https://doi.org/10.3389/fpsyg.2012.00193

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, *104*(51), 20600–20605. https://doi.org/10.1073/pnas.0705654104

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis –
connecting the branches of systems neuroscience. *Frontiers in Systems
Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep
Convolutional Neural Networks. *Advances in Neural Information Processing Systems*,
25.
https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c4
5b-Abstract.html

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the
ball. *Nature Neuroscience*, *3*(12), Art. 12. https://doi.org/10.1038/81887

Larsson, J., & Smith, A. T. (2012). fMRI Repetition Suppression: Neuronal Adaptation or
Stimulus Expectation? *Cerebral Cortex*, *22*(3), 567–576.
https://doi.org/10.1093/cercor/bhr119

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A
convolutional neural-network approach. *IEEE Transactions on Neural Networks*, *8*(1),
98–113. https://doi.org/10.1109/72.554195

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
https://doi.org/10.1038/nature14539

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L.
D. (1989). Handwritten Digit Recognition with a Back-Propagation Network.
Proceedings of the Advances in Neural Information Processing Systems (NIPS), 396–
404.
https://proceedings.neurips.cc/paper/1989/hash/53c3bce66e43be4f209556518c2fcb5
4-Abstract.html

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to
document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. Proceedings of
the IEEE. https://doi.org/10.1109/5.726791

Lee, J., & Geng, J. J. (2017). Idiosyncratic Patterns of Representational Similarity in

Prefrontal Cortex Predict Attentional Performance. *Journal of Neuroscience*, *37*(5),

1257–1268. https://doi.org/10.1523/JNEUROSCI.1407-16.2016

Lee, S.-M., Tibon, R., Zeidman, P., Yadav, P. S., & Henson, R. (2022). Effects of face

repetition on ventral visual stream connectivity using dynamic causal modelling of

fMRI data. *NeuroImage*, *264*, 119708.

https://doi.org/10.1016/j.neuroimage.2022.119708

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex.

Journal of the Optical Society of America A, *20*(7), 1434–1448.

https://doi.org/10.1364/JOSAA.20.001434

Leopold, D. A., Rhodes, G., Müller, K.-M., & Jeffery, L. (2005). The dynamics of visual

adaptation to faces. *Proceedings of the Royal Society B: Biological Sciences*,

272(1566), 897–904. https://doi.org/10.1098/rspb.2004.3022

Leopold, D., O'Toole, A., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape

encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94.

https://doi.org/10.1038/82947

Levine, S. M., & Schwarzbach, J. V. (2021). Individualizing Representational Similarity

Analysis. *Frontiers in Psychiatry*, *12*, 729457.

https://doi.org/10.3389/fpsyt.2021.729457

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in

anterior inferior temporal cortex. *Journal of Neurophysiology*, *69*(6), 1918–1929.

https://doi.org/10.1152/jn.1993.69.6.1918

Li, X., Arizpe, J., Rothlein, D., Esterman, M., & DeGutis, J. (2020). Deficient functional MRI

selectivity and connectivity in developmental prosopagnosia is specific to face

regions. *Journal of Vision*, *20*(11), 827. https://doi.org/10.1167/jov.20.11.827

Liddle, P. F., & Liddle, E. B. (2022). Imprecise Predictive Coding Is at the Core of Classical

Schizophrenia. *Frontiers in Human Neuroscience*, *16*, 818711.

https://doi.org/10.3389/fnhum.2022.818711

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. *Cortex*, *53*, 60–77. https://doi.org/10.1016/j.cortex.2014.01.013

Mann, D. L., Nakamoto, H., Logt, N., Sikkink, L., & Brenner, E. (2019). Predictive eye movements when hitting a bouncing ball. *Journal of Vision*, *19*(14):28, 1–21. https://doi.org/10.1167/19.14.28

Martin, D., Cairns, S. A., Orme, E., DeBruine, L. M., Jones, B. C., & Macrae, C. N. (2010). Form-Specific Repetition Priming for Unfamiliar Faces. *Experimental Psychology*, *57*(5), 338–345. https://doi.org/10.1027/1618-3169/a000040

Mattavelli, G., Cattaneo, Z., & Papagno, C. (2011). Transcranial magnetic stimulation of medial prefrontal cortex modulates face expressions processing in a priming task. *Neuropsychologia*, *49*(5), 992–998. https://doi.org/10.1016/j.neuropsychologia.2011.01.038

Mead, C. (1990). Neuromorphic electronic systems. *Proceedings of the IEEE*, *78*(10), 1629–1636. https://doi.org/10.1109/5.58356

Meng, M., Cherian, T., Singal, G., & Sinha, P. (2012). Lateralization of face processing in the human brain. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 2052–2061. https://doi.org/10.1098/rspb.2011.1784

Miller, E. K., Li, L., & Desimone, R. (1991). A Neural Mechanism for Working and Recognition Memory in Inferior Temporal Cortex. *Science*, *254*(5036), 1377–1379. https://doi.org/10.1126/science.1962197

Minar, M. R., & Naher, J. (2018). Recent Advances in Deep Learning: An Overview. *arXiv preprint*. https://doi.org/10.13140/RG.2.2.24831.10403

Mondloch, C. J., Maurer, D., & Ahola, S. (2006). Becoming a Face Expert. *Psychological Science*, *17*(11), 930–934. https://doi.org/10.1111/j.1467-9280.2006.01806.x

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Mueller, R., Utz, S., Carbon, C.-C., & Strobach, T. (2020). Face Adaptation and Face Priming as Tools for Getting Insights Into the Quality of Face Space. *Frontiers in Psychology*, *11*, 166. https://doi.org/10.3389/fpsyg.2020.00166

Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the Brain: A Diffusion Model Analysis of Prior Probability and Potential Payoff. *Journal of Neuroscience*, *32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012

Mumford, D. (1992). On the computational architecture of the neocortex. II The role of cortico-cortical loops. *Biological Cybernetics*, *66*(3), 241–251. https://doi.org/10.1007/BF00198477

Muukkonen, I., Ölander, K., Numminen, J., & Salmela, V. R. (2020). Spatio-temporal dynamics of face perception. *NeuroImage*, *209*, 116531. https://doi.org/10.1016/j.neuroimage.2020.116531

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, *108*(24), 9998–10003. https://doi.org/10.1073/pnas.1102433108

Nigam, T., & Schwiedrzik, C. M. (2024). Predictions enable top-down pattern separation in the macaque face-processing hierarchy. *Nature Communications*, *15*(1), 7196. https://doi.org/10.1038/s41467-024-51543-y

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLOS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Noyes, E., & Jenkins, R. (2019). Deliberate disguise in face identification. *Journal of Experimental Psychology: Applied*, *25*(2), 280–290. https://doi.org/10.1037/xap0000213

Noyes, E., Parde, C. J., Colón, Y. I., Hill, M. Q., Castillo, C. D., Jenkins, R., & O'Toole, A. J. (2021). Seeing through disguise: Getting to know you with a deep convolutional

neural network. *Cognition*, *211*, 104611.

https://doi.org/10.1016/j.cognition.2021.104611

Or, C. C.-F., Peterson, M. F., & Eckstein, M. P. (2015). Initial eye movements during face

identification are optimal and similar across cultures. *Journal of Vision*, *15*(13):12,

1–25. https://doi.org/10.1167/15.13.12

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face Space

Representations in Deep Convolutional Neural Networks. *Trends in Cognitive*

Sciences, *22*(9), 794–809. https://doi.org/10.1016/j.tics.2018.06.006

Pajani, A., Kouider, S., Roux, P., & de Gardelle, V. (2017). Unsuppressible Repetition

Suppression and exemplar-specific Expectation Suppression in the Fusiform Face

Area. *Scientific Reports*, *7*(1), 160. https://doi.org/10.1038/s41598-017-00243-3

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards

volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521–542.

https://doi.org/10.1037/bul0000097

Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep Face Recognition. *Procedings of*

the British Machine Vision Conference 2015, 41.1-41.12.

https://doi.org/10.5244/C.29.41

Pellicano, E., & Burr, D. (2012). When the world becomes 'too real': A Bayesian explanation

of autistic perception. *Trends in Cognitive Sciences*, *16*(10), 504–510.

https://doi.org/10.1016/j.tics.2012.08.009

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face

recognition tasks. *Proceedings of the National Academy of Sciences*, *109*(48),

E3314–E3323. https://doi.org/10.1073/pnas.1214269109

Pinsk, M. A., Arcaro, M., Weiner, K. S., Kalkus, J. F., Inati, S. J., Gross, C. G., & Kastner, S.

(2009). Neural Representations of Faces and Body Parts in Macaque and Human

Cortex: A Comparative fMRI Study. *Journal of Neurophysiology*, *101*(5), 2581–2600.

https://doi.org/10.1152/jn.91198.2008

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the

cortical face perception network. *Experimental Brain Research*, *209*(4), 481–493.

https://doi.org/10.1007/s00221-011-2579-1

Posch, C., Serrano-Gotarredona, T., Linares-Barranco, B., & Delbruck, T. (2014).

Retinomorphic Event-Based Vision Sensors: Bioinspired Cameras With Spiking

Output. *Proceedings of the IEEE*, *102*(10), 1470–1484.

https://doi.org/10.1109/JPROC.2014.2346153

Qawaqneh, Z., Mallouh, A. A., & Barkana, B. D. (2017). Deep Convolutional Neural Network

for Age Estimation based on VGG-Face Model. *arXiv preprint*.

https://doi.org/10.48550/arXiv.1709.01664

Rajimehr, R., Young, J. C., & Tootell, R. B. H. (2009). An anterior temporal face patch in

human cortex, predicted by macaque maps. *Proceedings of the National Academy of

Sciences*, *106*(6), 1995–2000. https://doi.org/10.1073/pnas.0807304106

Ramon, M., & Rossion, B. (2010). Impaired processing of relative distances between

features and of the eye region in acquired prosopagnosia—Two sides of the same

holistic coin? *Cortex*, *46*(3), 374–389. https://doi.org/10.1016/j.cortex.2009.06.001

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

Ratan Murty, N. A., Bashivan, P., Abate, A., DiCarlo, J. J., & Kanwisher, N. (2021).

Computational models of category-selective brain regions enable high-throughput

tests of selectivity. *Nature Communications*, *12*(1), 5540.

https://doi.org/10.1038/s41467-021-25409-6

Ratcliff, R. (1985). Theoretical interpretations of the speed and accuracy of positive and

negative responses. *Psychological Review*, *92*(2), 212–225.

https://doi.org/10.1037/0033-295X.92.2.212

Ratcliff, R., & Hacker, M. J. (1981). Speed and accuracy of same and different responses in

perceptual matching. *Perception & Psychophysics*, *30*(3), 303–307.

https://doi.org/10.3758/BF03214286

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-

Choice Decision Tasks. *Neural Computation*, *20*(4), 873–922.

https://doi.org/10.1162/neco.2008.12-06-420

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model:

Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281.

https://doi.org/10.1016/j.tics.2016.01.007

Rieth, C. A., & Huber, D. E. (2010). Priming and habituation for faces: Individual differences

and inversion effects. *Journal of Experimental Psychology: Human Perception and

Performance*, *36*(3), 596–618. https://doi.org/10.1037/a0018737

Roberts, D. J., Lambon Ralph, M. A., Kim, E., Tainturier, M.-J., Beeson, P. M., Rapcsak, S.

Z., & Woollams, A. M. (2015). Processing deficits for familiar and novel faces in

patients with left posterior fusiform lesions. *Cortex*, *72*, 79–96.

https://doi.org/10.1016/j.cortex.2015.02.003

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and

organization in the brain. *Psychological Review*, *65*(6), 386–408.

https://doi.org/10.1037/h0042519

Rossion, B. (2008). Clarifying the functional neuro-anatomy of face perception by single-case

neuroimaging studies of acquired prosopagnosia. In M. Jenkins & L. R. Harris (Eds.),

Cortical Mechanisms of Vision (pp. 171-207). Cambridge University Press.

https://dial.uclouvain.be/pr/boreal/object/boreal:110940

Rossion, B., & Lochy, A. (2021). Is human face recognition lateralized to the right

hemisphere due to neural competition with left-lateralized visual word recognition? A

critical review. *Brain Structure and Function*, *227*(2), 599–629.

https://doi.org/10.1007/s00429-021-02370-0

Scheller, E., Büchel, C., & Gamer, M. (2012). Diagnostic Features of Emotional Expressions

Are Processed Preferentially. *PLOS ONE*, *7*(7), e41792.

https://doi.org/10.1371/journal.pone.0041792

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of

familiar face recognition and visual scan paths for faces in a child with congenital

prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–729.

https://doi.org/10.1080/02643290802299350

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face

Recognition and Clustering. *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition (CVPR), 815–823. https://www.cv-

foundation.org/openaccess/content_cvpr_2015/html/Schroff_FaceNet_A_Unified_201

5_CVPR_paper.html

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye

movements during emotion recognition in faces. *Journal of Vision*, *14*(13):14, 1–16.

https://doi.org/10.1167/14.13.14

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (2001). Priming the access to names of

famous faces. *British Journal of Psychology*, *92*(2), 303–317.

https://doi.org/10.1348/000712601162202

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces.

Cortex, *80*, 141–153. https://doi.org/10.1016/j.cortex.2015.11.001

Schweinberger, S. R., Pfütze, E.-M., & Sommer, W. (1995). Repetition priming and

associative priming of face recognition: Evidence from event-related potentials.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *21*(3), 722–

736. https://doi.org/10.1037/0278-7393.21.3.722

Schwiedrzik, C. M., & Freiwald, W. A. (2017). High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron*, *96*(1), 89–97. https://doi.org/10.1016/j.neuron.2017.09.007

Schwind, V., Wolf, K., & Henze, N. (2017). FaceMaker—A Procedural Face Generator to Foster Character Design Research. In O. Korn & N. Lee (Eds.), *Game Dynamics: Best Practices in Procedural and Dynamic Game Content Generation* (pp. 95–113). Springer International Publishing. https://doi.org/10.1007/978-3-319-53088-8_6

Shehzad, Z., & McCarthy, G. (2019). Perceptual and Semantic Phases of Face Identification Processing: A Multivariate Electroencephalography Study. *Journal of Cognitive Neuroscience*, *31*(12), 1827–1839. https://doi.org/10.1162/jocn_a_01453

Shi, Y.-Q., & Sun, H. (2019). *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards* (3rd ed.). CRC Press. https://doi.org/10.1201/9781315097954

Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint*. https://doi.org/10.48550/arXiv.1409.1556

Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D., & Melloni, L. (2015). How previous experience shapes perception in different sensory modalities. *Frontiers in Human Neuroscience*, *9*, 594. https://doi.org/10.3389/fnhum.2015.00594

Spotorno, S., & Tatler, B. W. (2017). The elephant in the room: Inconsistency in scene viewing and representation. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(10), 1717–1743. https://doi.org/10.1037/xhp0000456

Stefanics, G., Stephan, K. E., & Heinzle, J. (2019). Feature-specific prediction errors for visual mismatch. *NeuroImage*, *196*, 142–151. https://doi.org/10.1016/j.neuroimage.2019.04.020

Stevenage, S. V., Hale, S., Morgan, Y., & Neil, G. J. (2014). Recognition by association: Within- and cross-modality associative priming with faces and voices. *British Journal of Psychology*, *105*(1), 1–16. https://doi.org/10.1111/bjop.12011

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., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science*, *314*(5803), 1311–1314. https://doi.org/10.1126/science.1132028

Summerfield, C., & Koechlin, E. (2008). A Neural Representation of Prior Information during Perceptual Inference. *Neuron*, *59*(2), 336–347. https://doi.org/10.1016/j.neuron.2008.05.021

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

Summerfield, C., Wyart, V., Mareike Johnen, V., & de Gardelle, V. (2011). Human Scalp Electroencephalography Reveals that Repetition Suppression Varies with Expectation. *Frontiers in Human Neuroscience*, *5*, 67. https://doi.org/10.3389/fnhum.2011.00067

Taigman, Y., Yang, M., Ranzato, M., & Wolf, L. (2014). DeepFace: Closing the Gap to Human-Level Performance in Face Verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1701–1708). https://doi.org/10.1109/CVPR.2014.220

Tavares, G., Perona, P., & Rangel, A. (2017). The Attentional Drift Diffusion Model of Simple Perceptual Decision-Making. *Frontiers in Neuroscience*, *11*, 468. https://doi.org/10.3389/fnins.2017.00468

Thomas, E. R., Haarsma, J., Nicholson, J., Yon, D., Kok, P., & Press, C. (2024). Predictions and errors are distinctly represented across V1 layers. *Current Biology*, *34*(10), 2265–2271. https://doi.org/10.1016/j.cub.2024.04.036

Thome, I., García Alanis, J. C., Volk, J., Vogelbacher, C., Steinsträter, O., & Jansen, A. (2022). Let's face it: The lateralization of the face perception network as measured

with fMRI is not clearly right dominant. *NeuroImage*, *263*, 119587.

https://doi.org/10.1016/j.neuroimage.2022.119587

Ting, C.-C., & Gluth, S. (2024). Unraveling information processes of decision-making with

eye-tracking data. *Frontiers in Behavioral Economics*, *3*, 1384713.

https://doi.org/10.3389/frbhe.2024.1384713

Todorova, L., & Neville, D. A. (2020). Associative and Identity Words Promote the Speed of

Visual Categorization: A Hierarchical Drift Diffusion Account. *Frontiers in Psychology*,

11, 955. https://doi.org/10.3389/fpsyg.2020.00955

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

Tsantani, M., Kriegeskorte, N., McGettigan, C., & Garrido, L. (2019). Faces and voices in the

brain: A modality-general person-identity representation in superior temporal sulcus.

NeuroImage, *201*, 116004. https://doi.org/10.1016/j.neuroimage.2019.07.017

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L.

(2021). FFA and OFA Encode Distinct Types of Face Identity Information. *Journal of

Neuroscience*, *41*(9), 1952–1969. https://doi.org/10.1523/JNEUROSCI.1449-20.2020

Tsao, D. Y., Moeller, S., & Freiwald, W. A. (2008). Comparing face patch systems in

macaques and humans. *Proceedings of the National Academy of Sciences*, *105*(49),

19514–19519. https://doi.org/10.1073/pnas.0809662105

Ufer, C., & Blank, H. (2023). Multivariate analysis of brain activity patterns as a tool to

understand predictive processes in speech perception. *Language, Cognition and

Neuroscience*, *39*(9), 1117–1133. https://doi.org/10.1080/23273798.2023.2166679

Underwood, G., Templeman, E., Lamming, L., & Foulsham, T. (2008). Is attention necessary

for object identification? Evidence from eye movements during the inspection of real-

world scenes. *Consciousness and Cognition*, *17*(1), 159–170.

https://doi.org/10.1016/j.concog.2006.11.008

Vaillant, R., Monrocq, C., & LeCun, Y. (1994). Original approach for the localisation of objects in images. *IEE Proceedings - Vision, Image, and Signal Processing*, *141*(4), 245–250. https://doi.org/10.1049/ip-vis:19941301

Van Belle, G., Ramon, M., Lefèvre, P., & Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. *Frontiers in Psychology*, *1*, 20. https://doi.org/10.3389/fpsyg.2010.00020

Van de Cruys, S., Evers, K., Van der Hallen, R., Van Eylen, L., Boets, B., de-Wit, L., & Wagemans, J. (2014). Precise minds in uncertain worlds: Predictive coding in autism. *Psychological Review*, *121*(4), 649–675. https://doi.org/10.1037/a0037665

van Dyck, L. E., & Gruber, W. R. (2023). Modeling Biological Face Recognition with Deep Convolutional Neural Networks. *Journal of Cognitive Neuroscience*, *35*(10), 1521–1537. https://doi.org/10.1162/jocn_a_02040

van Ravenzwaaij, D., Mulder, M., Tuerlinckx, F., & Wagenmakers, E.-J. (2012). Do the Dynamics of Prior Information Depend on Task Context? An Analysis of Optimal Performance and an Empirical Test. *Frontiers in Psychology*, *3*, 132. https://doi.org/10.3389/fpsyg.2012.00132

Varela, V. P. L., Towler, A., Kemp, R. I., & White, D. (2023). Looking at faces in the wild. *Scientific Reports*, *13*(1), 783. https://doi.org/10.1038/s41598-022-25268-1

Vater, C., & Mann, D. L. (2023). Are predictive saccades linked to the processing of peripheral information? *Psychological Research*, *87*(5), 1501–1519. https://doi.org/10.1007/s00426-022-01743-2

Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., & Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. *Scientific Reports*, *11*(1), 2661. https://doi.org/10.1038/s41598-021-81987-x

Visconti di Oleggio Castello, M., Halchenko, Y. O., Guntupalli, J. S., Gors, J. D., & Gobbini, M. I. (2017). The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Scientific Reports*, *7*(1), 12237. https://doi.org/10.1038/s41598-017-12559-1

Vitkovitch, M., Potton, A., Bakogianni, C., & Kinch, L. (2006). Will Julia Roberts harm Nicole Kidman? Semantic priming effects during face naming. *Quarterly Journal of Experimental Psychology*, *59*(6), 1134–1152. https://doi.org/10.1080/02724980543000178

Vladeanu, M., Lewis, M., & Ellis, H. (2006). Associative priming in faces: Semantic relatedness or simple co-occurrence? *Memory & Cognition*, *34*(5), 1091–1101. https://doi.org/10.3758/BF03193255

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*(3):24, 1–15. https://doi.org/10.1167/9.3.24

Võ, M. L.-H., & Henderson, J. M. (2011). Object–scene inconsistencies do not capture gaze: Evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742–1753. https://doi.org/10.3758/s13414-011-0150-6

Volfart, A., Yan, X., Maillard, L., Colnat-Coulbois, S., Hossu, G., Rossion, B., & Jonas, J. (2022). Intracerebral electrical stimulation of the right anterior fusiform gyrus impairs human face identity recognition. *NeuroImage*, *250*, 118932. https://doi.org/10.1016/j.neuroimage.2022.118932

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K., & Schilbach, L. (2016). Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150373. https://doi.org/10.1098/rstb.2015.0373

Von Helmholtz, H., & Nagel, W. A. (1909). *Handbuch der physiologischen Optik*. Leopold Voss.

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology*, *21*(5), 641–671. https://doi.org/10.1080/09541440802205067

Walsh, K. S., & McGovern, D. P. (2018). Expectation Suppression Dampens Sensory

Representations of Predicted Stimuli. *The Journal of Neuroscience*, *38*(50), 10592–

10594. https://doi.org/10.1523/JNEUROSCI.2133-18.2018

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

Walther, C., Schweinberger, S. R., Kaiser, D., & Kovács, G. (2013). Neural correlates of

priming and adaptation in familiar face perception. *Cortex*, *49*(7), 1963–1977.

https://doi.org/10.1016/j.cortex.2012.08.012

Walther, C., Schweinberger, S. R., & Kovács, G. (2013). Adaptor Identity Modulates

Adaptation Effects in Familiar Face Identification and Their Neural Correlates. *PLOS

ONE*, *8*(8), e70525. https://doi.org/10.1371/journal.pone.0070525

Wang, Q., Xiao, N. G., Quinn, P. C., Hu, C. S., Qian, M., Fu, G., & Lee, K. (2015). Visual

scanning and recognition of Chinese, Caucasian, and racially ambiguous faces:

Contributions from bottom-up facial physiognomic information and top-down

knowledge of racial categories. *Vision Research*, *107*, 67–75.

https://doi.org/10.1016/j.visres.2014.10.032

Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I. (2020). Rapid and dynamic

processing of face pareidolia in the human brain. *Nature Communications*, *11*(1),

4518. https://doi.org/10.1038/s41467-020-18325-8

Wegrzyn, M., Garlichs, A., Heß, R. W. K., Woermann, F. G., & Labudda, K. (2019). The

hidden identity of faces: A case of lifelong prosopagnosia. *BMC Psychology*, *7*(4), 1–

15. https://doi.org/10.1186/s40359-019-0278-z

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb

activations in human ventral temporal cortex. *NeuroImage*, *52*(4), 1559–1573.

https://doi.org/10.1016/j.neuroimage.2010.04.262

Wickens, T. D. (2001). *Elementary Signal Detection Theory*. Oxford University Press.

Wiese, H., & Schweinberger, S. R. (2008). Event-related potentials indicate different processes to mediate categorical and associative priming in person recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1246–1263. https://doi.org/10.1037/a0012937

Wilson, C. E., Palermo, R., & Brock, J. (2012). Visual Scan Paths and Recognition of Facial Identity in Autism Spectrum Disorder and Typical Development. *PLOS ONE*, *7*(5), e37681. https://doi.org/10.1371/journal.pone.0037681

Wolfe, J. M. (2020). Visual Search: How Do We Find What We Are Looking For? *Annual Review of Vision Science*, *6*(1), 539–562. https://doi.org/10.1146/annurev-vision-091718-015048

Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Wong, B., Cronin-Golomb, A., & Neargarder, S. (2005). Patterns of Visual Scanning as Predictors of Emotion Identification in Normal Aging. *Neuropsychology*, *19*(6), 739–749. https://doi.org/10.1037/0894-4105.19.6.739

Wu, D., Zhang, P., Liu, N., Sun, K., & Xiao, W. (2021). Effects of High-Definition Transcranial Direct Current Stimulation Over the Left Fusiform Face Area on Face View Discrimination Depend on the Individual Baseline Performance. *Frontiers in Neuroscience*, *15*, 704880. https://doi.org/10.3389/fnins.2021.704880

Yan, X., Young, A. W., & Andrews, T. J. (2017). The automaticity of face perception is influenced by familiarity. *Attention, Perception, & Psychophysics*, *79*(7), 2202–2211. https://doi.org/10.3758/s13414-017-1362-1

Yang, H., Susilo, T., & Duchaine, B. (2016). The Anterior Temporal Face Area Contains Invariant Representations of Face Identity That Can Persist Despite the Loss of Right FFA and OFA. *Cerebral Cortex*, *26*(3), 1096–1107. https://doi.org/10.1093/cercor/bhu289

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/h0027474

Yon, D., Zainzinger, V., de Lange, F. P., Eimer, M., & Press, C. (2021). Action biases perceptual decisions toward expected outcomes. *Journal of Experimental Psychology: General*, *150*(6), 1225–1236. https://doi.org/10.1037/xge0000826

Young, A. W., Flude, B. M., Hellawell, D. J., & Ellis, A. W. (1994). The nature of semantic priming effects in the recognition of familiar people. *British Journal of Psychology*, *85*(3), 393–411. https://doi.org/10.1111/j.2044-8295.1994.tb02531.x

Yovel, G., & Kanwisher, N. (2005). The Neural Basis of the Behavioral Face-Inversion Effect. *Current Biology*, *15*(24), 2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Zaragoza-Jimenez, N., Niehaus, H., Thome, I., Vogelbacher, C., Ende, G., Kamp-Becker, I., Endres, D., & Jansen, A. (2023). Modeling face recognition in the predictive coding framework: A combined computational modeling and functional imaging study. *Cortex*, *168*, 203–225. https://doi.org/10.1016/j.cortex.2023.05.021

Zhao, Y., Zhen, Z., Liu, X., Song, Y., & Liu, J. (2018). The neural network for face recognition: Insights from an fMRI study on developmental prosopagnosia. *NeuroImage*, *169*, 151–161. https://doi.org/10.1016/j.neuroimage.2017.12.023

Zhao, Z., Chen, J., Lin, Z., & Ying, H. (2024). Face processing emerges from object-trained convolutional neural networks. *arXiv preprint*. https://doi.org/10.48550/arXiv.2405.18800

Zhaoping, L. (2006). Theoretical understanding of the early visual processes by data compression and data selection. *Network: Computation in Neural Systems*, *17*(4), 301–334. https://doi.org/10.1080/09548980600931995

Zhen, Z., Yang, Z., Huang, L., Kong, X., Wang, X., Dang, X., Huang, Y., Song, Y., & Liu, J. (2015). Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. *NeuroImage*, *113*, 13–25. https://doi.org/10.1016/j.neuroimage.2015.03.010

Zhu, T. (2022). Accounting for the last-sampling bias in perceptual decision-making. *Cognition*, *223*, 105049. https://doi.org/10.1016/j.cognition.2022.105049

Zietz, D., & Hollands, M. (2009). Gaze Behavior of Young and Older Adults During Stair Walking. *Journal of Motor Behavior*, *41*(4), 357–366. https://doi.org/10.3200/JMBR.41.4.357-366

15. Erklärung des Eigenanteils an den Publikationen

Garlichs, A., Lustig, M., Gamer, M., & Blank, H. (2024). Expectations guide predictive eye movements and information sampling during face recognition. *iScience, 27*(10). https://doi.org/ 10.1016/j.isci.2024.110920

A.G.: conceptualization, methodology, formal analysis, data curation, writing – original draft, visualization, and project administration. M.L.: conceptualization, methodology, validation, investigation, and writing – review and editing. M.G.: conceptualization, methodology, resources, and writing – review and editing. H.B.: conceptualization, methodology, writing – original draft, supervision, project administration, and funding acquisition.

Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. *Nature Communications*, *15*(1), 3407.

A.G. and H.B. designed the project; A.G. performed experiments and analysed data; A.G. and H.B. wrote the paper.

16. Lebenslauf

Personal information

Name	Annika Garlichs
Date of birth	19.07.1995
Place of birth	Celle

Working experience

10/2020 – 07/2024	Research associate
	Department of Systems Neuroscience
	University Medical Center Hamburg-Eppendorf
	Hamburg, Germany
09/2019 – 03/2020	Student assistant
	Department of Psychology
	Münster University, Germany
01/2017 – 06/2017	Student assistant
	Department of Clinical Psychology With a Focus on Epilepsy Research
	Bielefeld University, Germany

Education

10/2020 – 04/2025	PhD student
	Department of Systems Neuroscience
	University Medical Center Hamburg-Eppendorf
	Hamburg, Germany
10/2018 – 09/2020	M.Sc. in psychology with a focus on cognitive neurosciences
	Thesis: "Neural Correlates of Action Segmentation: The Influence of Object Information in Action Observation"
	Münster University, Germany
10/2013 – 09/2017	B.Sc. in psychology
	Thesis: „Neuropsychologische Erfassung des Wiedererkennens von Gesichtern bei kongenitaler Prosopagnosie: eine Fallstudie"
	Bielefeld University, Germany
08/2005 – 07/2013	Abitur
	Mariengymnasium Warendorf, Germany
2009 – 2012	Foreign language correspondent in English (school-accompanying training)
	IHK Dortmund, Germany

17. Danksagung

Ein sehr großer Dank gilt außerdem unserem gesamten Institut unter der Leitung von Prof. Dr. Christian Büchel. Das ISN hat mir stets einen wunderbaren Arbeitsort mit tollen Kolleg:innen geboten und ich bin dankbar für die Zeit, die ich am UKE verbringen durfte. Ich habe wertvolle Freundschaften geschlossen und möchte mich an dieser Stelle insbesondere bei Lieven Schenk, Karita Ojala, Carina Jaap, Marike Maack, Jan Ostrowski, Maren Wolf, Christoph Wittkamp, Maria Bruntsch, Ying Chu, Sepideh Khoneiveh und Ivana Tanasic bedanken. Vielen Dank für eure Unterstützung in den letzten Jahren und die vielen schönen Erinnerungen, die wir zusammen erlebt haben, seien es Spieleabende im Institut, Cake Wednesday, Escape Rooms, Kino- oder Restaurant-Besuche. Ihr ward nicht nur Kolleg:innen, sondern seid wertgeschätzte Freund:innen, und ich schätze mich glücklich, gleichzeitig mit euch am Institut gewesen zu sein. Natürlich auch ein großes Dankeschön an alle weiteren Personen am Institut, die ich hier nicht alle namentlich erwähnen kann. Das ISN war und bleibt ein Ort von Geborgenheit und ich würde mich freuen, viele von euch weiterhin regelmäßig sehen zu können.

Abschließend möchte ich noch ein großes Dankeschön an meine Familie aussprechen. Vielen Dank an meine Mutter, meine Schwester und meinen Bruder für euren durchgängigen emotionalen Support, der mich auch in schwierigen Zeiten gestützt hat. Ihr habt mich auf dem gesamten Weg begleitet und ich bin stolz mit euch den Abschluss der Doktorandenzeit erleben zu können.

Vielen Dank!

18. Reprints

18.1 Study 1: Expectations guide predictive eye movements and information sampling during face recognition

Garlichs, A., Lustig, M., Gamer, M., & Blank, H. (2024). Expectations guide predictive eye movements and information sampling during face recognition. *iScience, 27*(10), 110920. https://doi.org/ 10.1016/j.isci.2024.110920

Article

Expectations guide predictive eye movements and information sampling during face recognition
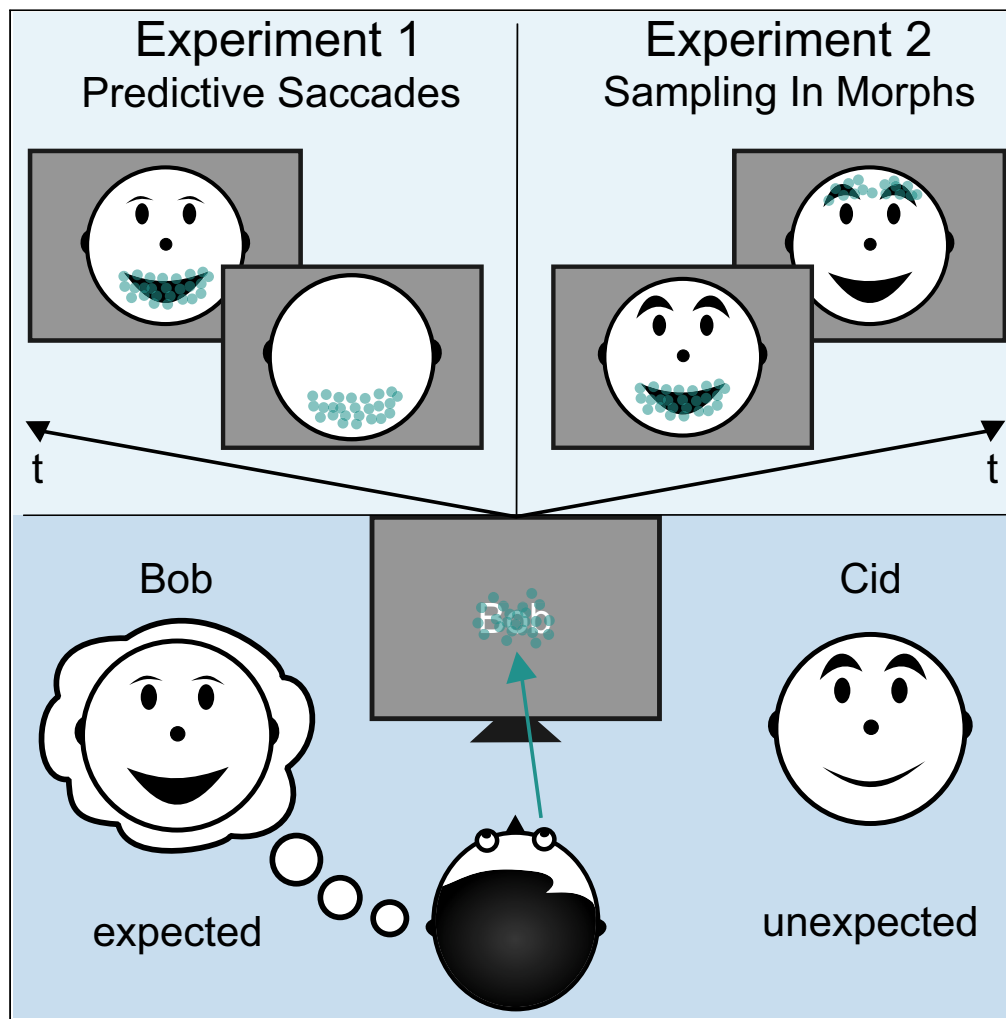
Annika Garlichs,
Mark Lustig,
Matthias Gamer,
Helen Blank

a.garlichs@uke.de (A.G.)
h.blank@uke.de (H.B.)

Highlights
Predictive coding suggests an active sampling of expected information

We studied visual information sampling during face perception

Participants performed predictive saccades and early fixations of expected features

Expectations guide eye movements toward locations of interest

Article

Expectations guide predictive eye movements and information sampling during face recognition

Annika Garlichs,[1,2,*] Mark Lustig,[1,3] Matthias Gamer,[4] and Helen Blank[1,2,5,6,*]

SUMMARY

Context information has a crucial impact on our ability to recognize faces. Theoretical frameworks of predictive processing suggest that predictions derived from context guide sampling of sensory evidence at informative locations. However, it is unclear how expectations influence visual information sampling during face perception. To investigate the effects of expectations on eye movements during face anticipation and recognition, we conducted two eye-tracking experiments (*n* = 34, each) using cued face morphs containing expected and unexpected facial features, and clear expected and unexpected faces. Participants performed predictive saccades toward expected facial features and fixated expected more often and longer than unexpected features. In face morphs, expected features attracted early eye movements, followed by unexpected features, indicating that top-down as well as bottom-up information drives face sampling. Our results provide compelling evidence that expectations influence face processing by guiding predictive and early eye movements toward anticipated informative locations, supporting predictive processing.

INTRODUCTION

Context provides important information for recognizing faces as one of the most important stimuli in everyday life.[1–3] While behavioral evidence indicates that expectations can facilitate face recognition, especially when the face is degraded or ambiguous,[4–7] it is unclear whether context already guides how visual information is sampled when we look at a face. Frameworks of predictive processing suggest an active sampling of informative locations, guided by prior beliefs.[8,9] This process may direct eye movements toward expected sources of sensory input, aiming to reduce uncertainty in predictions.[10]

Context efficiently influences gazing behavior in complex everyday scenes, making one faster at locating a pot than a printer in a kitchen due to semantic and episodic knowledge about typical occurrences of objects[11–14]; for review, see Wolfe.[15] Expectations can also guide predictive saccades ahead of time to anticipated regions of interest, e.g., toward ball locations during fast-paced games[16–18] or prospective steps when walking down a staircase.[19] While examining objects or scenes, expectations can lead to an earlier and increased sampling of either expected information[14,20] or unexpected information.[21,22]

Previous eye-tracking studies on face perception revealed that humans typically first look at a point slightly below the eyes,[23,24] before fixating on the eyes and mouth,[25,26] areas specifically relevant for emotion recognition.[27,28] Crucially, gaze patterns to faces can be modulated by context such as emotional content (for review, see Aviezer et al.[29]), familiarity,[30–33] or ethnicity of the anticipated face.[34] For instance, fixation patterns for angry faces are altered if the face is shown in a disgust context (e.g., the person is holding a trash bag), counterbalancing the typical bias toward the eye region. Conversely, if a disgusted face is shown in an anger context (e.g., the person is showing a fist), the usually symmetrical scanning pattern to the mouth and eye regions is shifted toward the eyes, which is already evident at the first fixation.[35] Effects of familiarity on viewing behavior during face identity recognition are mixed. In some studies, there were more fixations toward the eye region for novel or unfamiliar faces, a location allowing holistic face processing, whereas famous or personally familiar faces had overall reduced upper-face fixations.[31,33] In contrast, other studies using name recall and fame judgment instead of familiarity judgment tasks reported an increased sampling of the eye region for familiar or famous compared to unfamiliar faces.[30,32] Gaze patterns toward facial features also vary depending on whether viewers anticipate a displayed face morph to be Chinese or Caucasian, directing the gaze toward either the nose or the eyes, respectively.[34] Taken together, these findings suggest that expectations derived from surrounding emotional context, familiarity, or general knowledge about racial categories influence how people look at face images.

---

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[2]Hamburg Brain School, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[3]Department of Psychology, University of Hamburg, Hamburg, Germany
[4]Department of Psychology, University of Würzburg, Würzburg, Germany
[5]Predictive Cognition, Research Center One Health Ruhr of the University Alliance Ruhr, Faculty of Psychology, Ruhr-University Bochum, Bochum, Germany
[6]Lead contact
*Correspondence: a.garlichs@uke.de (A.G.), h.blank@uke.de (H.B.)
https://doi.org/10.1016/j.isci.2024.110920

We investigated how expectations about an upcoming face with particular facial features affect eye movements during anticipation and perception of that face in two preregistered experiments. Participants learned to associate images of four male faces with names in both experiments. Each face was characterized by one distinct facial feature, such as a high forehead or a wide chin. In each trial, expectations about the upcoming face and hence its distinct facial feature were induced by a name prior. Furthermore, we showed face morphs between two identities, containing the expected and an unexpected identity. In Experiment 1, we hypothesized that participants would use context information to perform predictive saccades toward locations containing the expected facial feature. To enhance the use of predictive information, we inserted a long time interval before image presentation and provided only limited sensory information by restricting stimulus duration (Figure 1D). In Experiment 2, we investigated whether context information influences how participants sample face information. We hypothesized that expectations modulate active sampling, such that locations associated with the expected identity are preferentially sampled, evident by (1) an initial fixation at the location of the expected facial feature in all faces as well as (2) more fixations and longer dwell time on the expected feature in face morphs, containing expectation-compliant as well as -incompliant information (Figures 1B and 1E).

RESULTS

Facilitation and assimilation effect due to expectation

In both experiments, prior expectations induced a facilitation effect, i.e., faster classification of expected compared to (partially) unexpected faces as evident by the main effects of "condition" (*match, mismatch, partial*) on reaction times (RTs) (Experiment 1: $F(1.97, 65.09) = 77.72$, $p < 0.001$, $\eta^2 = 0.10$; Experiment 2: $F(1.43, 47.15) = 182.40$, $p < 0.001$, $\eta^2 = 0.47$; Figures 2A and 2C). Specifically, participants were faster in classifying a clear face as expected or unexpected if the face matched their expectation compared to when it did not match their expectation at all (*match* vs. *mismatch*: Experiment 1: *match*: $M = 670.07$ ms, $SD = 101.16$ ms; *mismatch*: $M = 719.86$ ms, $SD = 102.22$ ms; $t(33) = -8.05$, $p < 0.001$, 95% confidence interval [CI] [$-65.40, -34.20$], $d = 1.38$, Figure 2A; Experiment 2: *match*: $M = 1439.10$ ms, $SD = 290.96$ ms; *mismatch*: $M = 1761.00$ ms, $SD = 321.00$ ms; $t(33) = -12.19$, $p < 0.001$, 95% CI [$-389, -255$], $d = 2.09$, Figure 2C). Similarly, expected faces were classified faster than face morphs containing the expected as well as an unexpected facial feature (*match* vs. *partial*: Experiment 1: *partial*: $M = 751.16$ ms, $SD = 103.96$ ms; $t(33) = -11.84$, $p < 0.001$, 95% CI [$-98.40, -63.80$], $d = 2.03$; Experiment 2: *partial*: $M = 2207.50$ ms, $SD = 398.45$ ms; $t(33) = -15.43$, $p < 0.001$, 95% CI [$-894, -643$], $d = 2.65$). Lastly, unexpected faces that completely mismatched expectations were classified faster than face morphs that contained information of two identities (*mismatch* vs. *partial*: Experiment 1: $t(33) = -4.72$, $p < 0.001$, 95% CI [$-48.00, -14.60$], $d = 0.81$, Figure 2A; Experiment 2: $t(33) = -10.76$, $p < 0.001$, 95% CI [$-551, -342$], $d = 1.85$, Figure 3A).

Prior expectations led to an assimilation effect in both experiments; i.e., face morphs were more often classified as the expected identity (Experiment 1: $t(33) = 4.12$, $p < 0.001$, 95% CI [54.02, 61.85], $d = 0.71$, Figure 2B; Experiment 2: $t(33) = 6.75$, $p < 0.001$, 95% CI [58.11, 65.10], $d = 1.16$, Figure 2D).

In addition, participants performed well in identifying the clear faces in both experiments. The presented face was correctly identified in the *match* (Experiment 1: $M = 86.43\%$, $SD = 9.84\%$, $Z = 5.08$, $p < 0.001$, Wilcoxon's $r = 0.87$; Experiment 2: $M = 98.07\%$, $SD = 1.52\%$, $Z = 5.11$, $p < 0.001$, Wilcoxon's $r = 0.88$) as well as in the *mismatch* condition (Experiment 1: $M = 83.86\%$, $SD = 11.18\%$, $Z = 5.04$, $p < 0.001$, Wilcoxon's $r = 0.87$; Experiment 2: $M = 87.93\%$, $SD = 10.65\%$, $Z = 5.08$, $p < 0.001$, Wilcoxon's $r = 0.87$). In Experiment 1, the accuracies of both conditions did not differ significantly from each other (Experiment 1: $Z = 1.41$, $p = 0.16$, Wilcoxon's $r = 0.24$), while, in Experiment 2, the accuracies in the *match* condition were higher compared to the accuracies in the *mismatch* condition ($Z = 5.01$, $p < 0.001$, Wilcoxon's $r = 0.86$).

Predictive saccades toward expected facial feature

Experiment 1 investigated whether participants use expectations to perform anticipatory eye movements. Indeed, expectations led to predictive saccades during the inter-stimulus interval (ISI) toward face locations associated with the expected distinct facial feature. A two-way ANOVA revealed the main effect of "expectation" ($F(1,33) = 157.86$, $p < 0.001$, $\eta_p^2 = 0.83$), as well as a significant interaction ($F(1,33) = 14.27$, $p < 0.001$, $\eta_p^2 = 0.30$). Specifically, participants performed more saccades to a region of interest (ROI) if the corresponding facial feature was expected (main effect "expectation": expected: $M = 51.70\%$, $SD = 18.20\%$; unexpected: $M = 7.22\%$, $SD = 4.17\%$, Figures 3A and 3B). There was no main effect of "saccade" ($F(1,33) = 0.01$, $p = 0.91$, $\eta_p^2 = 0.0004$). We report the corresponding *post hoc t* tests (Bonferroni-corrected) for completeness. The first and second saccade landed more often in an ROI in trials in which its facial feature had been expected compared to when it had not been expected, respectively (first/expected: $M = 50.30\%$, $SD = 16.80\%$; first/unexpected: $M = 8.71\%$, SD = 4.20%; $t(33) = 11.98$, $p < 0.001$, 95% CI [31.86, 51.35], $d = 2.06$; second/expected: $M = 53.10\%$, $SD = 19.70\%$; second/unexpected: $M = 5.74\%$, $SD = 3.63\%$; $t(33) = 12.59$, $p < 0.001$, 95% CI [36.77, 57.87], $d = 2.16$. Furthermore, this expectation effect was also evident when directly comparing the first and second saccade; i.e., the first saccade landed more often in an ROI if its facial feature had been expected compared to the second saccade in trials in which it had not been expected (first/expected vs. second/unexpected: $t(33) = 13.24$, $p < 0.001$, 95% CI [35.12, 54.03], $d = 2.12$). This was also evident for the reverse effect; i.e., the second saccade landed more often in an ROI in trials in which its facial feature had been expected compared to the first saccade in trials in which its facial feature had not been expected (second/expected vs. first/unexpected: $t(33) = 11.23$, $p < 0.001$, 95% CI [33.27, 55.43], $d = 1.93$). Lastly, for expected facial features, there was no difference in how often the first and second saccade landed in the respective ROI (first/expected vs. second/expected: $t(33) = -1.63$, $p = 0.68$, 95% CI [$-7.47, 1.99$], $d = -0.28$), while second saccades landed more frequently in the unexpected ROIs compared to first saccades ($t(33) = 6.79$, $p < 0.001$, 95% CI [1.74, 4.20], $d = 1.16$).
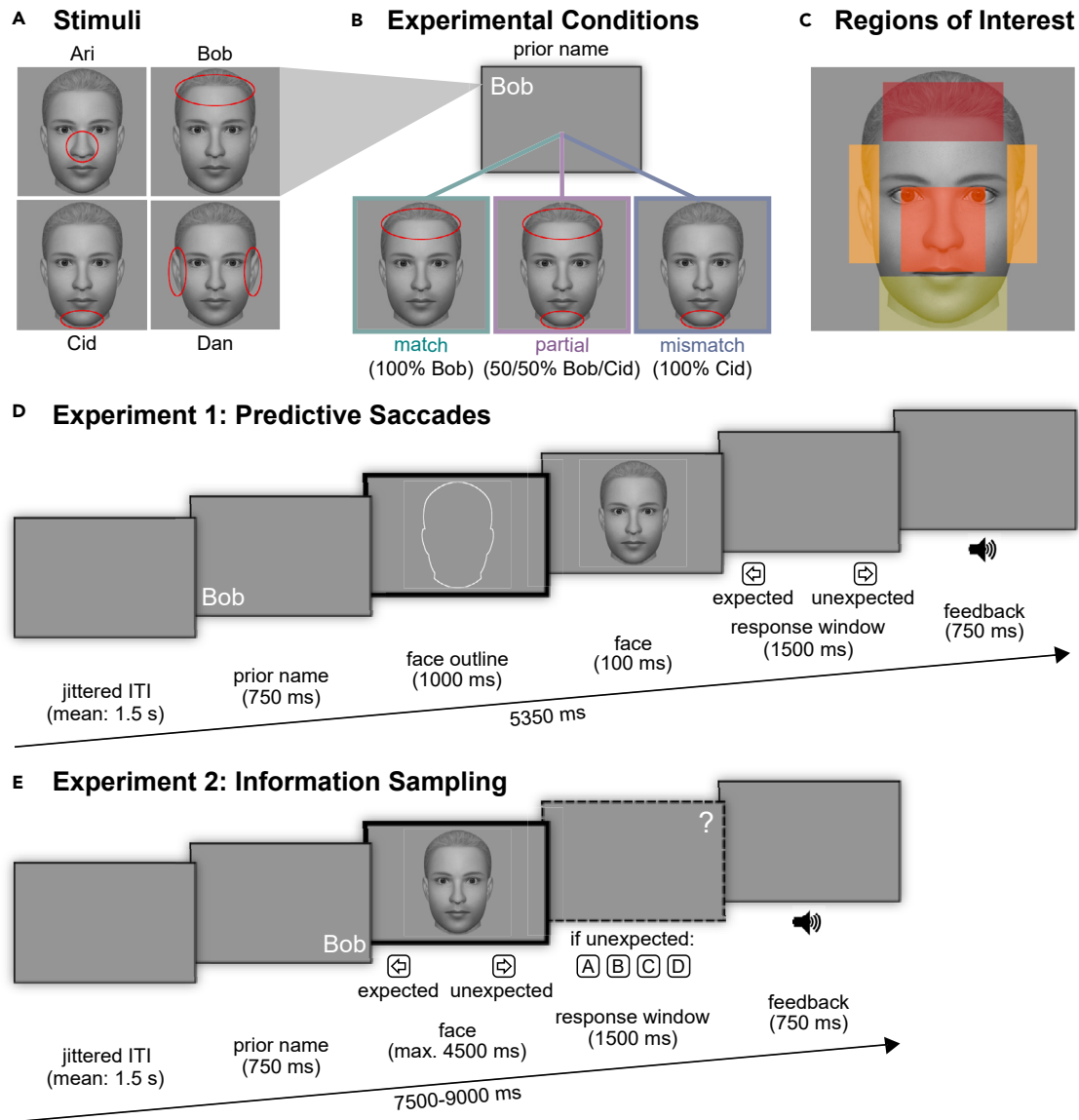
Figure 1. Experimental procedure

(A) Stimuli: Images of four identities were used. Each identity had one distinct facial feature (here shown with a red circle).

(B) Experimental conditions: in both experiments, a name prior was depicted randomly in one of the four corners. Afterward, a face was presented. The face could be expected (*match*), unexpected (*mismatch*), or a morph containing the expected as well as an unexpected identity (*partial*; 50/50% morphs).

(C) Regions of interest (ROIs): We used four ROIs of identical size that covered the distinct features of the four identities.

(D) Exemplary trial of Experiment 1: after the presentation of a name prior, an outline of the "base face" was presented. Next, a face was shown for 100 ms. The task for the participant was to indicate whether the face had been "expected' or "unexpected" based on the preceding name. Participants received auditory feedback on whether they answered correctly, incorrectly, or too slowly. The thick black frame indicates the time window for our eye-tracking analyses. ITI, inter-trial interval (jittered between 1.25 and 1.75 s, mean 1.5 s).

(E) Exemplary trial of Experiment 2: in contrast to Experiment 1, the face was presented up to 4,500 ms or until a button press. The first task for the participants was to indicate whether the presented face was "expected" or "unexpected" based on the preceding name. If participants answered "unexpected," participants were required to answer with one out of four buttons which identity they saw in the presented face. Auditory feedback (too slow) was provided.

Expectations guide fixations during face recognition

In Experiment 2, across all conditions (i.e., *match*, *mismatch*, and *partial*), expectations influenced the order in which the expected ROI was fixated (χ^2(3, N = 34) = 27.47, $p < 0.001$, V = 0.03; Figure 4A). *Post hoc* tests revealed that this effect was mainly driven by the expected ROI being fixated first or second more often than third or fourth (first vs. third: $z = 4.83$, $p < 0.001$, 95% CI [0.57, 0.67], $h = 0.24$; first
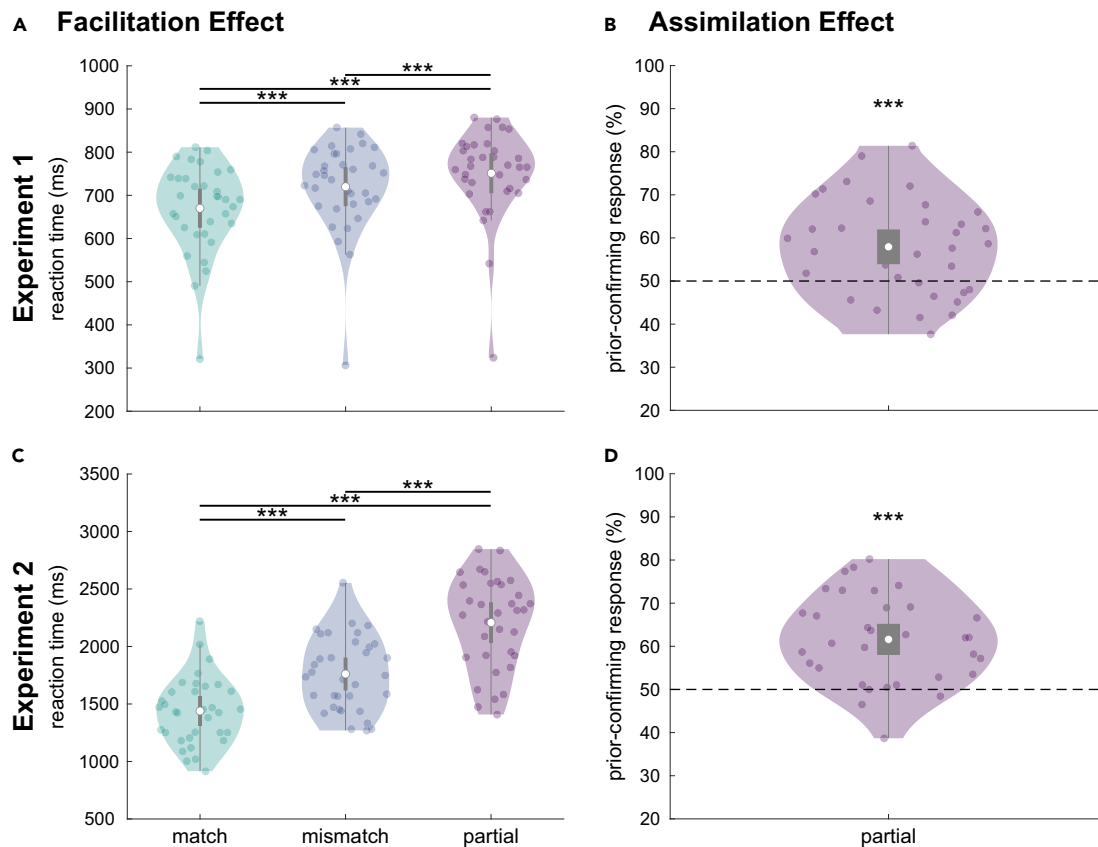
Figure 2. Behavioral results of experiments 1 and 2

(A and C) Facilitation effect: participants reacted faster to expected (*match* condition) compared to unexpected (*mismatch* condition) or morphed (*partial* condition) faces. Dots represent single participants. The white dots represent means, gray rectangles 95% confidence intervals, and the lower and upper whiskers $Q_{1/3}$ −/+ 1.5*interquartile range. In (A) and (C), black lines indicate $p < 0.001$.

(B and D) Assimilation effect: participants responded more often to have perceived the expected identity in a face morph (partial condition). The dashed line represents the chance level, i.e., perceiving the expected or unexpected identity equally often. Asterisks indicate $p < 0.001$. The upper and lower rows depict the results for experiments 1 and 2, respectively.

vs. fourth: $z = 4.84$, $p < 0.001$, 95% CI [0.59, 0.70], $h = 0.31$; second vs. third: $z = 5.27$, $p < 0.001$, 95% CI [0.62, 0.76], $h = 0.39$; second vs. fourth: $z = 5.25$, $p < 0.001$, 95% CI [0.63, 0.79], $h = 0.45$). The number of times the expected ROI was fixated first or second differed ($z = -5.06$, $p < 0.001$, 95% CI [0.39, 0.45], $h = 0.16$), whereas the number of times it was fixated third or fourth did not differ ($z = 2.18$, $p = 0.17$, 95% CI [0.50, 0.57], $h = 0.15$).

Similarly, in face morphs, the expected ROI was fixated earlier than the unexpected ROI ($z = 4.98$, $p < 0.001$, 95% CI [0.56, 0.64], $h = 0.20$) as well as the other ROIs containing no distinct feature ($z = 5.54$, $p < 0.001$, 95% CI [0.39, 0.44], $h = 0.17$). The unexpected ROI was also fixated earlier than the other ROIs ($z = 3.37$, $p < 0.001$, 95% CI [0.34, 0.37], $h = 0.08$, Figure 4B), indicating that, after initial guidance by expectations, bottom-up deviations influenced subsequent fixations.

In *mismatch* trials, the expectation guidance of initial fixations was partially counterbalanced by bottom-up information: While the expected ROI was still fixated earlier than the other two ROIs ($z = 4.37$, $p < 0.001$, 95% CI [0.36, 0.41], $h = 0.11$), it was not fixated earlier than the unexpected ROI ($z = -1.71$, $p = 0.09$, 95% CI [0.44, 0.50], $h = 0.14$). Rather, the unexpected ROI was also fixated earlier than the other two ROIs ($z = 5.33$, $p < 0.001$, 95% CI [0.37, 0.41], $h = 0.12$), suggesting that, if the sensory information distinctly differed from the expectation, when not a face morph but a completely unexpected face was presented, top-down as well as bottom-up information guided eye movements to informative locations.

The second aim of Experiment 2 was to investigate whether expectations influence how often and how long information is sampled from the expected and unexpected facial features. There were differences in the number of fixations and dwell time in line with a congruency effect; i.e., expected facial features in face morphs were fixated more often and longer than unexpected and other features (number of fixations: expected: $M = 26.58\%$, $SD = 3.91\%$; unexpected: $M = 21.99\%$, $SD = 2.78\%$; others: $M = 15.34\%$, $SD = 2.09\%$; expected vs. unexpected: $V = 582$, $p < 0.001$, 95% CI [2.82, 5.74], Wilcoxon's $r = 0.83$; expected vs. others: $V = 595$, $p < 0.001$, 95% CI [9.53, 12.54], Wilcoxon's

A Expectation Effect on Saccades

B First and Second Saccades During ISI

Figure 3. Eye-tracking results of Experiment 1

(A) Expectation-induced predictive saccades: participants performed saccades more often toward a region of interest (ROI) if the corresponding facial feature had been expected compared to when it had not been expected (main effect "expectation"). Each dot displays, for each participant, the a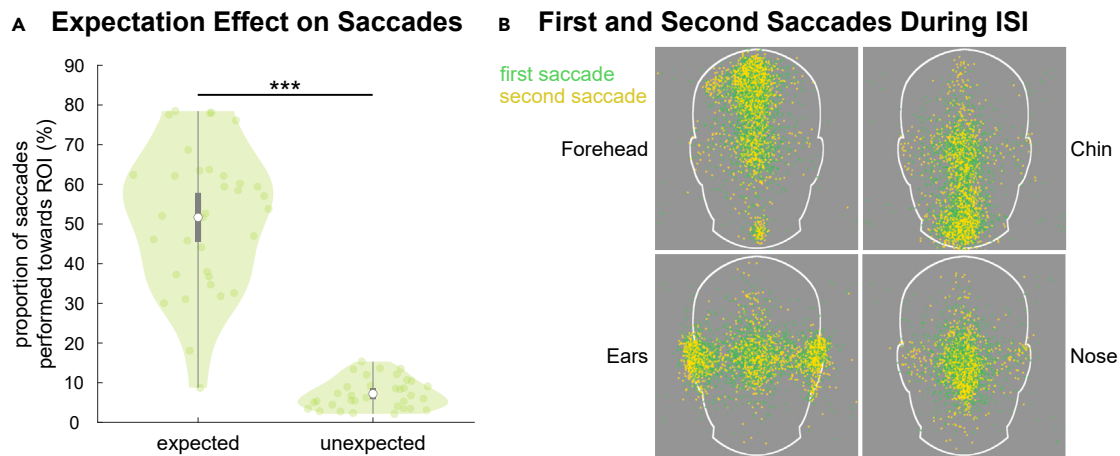verage percentage of fixations toward an ROI if its facial feature has been expected or not (averaged across the four ROIs). The white dots represent means, gray rectangles 95% confidence intervals, and the lower and upper whiskers $Q_{1/3}$ −/+ 1.5*interquartile range. The black line indicates $p < 0.001$.

(B) Visualization of predictive saccades: predictive saccades clustered depending on where the distinct facial feature was expected (forehead, chin, ears, and nose). Single dots represent the endpoints of all participants' first and second saccades during the presentation of the face outline in the inter-stimulus interval (ISI) between the name prior and a presented face. Green, first saccade; yellow, second saccade.

$r = 0.87$; dwell time: expected: $M = 22.86\%$, $SD = 3.73\%$; unexpected: $M = 18.84\%$, $SD = 2.23\%$; others: $M = 12.04\%$, $SD = 1.91\%$; expected vs. unexpected: $V = 559$, $p < 0.001$, 95% CI [2.29, 5.38], Wilcoxon's $r = 0.77$; expected vs. others: $V = 595$, $p < 0.001$, 95% CI [9.19, 12.20], Wilcoxon's $r = 0.87$).

Further exploratory analyses of fixation time courses during face presentation revealed an interaction effect of expectation and time on the number of fixations and dwell times, with an initial preferred sampling of the expected feature in the first 1,000 ms of face presentation, which reversed to an increased sampling of the unexpected feature in the later time window of 1,500–2,000 ms (Figure S1). Participants differentially fixated the expected, unexpected, and other ROIs (main effect "ROI": number of fixations: $F(1.92, 63.20) = 138.94$, $p < 0.001$, $\eta_p^2 = 0.81$; dwell time: $F(1.88, 62.13) = 132.56$, $p < 0.001$, $\eta_p^2 = 0.80$). In addition, there was a difference in the number of fixations toward any of the four ROIs over time (main effect "bin": $F(1.70, 55.98) = 9.80$, $p < 0.001$, $\eta_p^2 = 0.23$), although this was not evident in dwell times ($F(1.85, 60.99) = 1.94$, $p = 0.16$, $\eta_p^2 = 0.06$). For both, i.e., number of fixations and dwell time, there was an interaction between "ROI" and "bin" (number of fixations: $F(3.49, 115.09) = 26.07$, $p < 0.001 = 0.18$, $\eta_p^2 = 0.44$; dwell time: $F(3.53, 116.35) = 23.26$, $p < 0.001$, $\eta_p^2 = 0.41$), showing a preferred sampling of the expected feature in the first two bins, which reversed into an increased sampling of the unexpected feature in the fourth bin (number of fixations: expected vs. unexpected: first bin: $t(33) = 7.94$, $p < 0.001$, 95% CI [5.97, 15.02], $d = 1.36$; second bin: $t(33) = 5.47$, $p < 0.001$, 95% CI [3.26, 14.20], $d = 0.94$; fourth bin: $t(33) = -4.00$, $p = 0.01$, 95% CI [−10.28, −0.80], $d = -0.69$; dwell time: expected vs. unexpected: first bin: $t(33) = 6.77$, $p < 0.001$, 95% CI [3.89, 11.85], $d = 1.16$; second bin: $t(33) = 4.11$, $p = 0.008$, 95% CI [1.09, 12.03], $d = 0.70$; fourth bin: $t(33) = -4.08$, $p = 0.008$, 95% CI [−9.04, −0.79], $d = -0.70$).

Moreover, an exploratory analysis of fixation durations showed that fixations on the expected and unexpected ROIs were longer compared to the other ROIs, reflecting an increased sampling of informative features in face morphs (Figure S2). There were differences in the fixation duration depending on expectation (main effect "ROI": $F(2, 66) = 44.63$, $p < 0.001$, $\eta^2 = 0.122$), with longer fixations on features distinctive for the two identities within a face morph compared to the other ROIs (expected: $M = 239.58$ ms, $SD = 34.21$ ms; unexpected: $M = 244.88$ ms, $SD = 37.78$ ms; others: $M = 215.68$ ms, $SD = 31.67$ ms; expected vs. others: $t(33) = 7.15$, $p < 0.001$, 95% CI [15.50, 32.32], $d = 1.23$; unexpected vs. others: $t(33) = 9.09$, $p < 0.001$, 95% CI [21.10, 37.30], $d = 1.56$). There was no difference in the fixation duration between the expected and unexpected ROI ($t(33) = -1.59$, $p = 0.36$, 95% CI [−13.70, 3.09], $d = -0.27$).

Linking eye movements to behavior

Finally, we investigated the relationship between eye movements and participants' responses. In Experiment 1, the accuracy in identifying a clear face as expected or unexpected was higher if participants fixated on the expected facial feature at face onset, in the *match* ($M = 91.40\%$, $SD = 9.22\%$ vs. $M = 76.40\%$, $SD = 12.40\%$; $V = 584$, $p < 0.001$, 95% CI [10.82, Inf], Wilcoxon's $r = 0.84$) and in the *mismatch* condition ($M = 87.40\%$, $SD = 11.90\%$ vs. $M = 76.80\%$, $SD = 14.90\%$; $V = 529$, $p < 0.001$, [5.90, Inf], Wilcoxon's $r = 0.68$). Interestingly, in *partial* trials containing face morphs, participants were more likely than chance level to indicate the expected face identity if they fixated on the expected ROI at face onset ($M = 89.73\%$, $SD = 14.25\%$; $V = 592$, $p < 0.001$, [87.31, Inf], Wilcoxon's $r = 0.86$), possibly linking the assimilation effect to the gathering of expectation-compliant information at the expected face location.
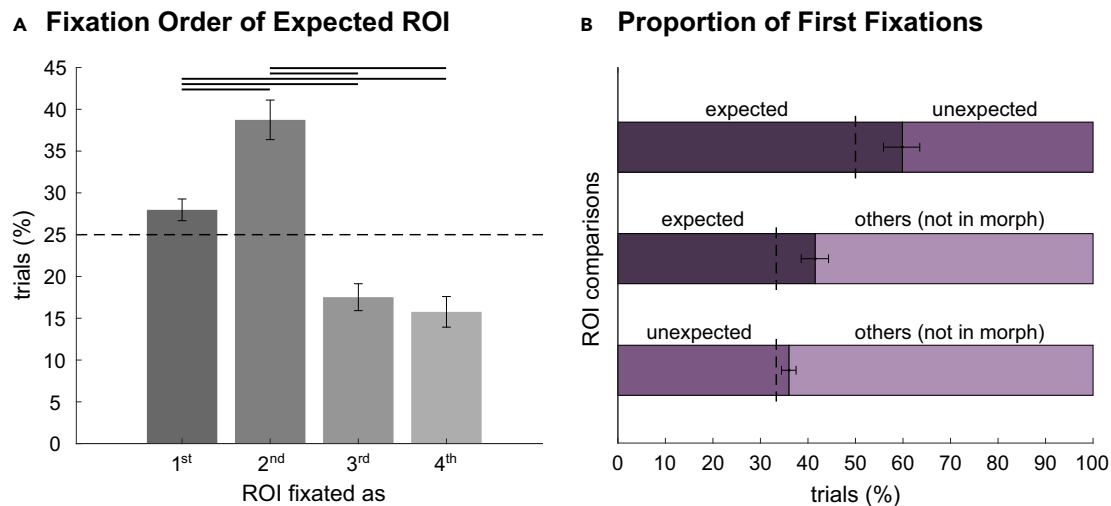
A Fixation Order of Expected ROI

B Proportion of First Fixations

Figure 4. Results of Experiment 2

(A) Early fixation of the expected region of interest (ROI): in all trials (i.e., *match*, *mismatch*, *partial*), participants fixated the expected ROI more often as first or second out of all four ROIs. Error bars indicate 95% CIs. The dashed line represents the chance level (π = 0.25). Black lines indicate $p < 0.001$.

(B) Fixation order in partial trials: in face morphs, participants fixated the ROI containing the expected feature earlier than the ROI containing the feature of the unexpected identity and the other two ROIs. They also fixated on the unexpected feature earlier than on the other two ROIs. Error bars indicate 95% CIs. The dashed lines represent the chance level (π = 0.50 and π = 0.33, respectively). See also Figure S3.

In Experiment 2, there was a last-sampling bias; i.e., participants perceived the identity in a face morph more often than chance if they fixated its distinct ROI last (z = 7.09, $p < 0.001$, 95% CI [0.36, 0.45], h = 0.35).

DISCUSSION

In two preregistered eye-tracking studies, we investigated the influence of expectations, induced by name cues, on gazing behavior when viewing faces. We created ambiguous sensory information by morphing two faces containing expected as well as unexpected facial features. In the first experiment, participants performed predictive saccades toward expected facial features. In the second experiment, expectations guided fixations toward expected facial features in face morphs, which were reversed toward unexpected features over time. In both experiments, participants were faster in recognizing an expected presented face compared to unexpected or morphed ones. Furthermore, the name prior shifted the identification of the ambiguous face morphs toward the expected identity. Overall, our results show that context shapes information sampling during face recognition, particularly during the early examination of a face. Guiding the eyes toward locations with expected facial features may help to evaluate whether the provided sensory information matches the prediction. We thereby support the established view that perception is an active process in which bottom-up sensory information and top-down expectations are combined.[7–9]

In our first experiment, as hypothesized, participants performed predictive saccades toward expected facial features, effectively translating context information into anticipatory eye movements. While previous literature on eye movements during face perception showed that humans tend to initially fixate on a point slightly below the eyes[23,24,36,37] and preferentially fixate on the eyes and mouth,[27,28] our findings show that higher-order contextual effects can revoke this automatism during anticipation. Expectations guide fixations toward facial features that are more informative for face identification, as has been shown for familiar faces compared to unfamiliar ones.[33] An important determinant of whether top-down information can predictively guide eye movements could be task relevance.[38,39] In our experiment, predictive saccades toward expected features allowed the sampling of relevant information to perform the identification task, further highlighting the possibility of task-relevant expectations to modify typical viewing behavior of faces. Our findings of expectation-dependent eye movements have wider implications for our understanding of face perception within the framework of "predictive perception" and active inference.[10] According to this view of the human brain as a pro-active Bayesian hypothesis tester, gaze control provides a strategy to actively sample evidence at locations where important and relevant information is anticipated.[40,41] Our findings support the theoretical assumption that action, here in the form of saccadic eye movements, tests perceptual hypotheses.[10] This active visual sampling may partly explain why the perception and early neural processing of face identity and facial expression are context dependent.[1] Translating this effect to face recognition in everyday situations, this confirmatory gaze strategy would lead to faster recognition of expected faces, while potentially overlooking additional deviating information within a face, as indicated by the assimilation effect for face morphs in our study—unless deviating sensory information leads to a rejection of the perceptual hypothesis as in the current *mismatch* trials.

In our second experiment, we found an order effect of expectation; i.e., expected facial features were fixated early in all trials, as well as earlier than the unexpected facial feature in *partial* trials. This is in line with the results of our first experiment, showing an early guidance of eye movements by expectations. The first fixations on a face are especially important to determine how the face is identified.[23,42] Similarly, there

were reduced latencies to the first target fixation when an object's semantic surroundings have been primed[43] and when an object semantically fitted to its surroundings (e.g., a pot compared to a printer in a kitchen[14]), supporting the idea of an active sampling of expected information.[10] However, there have also been several contradictory reports of an earlier fixation on semantically incongruent information.[21,44–46] Even when controlling for low-level stimulus properties, scene-incongruent information was recognized faster and more accurately, which has been taken as evidence for the preferred processing of unexpected information.[47] In *partial* trials, we observed earlier fixations of expected compared to unexpected features as well as of unexpected compared to the other two features. Correspondingly in *mismatch* trials, participants fixated on the unexpected facial feature earlier than on the other two ROIs, with no difference between the fixation order of the expected and the unexpected ROI. This implies that, if sensory information sufficiently differs from expectations, it can also attract eye movements to unexpected locations. We speculate that the weighted interplay of both bottom-up deviations and top-down context-driven goals may provide a useful strategy during visual sampling while less important deviations in the environment are investigated after processing of the expected information.

In line with the expectation effect on fixation order, we found a congruency effect on fixation number and dwell time; i.e., participants fixated on the expected facial feature more often and longer than on the unexpected facial feature in a face morph. An active sampling of expected information is in line with previous literature reporting more fixations on semantically congruent scenes[20] but contradicts other experiments demonstrating longer dwell times on scene-incongruent information.[21,22] Further exploratory analyses of fixation time courses in 500 ms bins revealed that this preference for the expected facial feature was especially evident during the first 1,000 ms, followed by increased fixations of the unexpected features during 1,500–2,000 ms.

On the behavioral level, we found the typical facilitation effect with faster RTs for expected compared to unexpected clear or morphed faces.[4–7] Furthermore, there was an assimilation effect; i.e., participants indicated the expected identity more often in a face morph, as anticipated for the short cue duration,[7] rather than contrastive after-effects in adaptation designs (for review, see Mueller et al.[48]).

Participants showed higher accuracy for identifying a presented face if they fixated on the expected feature at face onset (Experiment 1) and a last-sampling bias[49]; i.e., they chose an identity more often than chance if they fixated on its distinct ROI last (Experiment 2). These links between fixations and behavioral responses suggest that the perceptual decision about the identity of a face morph was reached by an information sampling strategy, in which first the expected informative features were tested and sampling at other locations continued until enough confirmatory information for the decision was obtained.[10]

Future research is needed to evaluate the ecological validity of expectations guiding viewing behavior in face identification. This could be realized in face-to-face experimental settings and mobile tracking of eye movements.[50,51] Furthermore, viewing patterns of both super-recognizers, i.e., individuals who are exceptionally skilled in face identification,[52,53] as well as developmental and acquired prosopagnosia, i.e., individuals who have difficulty recognizing familiar faces,[54,55] bear the potential to gain more insights into human face recognition. Superior performance might be related to more efficient processing of informative facial features to identify faces, whereas patients with prosopagnosia may fixate on less informative facial regions.[54,56,57] Our findings may also have broader implications for understanding atypical face-viewing behavior, such as that observed in individuals with autism spectrum disorders.[58–60] Understanding the effect of different viewing strategies might provide the potential to train and improve face recognition skills by focusing on expected informative facial features and help to understand why observers deploy idiosyncratic strategies.[61,62]

Overall, we were able to show that context-induced expectations guide predictive saccades toward and an early sampling of expected features during face recognition. Hence, expectations can influence the way we look at expected faces and direct the extraction of visual information from them. In face morphs, after sampling of expected information, unexpected features captured eye movements, suggesting that bottom-up information is additionally considered during the perception of visually ambiguous sensory information. Our results offer compelling evidence that expectations shape sampling of visual information, contributing empirical evidence to the influential theoretical framework of predictive processing.

Limitations of the study

Finally, some limitations of our experimental design should be acknowledged. Firstly, in both experiments, the name cue was task relevant as participants had to indicate whether a face was expected or unexpected. It will be interesting to test whether our observed expectation-dependent gaze patterns are also present in implicit task settings. Secondly, in Experiment 2, participants might have responded "expected" more frequently to shorten the duration of the experiment, possibly contributing to the behavioral assimilation effect. However, this should not affect the initial guidance of fixations. Thirdly, the manipulated facial features differed in saliency and could have affected gazing behavior in a purely bottom-up fashion during the second experiment. While this holds for the *match* trials, it does not explain the initial guidance to expected features in *partial* trials in which faces contained salient information in two regions. In *mismatch* trials, we observed preferred initial fixation of the expected and unexpected ROIs compared to the two regions of no interest, hinting toward a combination of bottom-up- and top-down-driven viewing behavior. Fourthly, the order analysis showed that overall the expected ROI was more often fixated second than first. This was due to first fixations landing predominantly in the center of the face, in line with a central viewing tendency[24,63] (Figure S3). Nevertheless, expectations clearly guided the second fixation to the other three ROIs in contrast to later (i.e., third and fourth) fixations (Figure S3). Lastly, sample sizes were preregistered and based on power analyses to reach a power of at least 0.80 to find effects of medium size at an alpha error probability of 5%. Studies powered to detect small effect sizes may reveal whether the currently observed group differences are robust. We did not perform analyses split up by gender because we investigated a general perceptual effect and our experimental samples were not powered for gender-based analyses.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Helen Blank (h.blank@uke.de).

Materials availability

The face images and auditory feedback stimuli generated in this study are available via the Open Science Framework (OSF) (Experiment 1: https://osf.io/7e38v/; Experiment 2: https://osf.io/tbdh6/).

Data and code availability

- All data have been deposited at the OSF and are publicly available as of the date of publication. DOIs are listed in the key resources table.
- All original code has been deposited at the OSF and is publicly available as of the date of publication. DOIs are listed in the key resources table.
- No additional information is required to reanalyze the data reported in this paper.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

A.G.: conceptualization, methodology, formal analysis, data curation, writing – original draft, visualization, and project administration. M.L.: conceptualization, methodology, validation, investigation, and writing – review and editing. M.G.: conceptualization, methodology, resources, and writing – review and editing. H.B.: conceptualization, methodology, writing – original draft, supervision, project administration, and funding acquisition.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Sample size
- METHOD DETAILS
  - Apparatus and stimuli
  - Training sessions
  - Procedure
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Data exclusion and sample size estimation
  - Behavioral analyses
  - Eye-tracking analyses
  - Number of fixations and dwell time analysis by bin
  - Fixation durations
  - Combined behavioral and eye-tracking analyses
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.110920.

REFERENCES

1. Hadders-Algra, M. (2022). Human face and gaze perception is highly context specific and involves bottom-up and top-down neural processing. Neurosci. Biobehav. Rev. *132*, 304–323. https://doi.org/10.1016/j.neubiorev.2021.11.042.

2. Webster, M.A., and MacLeod, D.I.A. (2011). Visual adaptation and face perception. Philos. Trans. R. Soc. Lond. B Biol. Sci. *366*, 1702–1725. https://doi.org/10.1098/rstb.2010.0360.

3. Blank, H., Wieland, N., and von Kriegstein, K. (2014). Person recognition and the brain: Merging evidence from patients and healthy individuals. Neurosci. Biobehav. Rev. *47*, 717–734. https://doi.org/10.1016/j.neubiorev.2014.10.022.

4. Laurence, S., Baker, K.A., Proietti, V.M., and Mondloch, C.J. (2022). What happens to our representation of identity as familiar faces age? Evidence from priming and identity aftereffects. Br. J. Psychol. *113*, 677–695. https://doi.org/10.1111/bjop.12560.

5. Walther, C., Schweinberger, S.R., and Kovács, G. (2014). Decision-dependent aftereffects

for faces. Vis. Res. *100*, 47–55. https://doi.org/10.1016/j.visres.2014.04.005.

6. Blank, H., Kiebel, S.J., and von Kriegstein, K. (2015). How the human brain exchanges information across sensory modalities to recognize other people. Hum. Brain Mapp. *36*, 324–339. https://doi.org/10.1002/hbm.22631.

7. Garlichs, A., and Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. Nat. Commun. *15*, 3407. https://doi.org/10.1038/s41467-024-47749-9.

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

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

10. Friston, K., Adams, R.A., Perrinet, L., and Breakspear, M. (2012). Perceptions as Hypotheses: Saccades as Experiments. Front. Psychol. *3*, 151. https://doi.org/10.3389/fpsyg.2012.00151.

11. Hwang, A.D., Wang, H.-C., and Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. Vis. Res. *51*, 1192–1205. https://doi.org/10.1016/j.visres.2011.03.010.

12. Malcolm, G.L., and Henderson, J.M. (2010). Combining top-down processes to guide eye movements during real-world scene search. J. Vis. *10*, 1–11. https://doi.org/10.1167/10.2.4.

13. Spotorno, S., and Tatler, B.W. (2017). The elephant in the room: Inconsistency in scene viewing and representation. J. Exp. Psychol. Hum. Percept. Perform. *43*, 1717–1743. https://doi.org/10.1037/xhp0000456.

14. Võ, M.L.-H., and Henderson, J.M. (2011). Object–scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. Atten. Percept. Psychophys. *73*, 1742–1753. https://doi.org/10.3758/s13414-011-0150-6.

15. Wolfe, J.M. (2020). Visual Search: How Do We Find What We Are Looking For? Annu Rev Vis Sci. *6*, 539–562. https://doi.org/10.1146/annurev-vision-091718-015048.

16. Hayhoe, M.M., McKinney, T., Chajka, K., and Pelz, J.B. (2012). Predictive eye movements in natural vision. Exp. Brain Res. *217*, 125–136. https://doi.org/10.1007/s00221-011-2979-2.

17. Land, M.F., and McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. Nat. Neurosci. *3*, 1340–1345. https://doi.org/10.1038/81887.

18. Mann, D.L., Nakamoto, H., Logt, N., Sikkink, L., and Brenner, E. (2019). Predictive eye movements when hitting a bouncing ball. J. Vis. *19*, 28. https://doi.org/10.1167/19.14.28.

19. Ghiani, A., Mann, D., and Brenner, E. (2024). Methods matter: Exploring how expectations influence common actions. iScience *27*, 109076. https://doi.org/10.1016/j.isci.2024.109076.

20. Bornstein, M.H., Mash, C., and Arterberry, M.E. (2011). Perception of object–context relations: Eye-movement analyses in infants and adults. Dev. Psychol. *47*, 364–375. https://doi.org/10.1037/a0021059.

21. Coco, M.I., Nuthmann, A., and Dimigen, O. (2020). Fixation-related Brain Potentials during Semantic Integration of Object–Scene Information. J. Cognit. Neurosci. *32*, 571–589. https://doi.org/10.1162/jocn_a_01504.

22. Cornelissen, T.H.W., and Võ, M.L.-H. (2017). Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior. Atten. Percept. Psychophys. *79*, 154–168. https://doi.org/10.3758/s13414-016-1203-7.

23. Hsiao, J.H.w., and Cottrell, G. (2008). Two Fixations Suffice in Face Recognition. Psychol. Sci. *19*, 998–1006. https://doi.org/10.1111/j.1467-9280.2008.02191.x.

24. Peterson, M.F., and Eckstein, M.P. (2012). Looking just below the eyes is optimal across face recognition tasks. Proc. Natl. Acad. Sci. USA *109*, E3314–E3323. https://doi.org/10.1073/pnas.1214269109.

25. Rogers, S.L., Speelman, C.P., Guidetti, O., and Longmuir, M. (2018). Using dual eye tracking to uncover personal gaze patterns during social interaction. Sci. Rep. *8*, 4271. https://doi.org/10.1038/s41598-018-22726-7.

26. Wegner-Clemens, K., Rennig, J., Magnotti, J.F., and Beauchamp, M.S. (2019). Using principal component analysis to characterize eye movement fixation patterns during face viewing. J. Vis. *19*, 2. https://doi.org/10.1167/19.13.2.

27. Scheller, E., Büchel, C., and Gamer, M. (2012). Diagnostic Features of Emotional Expressions Are Processed Preferentially. PLoS One *7*, e41792. https://doi.org/10.1371/journal.pone.0041792.

28. Schurgin, M.W., Nelson, J., Iida, S., Ohira, H., Chiao, J.Y., and Franconeri, S.L. (2014). Eye movements during emotion recognition in faces. J. Vis. *14*, 14. https://doi.org/10.1167/14.13.14.

29. Aviezer, H., Ensenberg, N., and Hassin, R.R. (2017). The inherently contextualized nature of facial emotion perception. Curr. Opin. Psychol. *17*, 47–54. https://doi.org/10.1016/j.copsyc.2017.06.006.

30. Althoff, R.R., and Cohen, N.J. (1999). Eye-movement-based memory effect: A reprocessing effect in face perception. J. Exp. Psychol. Learn. Mem. Cogn. *25*, 997–1010. https://doi.org/10.1037/0278-7393.25.4.997.

31. Barton, J.J.S., Radcliffe, N., Cherkasova, M.V., Edelman, J., and Intriligator, J.M. (2006). Information Processing during Face Recognition: The Effects of Familiarity, Inversion, and Morphing on Scanning Fixations. Perception *35*, 1089–1105. https://doi.org/10.1068/p5547.

32. Heisz, J.J., and Shore, D.I. (2008). More efficient scanning for familiar faces. J. Vis. *8*, 1–10. https://doi.org/10.1167/8.1.9.

33. Van Belle, G., Ramon, M., Lefèvre, P., and Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. Front. Psychol. *1*, 20. https://doi.org/10.3389/fpsyg.2010.00020.

34. Wang, Q., Xiao, N.G., Quinn, P.C., Hu, C.S., Qian, M., Fu, G., and Lee, K. (2015). Visual scanning and recognition of Chinese, Caucasian, and racially ambiguous faces: Contributions from bottom-up facial physiognomic information and top-down knowledge of racial categories. Vis. Res. *107*, 67–75. https://doi.org/10.1016/j.visres.2014.10.032.

35. Aviezer, H., Hassin, R.R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., and Bentin, S. (2008). Angry, Disgusted, or Afraid?: Studies on the Malleability of Emotion Perception. Psychol. Sci. *19*, 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x.

36. Han, N.X., Chakravarthula, P.N., and Eckstein, M.P. (2021). Peripheral facial features guiding eye movements and reducing fixational variability. J. Vis. *21*, 7. https://doi.org/10.1167/jov.21.8.7.

37. Or, C.C.-F., Peterson, M.F., and Eckstein, M.P. (2015). Initial eye movements during face identification are optimal and similar across cultures. J. Vis. *15*, 12. https://doi.org/10.1167/15.13.12.

38. Einhäuser, W., Rutishauser, U., and Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. J. Vis. *8*, 2. https://doi.org/10.1167/8.2.2.

39. Foulsham, T., and Underwood, G. (2007). How Does the Purpose of Inspection Influence the Potency of Visual Salience in Scene Perception? Perception *36*, 1123–1138. https://doi.org/10.1068/p5659.

40. Henderson, J.M. (2017). Gaze Control as Prediction. Trends Cognit. Sci. *21*, 15–23. https://doi.org/10.1016/j.tics.2016.11.003.

41. Otten, M., Seth, A.K., and Pinto, Y. (2017). A social Bayesian brain: How social knowledge can shape visual perception. Brain Cognit. *112*, 69–77. https://doi.org/10.1016/j.bandc.2016.05.002.

42. Miellet, S., Caldara, R., and Schyns, P.G. (2011). Local Jekyll and Global Hyde: The Dual Identity of Face Identification. Psychol. Sci. *22*, 1518–1526. https://doi.org/10.1177/0956797611424290.

43. Võ, M.L.-H., and Henderson, J.M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. J. Vis. *10*, 14. https://doi.org/10.1167/10.3.14.

44. Bonitz, V.S., and Gordon, R.D. (2008). Attention to smoking-related and incongruous objects during scene viewing. Acta Psychol. *129*, 255–263. https://doi.org/10.1016/j.actpsy.2008.08.006.

45. Brockmole, J.R., and Henderson, J.M. (2008). Prioritizing new objects for eye fixation in real-world scenes: Effects of object–scene consistency. Vis. Cognit. *16*, 375–390. https://doi.org/10.1080/13506280701453623.

46. Underwood, G., Humphreys, L., and Cross, E. (2007). Chapter 26 - Congruency, saliency and gist in the inspection of objects in natural scenes. In Eye Movements, R.P.G. Van Gompel, M.H. Fischer, W.S. Murray, and R.L. Hill, eds. (Elsevier), pp. 563–VII. https://doi.org/10.1016/B978-008044980-7/50028-8.

47. Spaak, E., Peelen, M.V., and de Lange, F.P. (2022). Scene Context Impairs Perception of Semantically Congruent Objects. Psychol. Sci. *33*, 299–313. https://doi.org/10.1177/09567976211032676.

48. Mueller, R., Utz, S., Carbon, C.-C., and Strobach, T. (2020). Face Adaptation and Face Priming as Tools for Getting Insights Into the Quality of Face Space. Front. Psychol. *11*, 166. https://doi.org/10.3389/fpsyg.2020.00166.

49. Zhu, T. (2022). Accounting for the last-sampling bias in perceptual decision-making. Cognition *223*, 105049. https://doi.org/10.1016/j.cognition.2022.105049.

50. Varela, V.P.L., Towler, A., Kemp, R.I., and White, D. (2023). Looking at faces in the wild. Sci. Rep. *13*, 783. https://doi.org/10.1038/s41598-022-25268-1.

51. Vehlen, A., Spenthof, I., Tönsing, D., Heinrichs, M., and Domes, G. (2021). Evaluation of an eye tracking setup for studying visual attention in face-to-face conversations. Sci. Rep. *11*, 2661. https://doi.org/10.1038/s41598-021-81987-x.

52. Dunn, J.D., Varela, V.P.L., Nicholls, V.I., Papinutto, M., White, D., and Miellet, S. (2022). Face-Information Sampling in Super-Recognizers. Psychol. Sci. 33, 1615–1630. https://doi.org/10.1177/09567976221096320.

53. Linka, M., Broda, M.D., Alsheimer, T., de Haas, B., and Ramon, M. (2022). Characteristic fixation biases in Super-Recognizers. J. Vis. 22, 17. https://doi.org/10.1167/jov.22.8.17.

54. Bobak, A.K., Parris, B.A., Gregory, N.J., Bennetts, R.J., and Bate, S. (2017). Eye-movement strategies in developmental prosopagnosia and "super" face recognition. Q. J. Exp. Psychol. 70, 201–217. https://doi.org/10.1080/17470218.2016.1161059.

55. Orban de Xivry, J.J., Ramon, M., Lefèvre, P., and Rossion, B. (2008). Reduced fixation on the upper area of personally familiar faces following acquired prosopagnosia. J. Neuropsychol. 2, 245–268. https://doi.org/10.1348/174866407X260199.

56. Lee, D.-H., Corrow, S.L., Pancaroglu, R., and Barton, J.J.S. (2019). The Scanpaths of Subjects with Developmental Prosopagnosia during a Face Memory Task. Brain Sci. 9, 188. https://doi.org/10.3390/brainsci9080188.

57. Schwarzer, G., Huber, S., Gruter, M., Gruter, T., Groß, C., Hipfel, M., and Kennerknecht, I. (2007). Gaze behaviour in hereditary prosopagnosia. Psychol. Res. 71, 583–590. https://doi.org/10.1007/s00426-006-0068-0.

58. Cuve, H.C., Castiello, S., Shiferaw, B., Ichijo, E., Catmur, C., and Bird, G. (2021). Alexithymia explains atypical spatiotemporal dynamics of eye gaze in autism. Cognition 212, 104710. https://doi.org/10.1016/j.cognition.2021.104710.

59. Palmer, C.J., Lawson, R.P., and Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. Psychol. Bull. 143, 521–542. https://doi.org/10.1037/bul0000097.

60. Wilson, C.E., Palermo, R., and Brock, J. (2012). Visual Scan Paths and Recognition of Facial Identity in Autism Spectrum Disorder and Typical Development. PLoS One 7, e37681. https://doi.org/10.1371/journal.pone.0037681.

61. Arizpe, J., Walsh, V., Yovel, G., and Baker, C.I. (2017). The categories, frequencies, and stability of idiosyncratic eye-movement patterns to faces. Vis. Res. 141, 191–203. https://doi.org/10.1016/j.visres.2016.10.013.

62. Mehoudar, E., Arizpe, J., Baker, C.I., and Yovel, G. (2014). Faces in the eye of the beholder: Unique and stable eye scanning patterns of individual observers. J. Vis. 14, 6. https://doi.org/10.1167/14.7.6.

63. Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. Vis. Res. 50, 2577–2587. https://doi.org/10.1016/j.visres.2010.08.016.

64. Schwind, V., Wolf, K., and Henze, N. (2017). FaceMaker—A Procedural Face Generator to Foster Character Design Research. In Game Dynamics: Best Practices in Procedural and Dynamic Game Content Generation, O. Korn and N. Lee, eds. (Springer International Publishing), pp. 95–113. https://doi.org/10.1007/978-3-319-53088-8_6.

65. Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behav. Res. Methods 41, 1149–1160. https://doi.org/10.3758/BRM.41.4.1149.

66. Rösler, L., Göhring, S., Strunz, M., and Gamer, M. (2021 Mar). Social anxiety is associated with heart rate but not gaze behavior in a real social interaction. J. Behav. Ther. Exp. Psychiatry 70, 101600. https://doi.org/10.1016/j.jbtep.2020.101600.

67. Rösler, L., and Gamer, M. (2019). Freezing of gaze during action preparation under threat imminence. Sci. Rep. 9, 17215. https://doi.org/10.1038/s41598-019-53683-4.

68. Boll, S., Bartholomaeus, M., Peter, U., Lupke, U., and Gamer, M. (2016). Attentional mechanisms of social perception are biased in social phobia. J. Anxiety Disord. 40, 83–93. https://doi.org/10.1016/j.janxdis.2016.04.004.

69. Gamer, M., Schmitz, A.K., Tittgemeyer, M., and Schilbach, L. (2013). The human amygdala drives reflexive orienting towards facial features. Curr. Biol. 23, R917–R918. https://doi.org/10.1016/j.cub.2013.09.008.

70. Lancry-Dayan, O.C., Gamer, M., and Pertzov, Y. (2021). Search for the Unknown: Guidance of Visual Search in the Absence of an Active Template. Psychol. Sci. 32, 1404–1415. https://doi.org/10.1177/0956797621996660.

71. Võ, M.L.H., and Henderson, J.M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. J. Vis. 9, 24. https://doi.org/10.1167/9.3.24.

72. Gregg, M., Datta, S., and Lorenz, D. (2022). htestClust: A Package for Marginal Inference of Clustered Data Under Informative Cluster Size. Rom. Jahrb. 14, 54–66. https://doi.org/10.32614/RJ-2022-024.

73. Willroth, E.C., and Atherton, O.E. (2024). Best Laid Plans: A Guide to Reporting Preregistration Deviations. Adv. Methods Pract. Psychol. Sci. 7. https://doi.org/10.1177/25152459231213802.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Experimental data (Experiment 1)	own	https://doi.org/10.17605/OSF.IO/7E38V
Stimuli (Experiment 1)	own	https://doi.org/10.17605/OSF.IO/7E38V
Material for experiments (Experiment 1)	own	https://doi.org/10.17605/OSF.IO/7E38V
Code for analyses (Experiment 1)	own	https://doi.org/10.17605/OSF.IO/7E38V
Code for experiments (Experiment 1)	own	https://doi.org/10.17605/OSF.IO/7E38V
Experimental data (Experiment 2)	own	https://doi.org/10.17605/OSF.IO/TBDH6
Stimuli (Experiment 2)	own	https://doi.org/10.17605/OSF.IO/TBDH6
Material for experiments (Experiment 2)	own	https://doi.org/10.17605/OSF.IO/TBDH6
Code for analyses (Experiment 2)	own	https://doi.org/10.17605/OSF.IO/TBDH6
Code for experiments (Experiment 2)	own	https://doi.org/10.17605/OSF.IO/TBDH6
Software and algorithms		
R (v4.2.0)	R Core Team	https://www.r-project.org
RStudio (v2022.02.2)	RStudio Team	https://posit.co
MATLAB (vR2020b)	Mathworks	https://de.mathworks.com
faceMaker	Schwind et al.[63]	http://facemaker.uvrg.org/
Audacity (v3.0.0)	Audacity Team	http://audacity.sourceforge.net/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Sample size

For Experiment 1, we report the results of 34 participants (17 females, self-reported gender) with a mean age of 26.5 years (SD = 4.38 years) (see quantification and statistical analysis). For Experiment 2, 34 participants (18 females) with a mean age of 25.09 years (SD = 4.91 years) were included. Participants had no history of neurological or psychiatric disorders. All experimental procedures were approved by the Ethics Committee of the Chamber of Physicians in Hamburg and participants provided written informed consent. We did not perform analyses split up by gender because we investigated a general perceptual effect and our experimental samples were not powered for gender-based analyses.

METHOD DETAILS

Apparatus and stimuli

Stimuli were presented on a Samsung SyncMaster 204B display (41.0 × 31.0 cm; 20.1″) with a resolution of 1600 × 1200 pixels and a refresh rate of 75 Hz. Eye movements were recorded from the participant's right eye using an EyeLink 1000 system at a sampling rate of 1000 Hz. The head location of the participant was fixed using a chin rest and forehead bar. Saccades were defined as periods in which the velocity exceeded 30°/sec or the acceleration 8000°/sec^2, respectively. The saccadic motion threshold was set to 0.1°. A 13-point-calibration was performed at the start of each experimental block in both experiments. Validation was repeated until the result was at least "good" according to the guidelines of the manufacturer (i.e., worst point error <1.5°, average error <1.0°). In case of calibration issues (e.g., due to dense eyelashes), a validated 9- or 5-point-calibration was used.

We used grey-scale images of four male faces created with FaceMaker (http://facemaker.uvrg.org/)[64] (Figure 1A). The four faces differed in one distinct feature (forehead, chin, ears, or nose) from an average 'base face'. For the 'base face', the parameters 'eyebrowsColor', 'hairColor', 'faceGender', and 'skinColor' were set to 1. The parameters of interest for creating different identities were set to: 'foreheadHeight' (0.7), 'jawChin' (0.575), 'jawLength' (0.55), 'earSize' (0.7), and 'noseWidth' (0.625). For the first identity with a high forehead, 'foreheadHeight' was set to 0.9 while keeping the other parameters constant. For the second identity with a wide chin, 'jawChin' was set to 0.65 and 'jawLength' to 0.6. For the third identity with large ears, 'earSize' was set to 0.9. For the fourth identity with a large nose, 'noseWidth' was set to 0.75. Six pairwise face morphs (50/50%) between all faces were created by adjusting the parameters of interest to the mean values between the base face and the two respective identities. The size of the images was 1100 × 1100 pixels, with the face covering approximately 1064 × 736 pixels of the screen. Distance from the eyes to the screen was ~560 mm, leading to a visual angle of 27.44 × 19.17° for the face stimuli.

For auditory feedback, we used three different tones to indicate 'correct', 'incorrect', or 'too slow' responses. All tones were generated with Audacity 3.0.0 (http://audacity.sourceforge.net/) with a duration of 200 ms and an amplitude of 0.7. The 'correct' tone was a sinus wave (880 Hz), and the 'incorrect' and 'too slow' tones were square waves (no aliasing) with 440 Hz and 220 Hz, respectively. The loudness of the tones was normalised.

Training sessions

In the training sessions of both experiments, no eye-tracking data were collected. The head location of the participant was fixed using a chin rest and forehead bar to ensure the same distance to the monitor and visual input as in the main experiment.

The first training was identical for both experiments (~15 min). Participants learned to associate the four distinct faces with their respective names (Ari, Bob, Cid, Dan). The training was divided into three blocks. In each trial of the first block (32 trials), a face was shown for 4500 ms. The task was to identify the face by pressing one out of four buttons with the right hand (index, middle, ring, and pinky finger). Afterward, feedback was provided by presenting the face with a red circle around its distinct feature, a tone indicating whether the response was correct, incorrect, or too slow, as well as written text at the bottom of the screen ('correct', 'incorrect', or 'too slow', and the correct name) for 3000 ms. Each face was shown eight times. In the second block (48 trials), the face presentation duration was shortened to 3500 ms, and visual as well as auditory feedback was provided for 2000 ms. Each face was shown 12 times. In the third block (48 trials), a face was presented for 3500 ms. Only auditory feedback was provided. The experimenter provided feedback about the accuracy score in the last block. If it was below 75%, the last training block was repeated up to two times. If the threshold could still not be surpassed, the participant did not proceed to the next training and the main experiment. In the second training, participants got accustomed to the task of the main experiment. For the training of Experiment 1, 12 trials were shown (~1–2 min). Each name prior was shown three times, once followed by the expected face (match), an unexpected face (mismatch), and a face morph containing the expected identity (partial). For the training of Experiment 2, 24 trials were shown (~2 min). Each prior name was shown six times, twice for each experimental condition (match, mismatch, partial). In between the experimental blocks (three and four blocks in Experiment 1 and 2, respectively), a short repetition of the first training was performed (~2 min) to ensure that participants still had a clear mental representation of the four distinct faces (16 trials, each face four times).

Procedure

In Experiment 1, participants first learned to associate the four faces with their respective names in training. In each trial of the main experiment, a name prior (Ari, Bob, Cid, or Dan) was presented as a cue for the upcoming face in one of the four corners of the screen (750 ms) (Figure 1D). Face-name associations were counterbalanced across participants. In the inter-stimulus interval (ISI), a face outline of the 'base face' (white line) was shown (1000 ms) so that participants could anticipate where the face would appear, followed by a brief presentation of a face (100 ms) and a response window (1500 ms). The presented face was either one of the four learned identities or one of the six pairwise morph combinations (e.g., 50/50% morph between Bob and Cid, Figure 1B). The task was to indicate with one of two buttons whether the presented face was either 'expected' or 'unexpected' based on the prior name. The allocation of the response options to the left and right keys was counterbalanced across participants. Auditory feedback indicated whether the response was correct, incorrect, or too slow. For face morphs, any response was counted as correct. The experiment was divided into three blocks with 144 trials each (~13 min per block). The ratio of trials per condition, i.e., *match* (face matched to the name), *mismatch* (face did not match to the name), and *partial* (face morph containing the expected and an unexpected face identity), was identical in all blocks (48 trials per condition). The trial order within each block was pseudo-randomized, ensuring that the same name cue was restricted to consecutively appear twice at maximum. Each name prior appeared equally often in the upper-left, upper-right, bottom-left, and bottom-right corner. The total duration of the experiment was approximately 90 min.

For Experiment 2, the face was presented longer than in Experiment 1 to allow for visual exploration. Participants started by learning the four identities and their distinct features. In the main experiment, as in Experiment 1, a name was randomly presented in one of the four corners of the screen, followed by the presentation of a clear or morphed face (Figure 1E). The face images were either presented for 4500 ms or until the button press. The first task in each trial was to indicate whether the presented face was 'expected' or 'unexpected' based on the preceding name. If participants answered 'unexpected', a question mark prompted them to respond which identity was perceived in the face. Auditory feedback for the first and second tasks was only provided for too-slow responses. Each trial lasted either maximally 7500 ms (in case of responding 'expected') or 9000 ms (in case of responding 'unexpected'). The experiment was divided into four blocks (~10 min). The ratio of trials per condition (*match*, *mismatch*, *partial*) was identical in all blocks (72 trials per block). In addition to the pseudo-randomization and counterbalancing described for Experiment 1, the question mark was randomly presented in one of the four corners of the screen. The total duration of the experiment was approximately 90 min.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analyses were performed using custom-written scripts in R/RStudio (https://www.r-project.org; https://posit.co) and MATLAB R2020b (https://de.mathworks.com). In case data were not normally distributed (Shapiro-Wilk tests, $p < 0.05$, and visual inspection of Q-Q plots) or outliers were present in the data (above or below $Q_{1/3} -/+ 1.5*IQR$), we calculated non-parametric tests instead of the preregistered parametric tests, except for the ANOVAs due to their robustness to slight violations of the normality assumption and a lack of non-parametric alternatives for two-factorial designs.

Data exclusion and sample size estimation

For Experiment 1, we measured 42 participants. Eight participants were excluded from final data analyses: two participants did not sufficiently learn the four identities and their features in the training (<75% accuracy) and six participants had less than 70% valid eye tracking trials, leading to a final sample of 34 participants. For Experiment 2, we measured 35 participants, one participant was excluded due to behavioral performance below chance level. Sample sizes for both experiments ($N = 34$, each) were preregistered and based on power analyses using G*Power.[65] For Experiment 1, the sample size was optimised to obtain 0.80 power to detect a medium effect size ($f = 0.25$) at 0.05 alpha error probability for testing whether participants perform more predictive saccades to a region of interest (ROI) if its facial feature has been expected compared to when it has not been expected, as the main effect 'expectation' in a 2 × 2 repeated measures ANOVA with the within-subject factors 'expectation' (expected, unexpected) and 'saccade' (first, second). For Experiment 2, we aimed to optimise the power for our main research questions regarding the *partial* trials (face morphs): we tested whether the dwell time on the expected ROI was significantly different from the dwell time on (1) the ROI associated with the other identity contained in a face morph, and on (2) the other ROIs using paired *t*-tests (two-sided) with a power of 0.80 to detect a medium effect size ($dz = 0.50$) at 0.05 alpha error probability.

Data exclusion criteria were preregistered. Firstly, single trials were excluded if participants did not fixate the name prior because awareness of the prior is a prerequisite for the hypotheses about prediction-guided saccades and fixations we aimed to test. Secondly, trials with no response were excluded from data analyses. Thirdly, single trials were excluded if the loss of eye-tracking data, e.g., due to blinks, in our time windows of main interest exceeded a certain threshold. For Experiment 1 ('predictive saccades'), trials were excluded if data loss during the ISI (face outline for 750 ms) exceeded 30%. In Experiment 2 ('information sampling'), trials were excluded if less than 50% valid fixation data were available during face presentation. Face presentation duration was variable in each trial and was determined by a button press of the participant or was 4500 ms at maximum.

Whole datasets (i.e., participants) were excluded in case of too many excluded trials (>30%).[66,67] This was the case for six participants in Experiment 1. Whole datasets were also excluded if the behavioral performance during the main experiment was below the chance level in the match and mismatch conditions. In Experiment 1, chance level accuracy was at 50% (correctly responding 'expected' or 'unexpected'). One participant had an average accuracy score of 25% as well as > 30% invalid trials and was therefore excluded from final data analyses. In Experiment 2, the chance level for mismatch was at 12.5%: participants needed to first correctly indicate that the presented face was 'unexpected' and then correctly identify it was one out of four identities (0.5*0.25 = 0.125). For match, we considered a response correct if participants either answered 'expected', or answered 'unexpected' but then correctly identified the face afterward. We chose the more conservative chance level of 0.5, also for the average match and mismatch scores. One participant in Experiment 2 had a mean accuracy score of 42.71% and was therefore excluded.

In addition to the preregistered exclusion criteria, a learning criterion during the training was important for the completion of the main experiment, making sure that participants learned the four identities and their respective distinct facial features. Participants needed at least 75% accuracy in identifying the four faces in the training to proceed to the main experiment. In Experiment 1, two participants were not able to surpass this cut-off despite two repetitions of the training phase (mean accuracies per participant: 22.92% and 65.97%) and did not participate in the main experiment.

Behavioral analyses

In both experiments, participants were asked to indicate whether a presented face was 'expected' or 'unexpected' given the preceding name. Firstly, a facilitation effect due to expectation, i.e., shorter reaction times (RT) for expected compared to unexpected clear and morphed faces, was tested with repeated-measures ANOVAs with the within-subject factor 'condition' (*match, mismatch, partial*). We report Greenhouse-Geisser corrected results due to sphericity violation as well as the generalised η^2 as a measure of effect size. Pairwise comparisons were based on post-hoc tests (Bonferroni-corrected) with Cohen's *d* as effect size estimate.

Secondly, we tested whether expectations shifted the perception of face morphs, either into the direction of the expected face (assimilation effect) or the unexpected face part (contrastive effect). For each participant, we calculated a perceived face identity score indicating which identity has been perceived in face morphs (*partial* condition) depending on the preceding name prior. For each face morph and name combination (e.g.,$A_{prior}AB_{morph}$), the number of expectation-noncompliant button responses was subtracted from the number of expectation-compliant responses and divided by the number of possible combinations ($N = 12$). This index was converted to percentage values (0–100%): A difference score of zero (no prior effect on morph identification) was converted to 50%, i.e., the participant equally often identified face morphs as the expected or unexpected identity. Values above 50% were indicative of an assimilation effect and values below 50% of a contrastive effect. On the group level, we tested against chance level using a one-sample *t*-test (two-sided).Thirdly, for both experiments, we calculated accuracy scores for the *match* and *mismatch* conditions. On the one hand, we tested each condition's accuracy scores against chance level using one-sample Wilcoxon signed rank tests (one-sided): In Experiment 1, the chance level equaled a probability of 50%. In Experiment 2, the chance level for *mismatch* trials equaled a probability of 12.5% (first task: 'unexpected'; second task: answering the correct ID out of four possibilities → 0.5*0.25 = 0.125). For *match* trials, we considered trials as correct if participants either responded 'expected' or 'unexpected' and then correctly identified the person. We used the more conservative probability of 50% as chance level for the *match* condition. On the other hand, we tested whether there was a difference in accuracy scores between the two conditions using paired Wilcoxon signed rank tests (two-sided).

Eye-tracking analyses

The aim of Experiment 1 was to investigate whether participants use context to perform predictive eye movements if limited sensory information is available. We hypothesised to see predictive saccades during the ISI between name prior and face on the white face outline. We investigated the target locations of the first two saccades[68] and introduced a minimum latency criterion of 100 ms after face outline onset to exclude non-intentional saccades. Our ROIs were four rectangular regions with the same area, covering the distinct features of the four identities (forehead, chin, ears, and nose; Figure 1C). We computed a 2 ✕ 2 repeated measures ANOVA with the within-subject factors 'expectation' (expected, unexpected) and 'saccade' (first, second). We tested whether significantly more first or second saccades were performed toward the ROI with the expected facial feature (hereafter: 'expected ROI') compared to how often the ROI was fixated in trials in which its facial feature was not expected. We calculated relative frequencies to account for the different number of trials in which the facial feature of an ROI (e.g., the nose) was expected vs. unexpected (1:3). Relative frequencies (%) of the different ROIs were averaged, yielding mean percentages for each participant for the different combinations of 'saccade' and 'expectation'. Post-hoc tests (Bonferroni-corrected) were performed.

The aim of Experiment 2 was to investigate whether context influences the order of fixations as well as the time spent looking at expected or unexpected facial features. We used three measures to evaluate an expectation bias: (1) Order and ordinal number of fixations, (2) number of fixations, and (3) dwell time.[69–71] To evaluate an order effect (1), we conducted two analyses: firstly, in all trials (i.e., *match*, *mismatch*, *partial*), we tested whether participants fixated the expected ROI earliest out of all the four ROIs. We considered trials in which at least one of the ROIs was fixated and assigned ordinal numbers of 1–4. In case of missing fixations, ROIs that were not fixated were randomly assigned to one of the missing ordinal numbers in that trial.[70] We evaluated whether the distribution of ordinal numbers differed from a uniform distribution using a subject-level chi-square goodness of fit test (R-package {htestClust}[72]) and calculated Cramer's *V* as an effect size. Post-hoc proportional tests for clustered data were performed, Bonferroni-corrected for the number of tests ($N = 6$). As an estimate of effect size, we averaged the subject-level Cohen's *h*. Secondly, in *partial* and *mismatch* trials, separately, we tested whether the expected ROI, the ROI with the unexpected facial feature (hereafter: 'unexpected ROI'), or the other two ROIs (hereafter: 'other ROIs') were fixated more often before the other by calculating tests of marginal proportion for clustered data against binomial distributions,[72] tested against $\pi = 0.5$ and $\pi = 0.33$, respectively, to account for the twice as large area of the other ROIs. We considered trials in which at least one of the respective ROIs had been fixated. Next, we evaluated an expectation effect on the number and duration of fixations (2 and 3): We conducted paired Wilcoxon signed-rank tests (two-tailed) to test whether the number of fixations and/or the dwell time on the expected ROI differed compared to the unexpected ROI and the average across the other ROIs. The number of fixations and dwell times on each ROI were divided by the total number of fixations and face presentation duration, respectively, to yield proportions.

Number of fixations and dwell time analysis by bin

We explored whether expectations influenced information sampling in face morphs in early time windows of the presented face (Figure S1). Therefore, we analyzed the relative number of fixations and dwell time on the expected, unexpected, and other ROIs by splitting the face presentation duration into time windows of 500 ms. As the presentation duration in each trial depended on the response given by participants, data availability for different time windows varied. We considered the first four time windows (i.e., 2000 ms) given that the fourth window was the latest window in which every participant had more than one trial for data analysis. We conducted 3 ✕ 4 repeated-measures ANOVAs with the within-subject factors 'ROI' (expected, unexpected, others) and 'bin' (500 ms steps) on the number of fixations and dwell times as the dependent variables. We report Greenhouse-Geisser corrected results as well as partial η^2. Post-hoc tests were Bonferroni-corrected and restricted to comparisons of interest, i.e., across conditions within each time window and within conditions across time windows.

Fixation durations

We explored fixation durations and conducted a one-way repeated-measures ANOVA with the within-subject factor 'ROI' (expected, unexpected, others) on average fixation durations as the dependent variable. We report η^2 as a measure of effect size. Post-hoc tests were Bonferroni-corrected.

Combined behavioral and eye-tracking analyses

Lastly, we investigated whether there was a link between eye movements and the responses given by the participants. In Experiment 1, we tested whether accuracy for identifying faces as expected or unexpected (in *match* and *mismatch* trials) was higher in trials in which participants fixated the expected ROI at face onset compared to when they did not fixate it in paired Wilcoxon signed-rank tests (one-sided). Correspondingly, we investigated whether participants chose the expected identity in a face morph if they fixated the expected vs. the unexpected ROI at face onset using a Wilcoxon signed-rank test (one-sided).

In Experiment 2, we tested whether participants more often chose the identity in a face morph if they fixated their distinct ROI last[49] using a proportional test for clustered data against chance level ($\pi = 0.25$).[72]

ADDITIONAL RESOURCES

The hypotheses, methods, and analysis plan were preregistered via the OSF (Experiment 1: https://osf.io/c2ydh; Experiment 2: https://osf.io/vxyrg). For minor deviations from these preregistrations see the detailed description in Tables S1 and S2 (template by Willroth and Atherton[73]).

Supplemental information

Expectations guide predictive eye movements

and information sampling during face recognition

Annika Garlichs, Mark Lustig, Matthias Gamer, and Helen Blank

Supplemental Tables

Table S1. Preregistration Deviations for Experiment 1, related to STAR Methods.

Deviations

#	Details		Original Wording	Deviation Description	Reader Impact
1	Type	Sample		During participant recruitment, we considered this limitation as too strict and reformulated it to "fluent German skills" in the adverts. Furthermore, we included one participant who spoke English by translating the instructions.	By choosing less strict inclusion criteria, the results are more likely to represent a broader population. We repeated all analyses excluding the one English speaker, revealing that this did not influence the results.
	Reason	Other (Please Explain)	Inclusion criterion: "German as first language"		
	Timing	During data collection			
2	Type	Other (Please Explain)	"At the beginning of each block, a 9-point-calibration is performed."	We used a 13-point-calibration, if possible. In case of calibration issues (e.g., due to dense eyelashes), we used a validated 9- or 5-point-calibration.	A calibration with fewer points might lead to decreased accuracy of measured eye movements. In the final sample, a 9-point-calibration was used for four participants, and a 5-point-calibration for one participant.
	Reason	New knowledge			
	Timing	During data collection			
3	Type	Analysis	Behavioural analyses: "Secondly, we will test whether there is a significant difference in the accuracy scores of the conditions match and mismatch using a paired t-test. Furthermore, we will test the accuracy scores of	We used a paired Wilcoxon signed rank test (two-sided) as well as one-sample Wilcoxon signed rank tests (one-sided) to evaluate significance. The difference scores of *match* and *mismatch* accuracies were not	The non-parametric tests might have led to a decrease in power, making it harder to find an effect. A paired *t*-test (two-sided) would have led to the same non-significant
	Reason	New knowledge			
	Timing	After results known			

Unregistered Steps

#	Details	Unregistered Step Description	Reader Impact
		both conditions against chance level (.50) using one-sample t-tests (one-sided)."	normally distributed (Shapiro-Wilk test: *p* = .01), and neither were the accuracies of the conditions themselves (*match: p* = .0011; *mismatch: p* < .001). Furthermore, there were outliers present (*match:* Z = -3.23, Z = -1.86, Z = -1.77, Z = -1.68, Z = -1.48; *mismatch:* Z = -3.96, Z = -2.52).	result: *t*(33) = 1.24, *p* = 0.22, 95% CI [-1.65, 6.79]. One-sample *t*-tests (one-sided) would have led to the same significant results: *match: t*(33) = 21.58, *p* < .001, [83.57, Inf]; *mismatch: t*(33) = 17.66, *p* < .001, [80.62, Inf].
4	Type	Other (Please Explain)		
	Reason	Other (Please Explain)	We renamed the factor priming (primed, not primed) to expectation (expected, unexpected) for coherent wording in our manuscript.	This does not affect the results.
	Timing	After results known	Eye-tracking analysis: "2 × 2 repeated measures ANOVA with the within-subject factors saccade number (first, second) and priming (primed, not primed)"	
5	Type	Analysis	Combined behavioural and eye-tracking analysis: "We will investigate whether participants have significantly higher accuracies for match and mismatch trials in which they predictively fixated the ROI with the expected face property at face onset compared to trials in which they did not fixate it. For testing significance, we will use paired t-tests (one-sided)."	We used paired Wilcoxon signed rank tests (one-sided) due to outliers in the data (*match:* ROI fixated: Z = -3.59, Z = -2.31, Z = -1.99; *mismatch:* ROI fixated: Z = -4.10, Z = -2.44; ROI not fixated: Z = -2.47, Z = -1.98).
	Reason	New knowledge		The non-parametric tests might have led to a decrease in power, making it harder to find an effect.
	Timing	After results known		Paired *t*-tests (one-sided) would have led to the same significant results (*match: t*(33) = -7.60, *p* < .001, 95% CI [11.73, Inf]; *mismatch: t*(33) = 4.62, *p* < .001, [6.74, Inf]).

#				
1	Type	Sample	Not preregistered	Exclusion criterion: During data collection, we noticed that few participants did not sufficiently learn the four identities and their distinct features in the training. We defined a cut-off threshold of 75%. We allowed participants to repeat the last part of the training up to two times. Two participants were not able to proceed to the main experiment (accuracies of the second repetition: 25.00% and 68.75%).	The additional exclusion criterion might have enhanced the results by preventing participants who did not learn the identities from proceeding to the main experiment.
	Timing	During data collection			
2	Type	Analysis	Not preregistered	Combined behavioural and eye-tracking analysis: We exploratively investigated "whether participants chose the expected identity in a face morph if they fixated the expected vs. the unexpected ROI at face onset using a Wilcoxon signed rank test (one-sided)." This additional analysis provides valuable insights into the possible source of the observed assimilation effect.	The result provides an additional interpretation possibility for the assimilation effect.
	Timing	After results known			

Table S2. Preregistration Deviations for Experiment 2, related to STAR Methods.

Deviations

#	Details	Original Wording	Deviation Description	Reader Impact
1	Type	Sample	Other (Please Explain)	Inclusion criterion: "German as first language"	By choosing less strict inclusion criteria, the results are more likely to represent a broader population.
	Reason	Other (Please Explain)		
	Timing	During data collection	During participant recruitment, we considered this limitation as too strict and reformulated it to "fluent German skills" in the adverts.	
2	Type	Study Design	"At the beginning of each block, a 9-point-calibration is performed."	We used a 13-point-calibration, if possible. In case of calibration issues (e.g., due to dense eyelashes), we used a validated 9- or 5-point-calibration.	A calibration with fewer points might lead to decreased accuracy of measured eye movements. In the final sample, a 9-point-calibration was used for two participants, a 5-point-calibration for one participant.
	Reason	New knowledge		
	Timing	During data collection		
3	Type	Analysis	"Secondly, we will test whether there is a significant difference in the accuracy scores of the conditions match and mismatch using a paired t-test. Furthermore, we will test the accuracy scores of both conditions against chance level (.50) using one-sample t-tests (one-sided)."	Behavioural analyses: We used a paired Wilcoxon signed rank test (two-sided) as well as one-sample Wilcoxon signed rank tests (one-sided) to evaluate significance. The difference scores of match and mismatch accuracies were not normally distributed (Shapiro-Wilk test: $p < .001$), neither were the accuracies of the conditions themselves (match: $p = .009$; mismatch: $p < .001$).	Behavioural analyses: The non-parametric tests might have led to a decrease in power, making it harder to find an effect. A paired t-test (two-sided) would have led to the same significant result: $t(33) = 5.83$, $p < .001$, 95% CI [6.60, 13.68]. One-sample t-tests (one-sided) would have led to the same significant results: match: $t(33) = 184.40$, $p < .001$, [97.63, Inf]; mismatch: $t(33) = 20.76$, $p < .001$, [84.84, Inf].
	Reason	New knowledge		
	Timing	After results known		
4	Type	Analysis	"For partial match trials, we will test whether (a) participants spend a significantly different	Eye-tracking analyses: Instead of averaging over the three ROIs not containing the expected feature, we computed	This increased the clarity of our statistical analyses, ruling out confounds of bottom-up information in
	Reason	New knowledge		

Timing	After results known	dwell time on the expected ROI compared to the average proportion of dwell time on all other three ROIs. We will further test whether (b) participants spend a significantly different dwell time on the expected ROI compared to the ROI with the face property of the other identity in the face morph. … "Secondly, we will compare the proportion of fixations on the ROI with the expected face property to the proportion of fixations (a) on the average of all other ROIs, and (b) on the ROI with the face property of the other identity in the face morph. We will use paired t-tests (two-tailed) to test for significance."	comparisons between the expected ROI and the unexpected ROI or the other two ROIs, respectively. The unexpected ROI is qualitatively different from the other two ROIs not containing any salient information. Furthermore, we used paired Wilcoxon signed rank tests (two-sided) to evaluate significance due to outliers in the data (number of fixations: expected: $Z = 2.97$; other two: $Z = 2.77$, $Z = 2.17$; dwell time: other two: $Z = 2.71$, $Z = 2.56$). For comparability, we also calculated a Wilcoxon test for dwell time 'expected vs. unexpected'.	an averaging over the other ROIs. Paired t-tests would have led to the same results: Number of fixations: expected vs. unexpected: $t(33) = 6.73$, $p < .001$, 95% CI [3.20, 5.98], $d = 1.15$; expected vs. other two: $t(33) = 15.12$, $p < .001$, 95% CI [9.73, 12.75], $d = 2.59$. Dwell time: expected vs. unexpected: $t(33) = 5.96$, $p < .001$, 95% CI [2.65, 5.40], $d = 1.02$; expected vs. other two: $t(33) = 14.75$, $p < .001$, 95% CI [9.33, 12.32], $d = 2.53$.

Unregistered Steps

No.					
1	Type	Sample	Not preregistered		
	Timing	During data collection		Exclusion criterion: During data collection of the first experiment, we noticed that some participants did not sufficiently learn the four identities and their distinct features in the training. We defined a cut-off threshold of 75%. We allowed participants to repeat the last part of the training up to two times. All participants proceeded to the main experiment.	This did not affect the results as no participant was excluded.
2	Type	Analysis	Not preregistered		

No.	Type	Timing	Analysis	Justification
	Analysis	After results known	Eye-tracking analysis: We investigated in *mismatch* trials whether the expected or unexpected ROI were fixated more often before each other and more often before the other two ROIs, respectively. We realised that the mismatch trials allowed us to disentangle whether expectations would guide eye movements even in the presence of unexpected salient features.	The *mismatch* trials provide valuable insights into whether bottom-up saliency can counterbalance top-down expectations in guiding participants' eye movements.
3	Analysis	Not preregistered	Eye-tracking analysis: We investigated whether fixation order in each of the four ROIs differed depending on expectation.	This analysis allowed us to show that the first fixations predominantly landed in the nose ROI, followed by expectations guiding the second fixation.
3	Other (Please Explain)	Not preregistered, suggested during revision	Eye-tracking analysis: We explored early viewing patterns towards the expected, unexpected, and other two ROIs in the first 2000 ms, split up into bins of 500 ms. We computed two 3 × 4 repeated-measures ANOVAs with the within-subject factors 'ROI' (expected, unexpected, others) and 'bin' (500 ms steps), on the number of fixations and dwell times as the dependent variables.	This analysis allowed us to investigate whether the congruency effect in the number of fixations and dwell times, i.e., a preferred sampling of the expected facial feature in a face morph, varied within the face presentation duration.
4	Other (Please Explain)	Not preregistered, suggested during revision	Eye-tracking analysis: We explored whether single fixation durations towards the expected, unexpected, and other two ROIs were influenced by expectation. We conducted a one-way repeated-measures ANOVA with the within-subject factor 'ROI' (expected, unexpected, others) on average fixation durations (ms) as the dependent variable.	This analysis allowed us to investigate whether the processing of unexpected compared to expected facial features is linked to higher difficulty and cognitive processing.

Figure S1. Early viewing of face morphs in Experiment 2, related to STAR Methods. A) Relative number of fixations: Effects of expectation on early viewing behaviour were investigated by splitting the first 2000 ms of face presentation in *partial* trials into bins of 500 ms. There was an interaction between 'ROI' and 'bin', showing that after initial preferred sampling of the expected region of interest (ROI) in the first and second bin, participants preferentially fixated on the unexpected ROI in the fourth bin. Dots represent single participants, white dots means, grey rectangles 95% confidence intervals (CI), and the lower and upper whiskers $Q_{1/3}$ -/+ 1.5*IQR. Post-hoc tests were Bonferroni-corrected for the comparisons of interest, i.e., across conditions within each time window and within conditions across time windows Lines indicate $p < .05$ (black/solid: across conditions, within bin; dark purple/dashed: within 'expected', across bins; rose/dotted: within 'unexpected', across bins. **B) Relative dwell time**: Similar to the relative number of fixations, dwell time analyses revealed an interaction between 'ROI' and 'bin'. Light rose/dashed-dotted: within 'others', across bins.
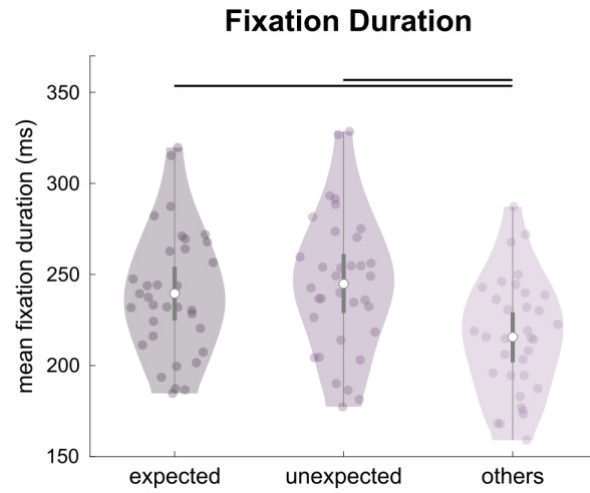
Figure S2. Average fixation duration for face morphs in Experiment 2, related to STAR Methods. For each participant, average fixation durations (ms) for the expected, unexpected, and other two ('others') regions of interest (ROI) were calculated. Dots represent single participants, white dots means, grey rectangles 95% confidence intervals (CI), and the lower and upper whiskers $Q_{1/3}$ -/+ 1.5*IQR. Black lines indicate $p < .05$.
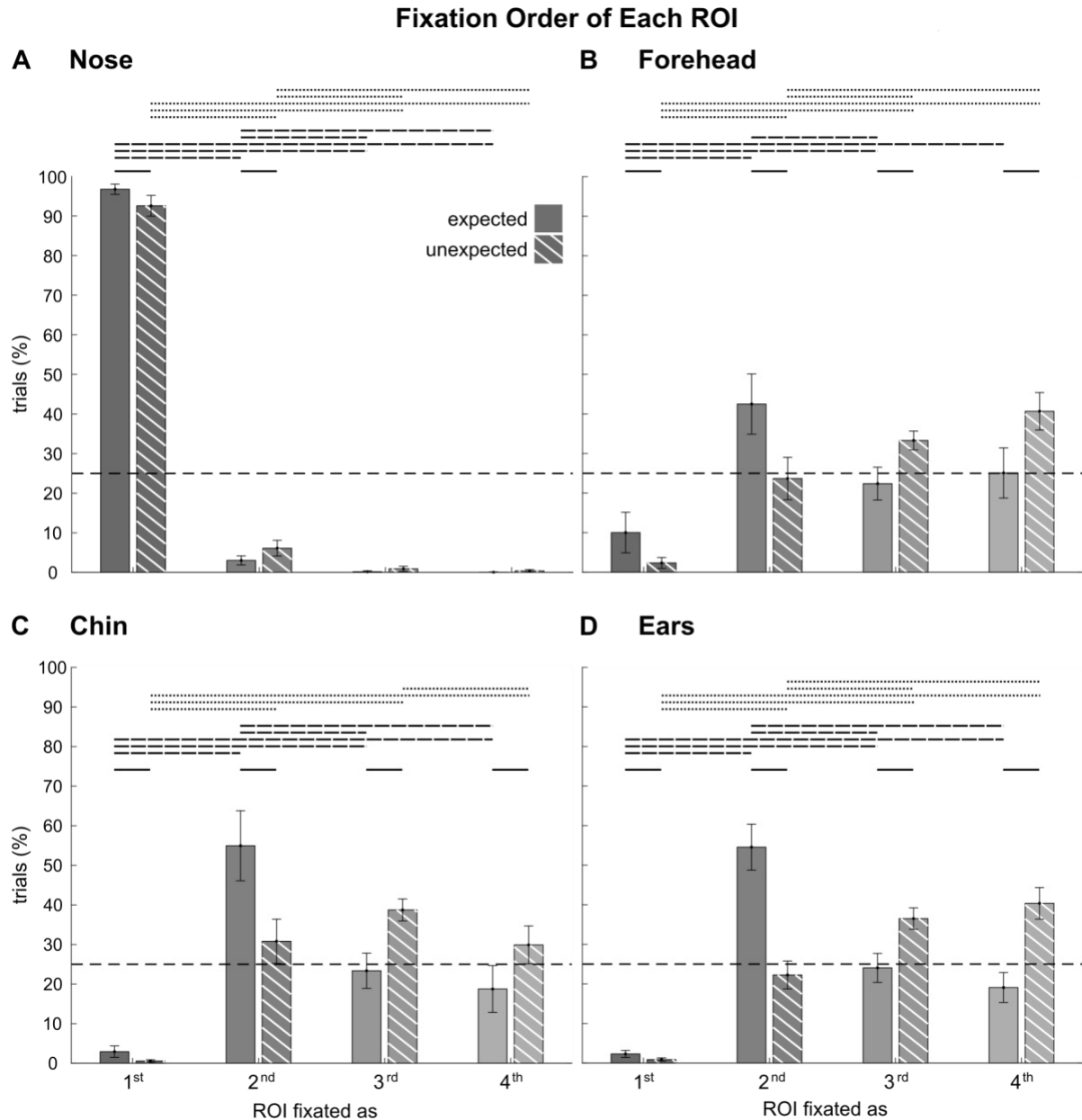
Figure S3. Order of fixations on the four regions of interest (ROI), related to Figure 4A. Depicted are percentages of how often each ROI was fixated as the first, second, third, or fourth out of all ROIs (**nose (A)**, **forehead (B)**, **chin (C)**, and **ears (D)**). Full and striped bars represent trials in which the ROI was expected or not expected, respectively. First fixations landed predominantly in the nose ROI, in line with a central viewing tendency. In the other three ROIs, expectations guided the second fixation towards an ROI. Error bars indicate 95% CIs. The dashed line represents the chance level ($\pi = .25$). Lines at the top indicate $p < .05$ (solid: expected vs. unexpected within each order number; dashed: within the condition 'expected'; dotted: within the condition 'unexpected'). The order analysis revealed that expectations guided eye movements towards the expected ROI so that it was fixated first or second more often than third or fourth (Figure 4A). In the following exploratory analysis, we further investigated whether fixation order differed depending on the ROI. In case of missing fixations within a trial, random ordinal numbers were assigned as described in the STAR Methods. Using subject-level chi-square goodness of fit tests, we evaluated whether the distribution of ordinal numbers differed from a uniform distribution within each ROI, for trials in which the ROI was either expected or unexpected. Cramer's *V* was calculated as an effect size. Post-hoc proportional tests for clustered data were performed within

each ROI, Bonferroni-corrected for the number of tests (expected: $N = 6$; unexpected: $N = 6$; expected vs. unexpected: $N = 4$). In case of missing values in both conditions of interest, post-hoc tests were performed with the remaining participants. As an estimate of effect size, we averaged the subject-level Cohen's h. The order in which a ROI was fixated differed, both when its facial feature was expected (nose: $\chi^2(3, N = 34) = 34.00$, $p < .001$, $V = 0.07$; forehead: $\chi^2(3, N = 34) = 23.37$, $p < .001$, $V = 0.06$; chin: $\chi^2(3, N = 34) = 33.21$, $p < .001$, $V = 0.07$; ears: $\chi^2(3, N = 34) = 33.61$, $p < .001$, $V = 0.07$), and when its feature was unexpected (nose: $\chi^2(3, N = 34) = 33.97$, $p < .001$, $V = 0.04$; forehead: $\chi^2(3, N = 34) = 33.09$, $p < .001$, $V = 0.04$; chin: $\chi^2(3, N = 34) = 33.98$, $p < .001$, $V = 0.04$; ears: $\chi^2(3, N = 34) = 33.93$, $p < .001$, $V = 0.04$). First fixations landed predominantly in the nose ROI, especially if it was expected ($z = 3.51$, $p = .002$, 95% CI [0.25, 0.26], $h = 1.03$). After this initial fixation, participants fixated second on the expected ROIs (expected vs. unexpected: 2nd: forehead: $z = 6.42$, $p < .001$, 95% CI [0.35, 0.44], $h = 0.74$; chin: $z = 6.43$, $p < .001$, 95% CI [0.34, 0.42], $h = 0.77$; ears: $z = 8.67$, $p < .001$, 95% CI [0.43, 0.53], $h = 0.57$). In line with our number of fixations and dwell time analyses over time (Figure S3), post-hoc tests further revealed a reversal in the later fixations, showing that unexpected ROIs attracted more third or fourth fixations (expected vs. unexpected: 3rd: forehead: $z = -2.95$, $p = .01$, 95% CI [0.14, 0.23], $h = 1.23$; chin: $z = -3.33$, $p = .003$, 95% CI [0.12, 0.22], $h = 1.26$; ears: $z = -3.25$, $p = .005$, 95% CI [0.15, 0.22], $h = 1.21$; 4th: forehead: $z = -3.28$, $p = .004$, 95% CI [0.11, 0.21], $h = 1.30$; chin: $z = -3.41$, $p = .003$, 95% CI [0.10, 0.21], $h = 1.30$; ears: $z = -3.58$, $p = .001$, 95% CI [0.08, 0.20], $h = 1.34$).

18.2 Study 2: Prediction error processing and sharpening of expected information across the face-processing hierarchy

Garlichs, A., & Blank, H. (2024). Prediction error processing and sharpening of expected information across the face-processing hierarchy. *Nature Communications*, *15*(1), 3407.

Prediction error processing and sharpening of expected information across the face-processing hierarchy

Annika Garlichs [1] & Helen Blank [1]

The perception and neural processing of sensory information are strongly influenced by prior expectations. The integration of prior and sensory information can manifest through distinct underlying mechanisms: focusing on unexpected input, denoted as prediction error (PE) processing, or amplifying anticipated information via sharpened representation. In this study, we employed computational modeling using deep neural networks combined with representational similarity analyses of fMRI data to investigate these two processes during face perception. Participants were cued to see face images, some generated by morphing two faces, leading to ambiguity in face identity. We show that expected faces were identified faster and perception of ambiguous faces was shifted towards priors. Multivariate analyses uncovered evidence for PE processing across and beyond the face-processing hierarchy from the occipital face area (OFA), via the fusiform face area, to the anterior temporal lobe, and suggest sharpened representations in the OFA. Our findings support the proposition that the brain represents faces grounded in prior expectations.

It is widely accepted that perception is a process of active inference in which incoming sensory information is combined with priors that were either learned or derived from the current context[1–3]. Expectations can enhance our ability to recognise familiar stimuli more quickly and accurately. For instance, recognising a colleague's face in the office is easier than spotting them at the beach. However, expectations can also introduce a bias in our perception when faced with ambiguous sensory information. For instance, from a distance, we might mistakenly categorise a distant individual as a friend due to their attire, even if they are, in fact, a stranger. The neural mechanism of how expectations influence representations of sensory information is still unclear. Here, we combined multivariate functional magnetic resonance imaging (fMRI) with neural network models to test whether face representations mainly rely on the processing of deviances from expectations (i.e., Prediction Errors) or sharpening of expected information[4–7].

Context effects on face perception have been studied extensively showing assimilative[8–16] as well as contrastive[17–19] behavioural effects.

In the brain, an expectation suppression effect, i.e., reduced neural activation for expected compared to unexpected face information, has been reported across a variety of different designs and brain measures (electroencephalography (EEG)[20], magnetoencephalography (MEG)[21], and fMRI[22–27]). However, it is still unclear how prior and incoming sensory information are combined. Different computational mechanisms could underlie reduced activation for expected faces: According to the hierarchical framework of predictive coding, higher-level 'representational units' generate backward predictions concerning anticipated sensory information, which are then compared with the actual sensory input in lower-level 'error units' to compute the prediction error (PE)[2,5,6]. These PEs may play a crucial role in updating prior expectations about incoming sensory information, thereby improving predictive accuracy[2,7,28,29]. Consequently, the phenomenon of expectation suppression may be explained by a diminished PE for expected faces relative to unexpected ones. Alternatively, this expectation effect could also be attributed to a computational mechanism

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany. ✉e-mail: a.garlichs@uke.de; h.blank@uke.de

focusing on the sharpening of expected information[4,7,30–33]. Under the Sharpening account, neurons encoding the expected features become more active, whereas neurons encoding unexpected features are suppressed. At the population level, this would result in a more selective response for expected stimuli with lower overall amplitude. Consequently, weaker univariate activity might reflect a "sharper" neural population response for expected sensory events and suppression of unexpected noise rather than a suppression of the expected signals[32–35]. Since both computational processes lead to decreased activation for expected stimuli compared to unexpected ones and are indistinguishable at the univariate analysis level, our study was designed to differentiate between them using multivariate analyses[4,30,36,37].

To do this, we investigated face representations in the well-established face-processing hierarchy along the ventral stream of the temporal lobe[38–44]. Previous studies have demonstrated that there is a progression of higher-level feature analysis in the processing of facial information, moving from lower to higher face-processing regions[9,42,45]. Specifically, the occipital face area (OFA) has shown sensitivity to low-level image properties such as the eyes, nose, and mouth[9,46,47]. The fusiform face area (FFA) processes a combination of low-level properties[48,49], as well as higher-level face properties, including traits, gender[47], and identity[9,50,51]. Finally, the face-sensitive anterior temporal lobe (aTL) specifically encodes identity information[9,51], which remains consistent across different images[45,52,53]. The influence of face priors has been observed in the form of expectation suppression effects in OFA[24] and FFA[23–26,54]. Increased activity to unexpected faces in the FFA has been taken as evidence for PEs, i.e., the difference between expected and presented faces[23,26,55]. In the macaque brain, which exhibits a face-processing hierarchy similar to humans, signals recorded at the lowest level ML (comparable to the human OFA) displayed identity-specific information derived from higher levels. This finding was considered as evidence for predictions transmitted from higher to lower levels, where incoming face information is represented as deviating information[56]. However, others did not observe any neural indication of repetition probability for faces

within face-responsive patches of the macaque IT[57,58]. In contrast, recent studies have provided evidence for the sharpening of prior information along the ventral processing stream. Our research demonstrated that the strength of face prior representations can be quantified through multivoxel fMRI patterns in the high-level face-sensitive aTL[39]. In addition, we identified multivariate representations of presented faces that increased with expectedness in the OFA, indicating a potential sharpening of expected low-level facial features. This finding is corroborated by a study that demonstrated the enhancement of prior information across the ventral stream in sensory-degraded Mooney face images, from early visual areas and extending throughout the lateral occipital cortex and the fusiform gyrus[31]. However, there is a lack of research that directly compares and tests alternative explanations and computational mechanisms against each other to determine how face images are represented based on prior information.

In this study, we tested how the representation of identical face images is changed by different prior expectations by investigating multivariate fMRI response patterns from a paradigm involving ambiguous face images that were created by morphing an expected and an unexpected face image (Fig. 1a–c). In a preceding training, participants learned to associate scene images with subsequently presented face images (Fig. 1a). During the following fMRI session, participants viewed the scene cues followed by expected, unexpected, or morphed ambiguous face images (Fig. 1b). Our design allowed us to differentiate whether the neural representations for the same face morph differed depending on the expectation and was better explained by a computational model based on PE processing or based on sharpened representations of expected face information (Fig. 1c). Deep convolutional neural networks (DCNN) can be viewed as advanced computational models for biological face recognition that process information hierarchically, closely resembling the neural face-recognition system found in humans and nonhuman primates[59–62]. Combining computational modelling with neural network activations based on the face-recognition DCNN VGG-Face[60,63] (Fig. 2a) allowed us to optimise our hypothesis models for the representational similarity

a Overall Procedure

b Exemplary Trial
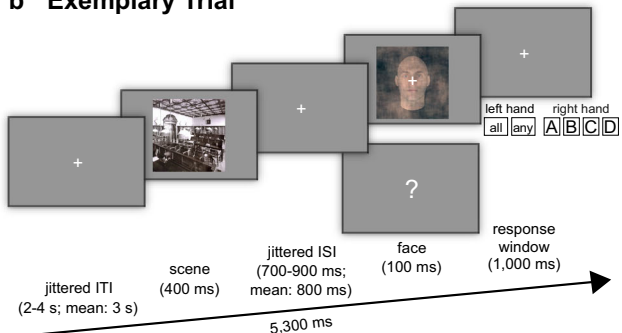
c Conditions and Research Question
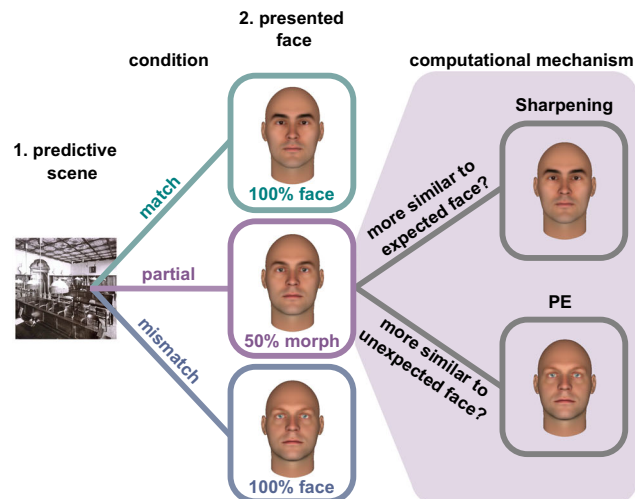
Fig. 1 | **Procedure, design, and conditions of the fMRI experiment. a** Procedure: Individual morph levels of face images were calibrated for each participant. After training to associate scenes with face images, the experiment took place in the scanner. **b** Trial: A scene was followed by a face image. The task was to identify the face and, in case of a question mark, to indicate which face was anticipated based on the scene. After the neutral scene, the task was to indicate with the left hand 'any' face or, in case of a question mark, that 'all' four faces had been anticipated. **c** Conditions and research question: There were four scenes predictive of the

upcoming face, and one neutral scene. After the four scenes, a clear face that matched or mismatched the prediction or an ambiguous face that contained the predicted face appeared. We aimed at differentiating whether the representation of face morphs depends on Prediction Errors (PE) or the Sharpening of expected facial features. The scene image shown is in the public domain and available at [https://commons.wikimedia.org], but not part of the original stimulus set due to copyright; the exact stimulus set is available at [https://osf.io/765jx/]. The face images were created using FaceGen Modeller Core 3.22.
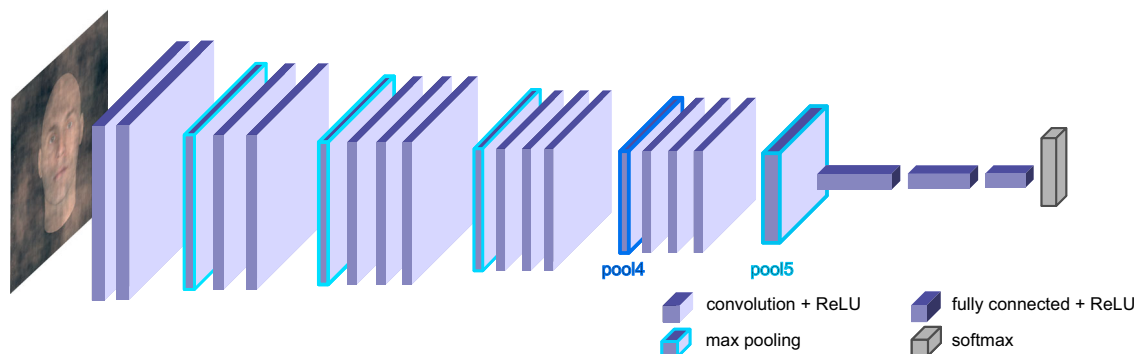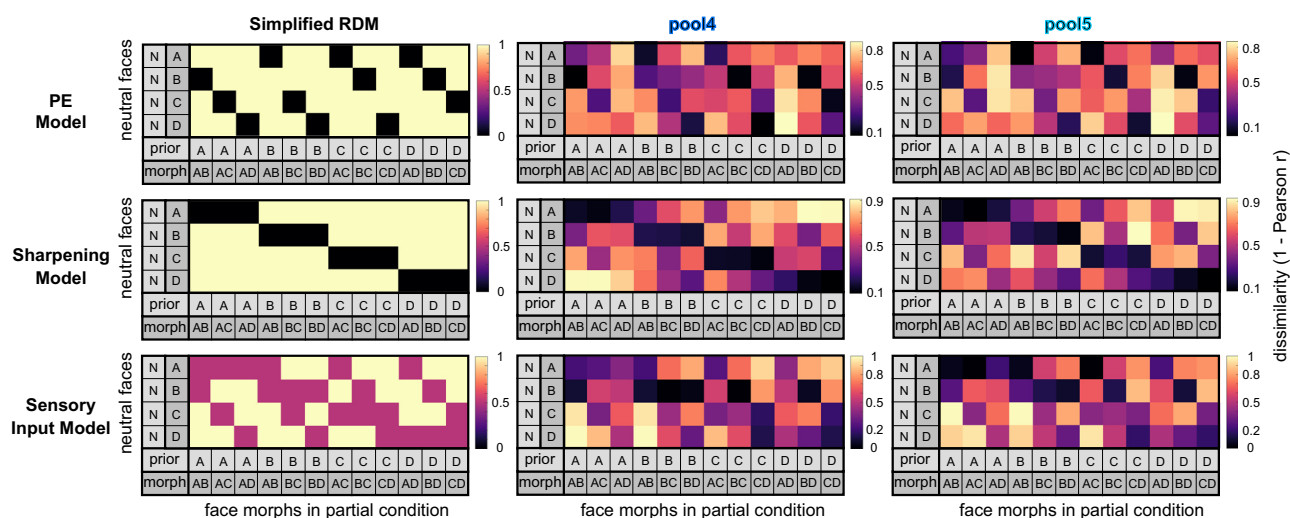
Fig. 2 | The neural network architecture of VGG-Face and hypothesis representational dissimilarity matrices (RDM). a Schematic architecture of the deep convolutional neural network VGG-Face[63]: ReLU rectification linear unit. **b** Hypothesis RDMs: The PE (top), Sharpening (middle), and pure Sensory Input (bottom) models were used as hypothesis models in the RSA[64,65]. For visualisation purposes, we included simplified RDMs (left panel) to demonstrate the theoretical dissimilarities of the models without network activations and behavioural weighting. For RDM creation, the activations of the neutral and morph images were extracted from the layers pool4 and pool5 of the network VGG-Face before calculating their dissimilarities. Specifically, for prior activations in the PE and Sharpening model, we utilised neutral face images weighted by individual behaviour based on how strongly each prior influenced the perception of the following morph. For visualisation, the displayed PE and Sharpening RDMs were averaged across individual (N = 43) RDMs.

analysis (RSA)[64,65] (Fig. 2b). We derived face activations from the final two max pooling steps of this network, namely pool4 and pool5, to construct our hypothesised representational dissimilarity matrices (RDM). This choice was informed by a recent intracranial electroencephalography study associating these layers with our brain regions of interest − specifically, pool4 was linked to lower-level inferior occipital gyrus, while pool5 was associated with higher-level face processing in the fusiform gyrus[60]. In addition, we tested two more DCNNs (i.e., VGG-16[66] and ResNet50[67]) to explore whether face representations in the brain also correlate with face representations from DCNNs that were not specifically trained on face images. All three networks have previously been linked to brain activations in studies using different methods, such as MEG[59] and fMRI[61], for a review see ref. 62. Furthermore, to take into account that individuals may differ in their usage of prior information during ambiguous face identification, we included individual prior weights by contrasting prior-confirming with prior-discarding responses for a face morph in the RSA[30,68,69]. We show that PE representations dominate along and beyond the face-processing hierarchy, while there was also evidence for the co-existence of sharpened expected face information in early face areas. These PE and sharpened representations indicate a predictive

mechanism through which the brain integrates prior knowledge with sensory input.

Results

Assimilation and facilitation in perception of expected faces

We recorded fMRI data from 43 participants while they viewed and identified face images that were preceded by scene images. The scene prior shifted the perception of ambiguous face morphs towards the expected face identity (assimilation effect). Specifically, in partial trials, in which images of morphed faces contained expected and unexpected face information, participants identified the expected face identity more frequently than the unexpected identity ($Z = 5.65$, $p < 0.001$, 95% CI [63.54, 69.01], Wilcoxon's $r = 0.86$; Fig. 3a).

In addition, reaction times (RT) showed a facilitation effect due to expectancy (main effect of condition (match, partial, mismatch, neutral): $\chi^2(3) = 110.08$, $p < 0.001$, Kendall's $W = 0.85$; Fig. 3b, Supplementary Results). RTs for expected faces were faster compared to unexpected faces (match: $M = 591.08$ ms, $SD = 47.06$ ms; mismatch: $M = 727.56$, $SD = 47.82$; $p < 0.001$, LB/UB [−2.41, −0.98]) and ambiguous faces (partial: $M = 716.06$, SD = 50.28; $p < .001$, [−2.30, −0.87]). The RTs
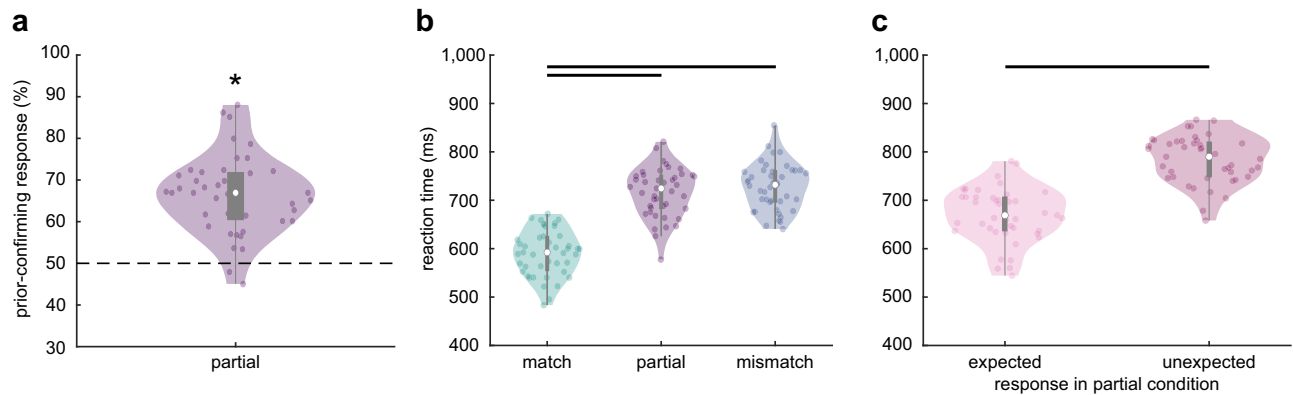
Fig. 3 | Behavioural results of the fMRI experiment. a Assimilation effect: Participants ($N = 43$) identified the expected face identity in morphed images (two-sided Wilcoxon signed rank test: $Z = 5.65$, $p < 0.001$, 95% CI [63.54, 69.01], Wilcoxon's $r = 0.86$). Dots represent single participants, the white dot the median, the grey rectangle the interquartile range (Q_1, Q_3), and the lower and upper whiskers $Q_1 - 1.5*IQR$ and $Q_3 + 1.5*IQR$, respectively. From (**a**) to (**c**), asterisks and black lines indicate $p < 0.001$. **b** Facilitation effect: Responses were faster for expected compared to (partially) unexpected faces ($N = 43$; Friedman: $\chi^2(3) = 110.08$, $p < 0.001$,

Kendall's $W = 0.85$; post-hoc tests (Tukey-corrected): match vs. mismatch: $p < 0.001$, LB/UB [−2.41, −0.98]; match vs. partial: $p < 0.001$, [−2.30, −0.87]; mismatch vs. partial: $p = 0.98$, [−0.60, 0.83]). **c** Reaction time analysis for face morphs: Partial trials were split into trials in which the expected face or the unexpected face had been identified in a morph ($N = 43$; two-sided paired Wilcoxon signed rank test: expected vs. unexpected: $Z = −5.71$, $p < 0.001$, 95% CI [−128.58, −96.72], Wilcoxon's $r = 0.87$). Source data are provided as a Source Data file.

between unexpected and morphed faces did not differ (mismatch vs. partial: $p = 0.98$, [−0.60, 0.83]).

Furthermore, there was a facilitation effect for face morphs depending on whether participants answered to have perceived the expected or the unexpected face identity. Partial trials were split into trials with prior-confirming responses (assimilation effect) and with responses favouring the unexpected identity contained in a morph (contrastive effect). RTs for trials with prior-confirming responses were faster ($M = 670.13$ ms, $SD = 55.96$ ms) than for trials with contrastive responses ($M = 783.83$, $SD = 48.67$, $Z = −5.71$, $p < 0.001$, 95% CI [−128.58, −96.72], Wilcoxon's $r = 0.87$, Fig. 3c), but still slower compared to responses in the match condition without face morphs ($Z = −5.64$, $p < 0.001$, [67.80, 93.07], Wilcoxon's $r = 0.86$).

Further control measures substantiated that participants effectively acquired knowledge of the associations and attentively considered both scene priors and face images (see Supplementary Results).

Expectations reduce evoked fMRI activations
To test which brain regions are overall differently activated by expected and unexpected faces, we conducted a univariate whole-brain analysis for unexpected vs. expected face images (contrast 'mismatch > match') that yielded a significant cluster along the ventral face-processing hierarchy in the left inferior and middle temporal gyrus (ITG/MTG, $p(FWE) = 0.005$; Fig. 4, Supplementary Table 1). Within the face-sensitive regions that were localised with the independent localiser (Fig. 4a), only the left posterior FFA (pFFA) showed an increased response to unexpected face images ($p(FWEsvc_{(small-volume\ corrected)}) = 0.042$). Additionally, this analysis revealed cluster activations in the bilateral anterior insula, superior parietal lobule (SPL) including the supramarginal gyrus and precuneus, left thalamus, and right caudate (Fig. 4b, Supplementary Table 1). Parts of this network are involved in surprise[39,70] as well as error processing[71].

Next, we tested which brain regions were overall differently activated during the presentation of face morphs depending on whether they were perceived as the expected or unexpected face. Therefore, we split the partial trials into trials in which participants answered to have perceived the expected and the unexpected part of the face morph. A bilateral cluster along the ventral stream resembling the 'mismatch > match' cluster was identified in the MTG (left:

$p(FWE) = 0.004$; right: $p(FWE) = 0.031$). In the ROIs along the ventral face-processing hierarchy, the right pFFA as well as the right aTL showed an increased response to morphed faces that were identified as the unexpected face ($p(FWEsvc) = 0.021$ and $p(FWEsvc) < 0.001$, respectively). Furthermore, the contrast 'unexpected > expected' yielded a similar activation network as the contrast 'mismatch > match', bilaterally in the SPL, angular gyrus (AnG), superior frontal gyrus, and right thalamus (all $p(FWE) < 0.05$ at the cluster level, Fig. 4b, Supplementary Table 2). Additional activation was found in the left anterior cingulate gyrus and bilaterally in the anterior orbital gyrus which are typically involved in decision-making processes[72,73].

Prediction error and sharpened representations of expected face information in face-sensitive regions
We used RSA to investigate how the information of a face prior was combined with the incoming face information[64,65]. Firstly, we computed theoretical representational dissimilarity matrices (i.e., hypothesis RDMs) based on activations from the layers pool4 and pool5 of the DCNN VGG-Face[63] for three computational approaches of how expected and presented face could be combined, i.e., PE, Sharpening, and a pure Sensory Input model without prior influence. The computational models were based on the face-recognition DCNN VGG-Face because the similarity structure of face-image transformations extracted from the layers pool4 and pool5 have been shown to correlate with the neural similarity structure of single-cell recordings in OFA and FFA, respectively[60]. Next, we compared the resulting hypothesis RDMs with the dissimilarity structure of our neural data (i.e., neural RDM). To obtain neural RDMs, we compared the multi-voxel representations of morphed faces measured in partial trials with the 'pure' face representations measured in the neutral trials.

By testing the correlation of hypothesis and neural RDMs, we found evidence for PE processing at each stage of the face-processing hierarchy (OFA: $M = 0.06$, SEMws$_{(within-subject)} = 0.01$, $p = 0.003$; pFFA: $M = 0.04$, $SEM_{ws} = 0.01$, $p = 0.0044$), indicated in aTL by higher correlations with the PE model compared to the Sharpening (PE: $M = 0.03$, $SEM_{ws} = 0.01$; Sharpening: $M = −0.02$, $SEM_{ws} = 0.02$, $p = 0.046$) and the Sensory Input model ($M = −0.01$, $SEM_{ws} = 0.01$, $p = 0.0154$; Fig. 4a, d–f, Supplementary Tables 3 and 4). Additionally, there was evidence for sharpened face representations in the OFA ($M = 0.03$, $SEM_{ws} = 0.01$, $p = 0.0232$). Furthermore, we investigated whether the reduced activation for expected faces observed in the univariate contrast
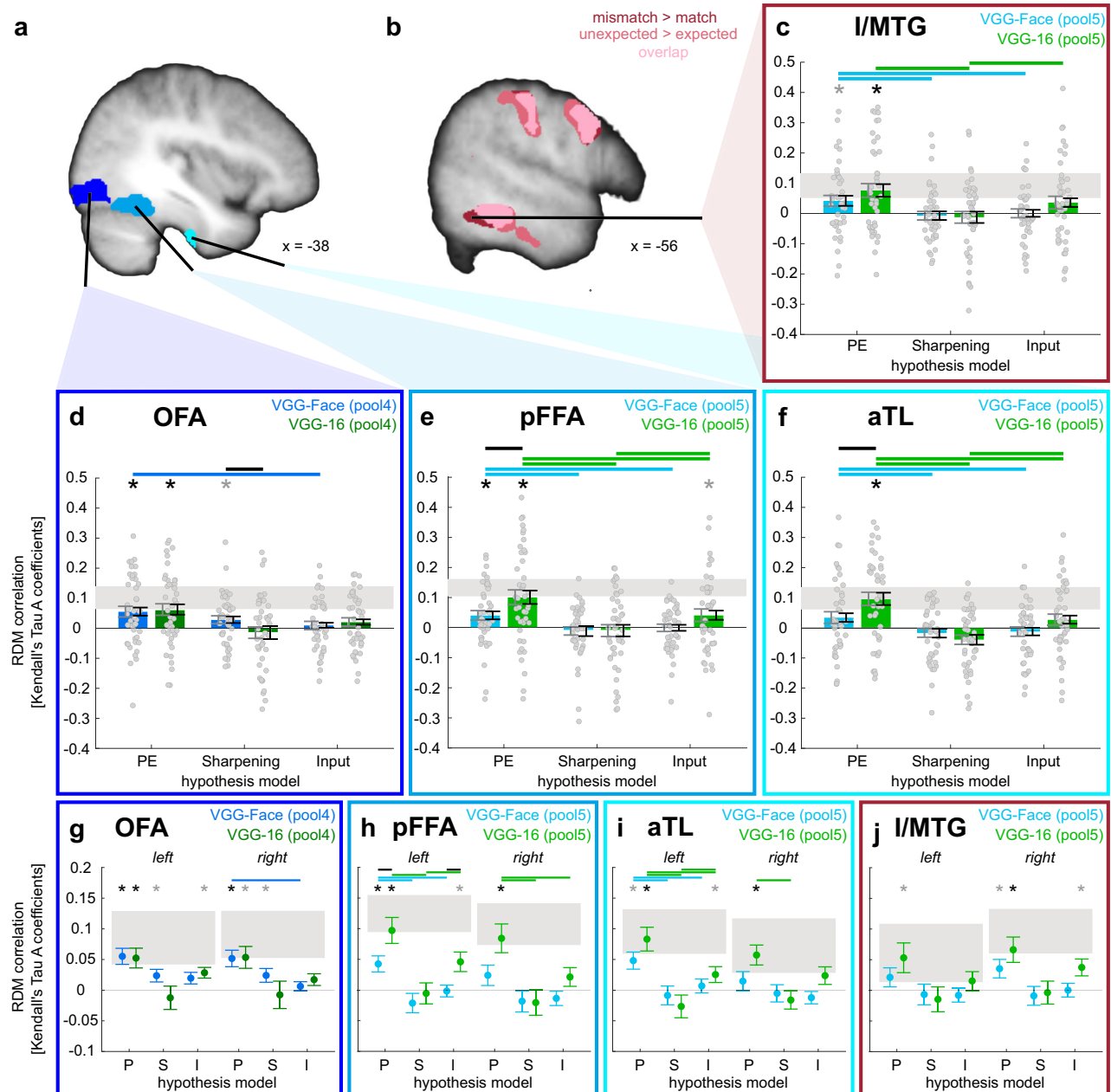
Fig. 4 | Multivariate representations of expected faces in face-sensitive regions. **a** Regions of interest (ROI): The face-sensitive ROIs along the ventral face-processing hierarchy, i.e., the occipital face area (OFA), posterior fusiform face area (pFFA), and anterior temporal lobe (aTL), were based on the independent functional localiser 'faces > scenes' (see Methods). **b** Reduced univariate activation for expected faces: The contrasts 'mismatch > match' and 'unexpected > expected' revealed enhanced activation in the inferior/middle temporal gyrus (ITG/MTG); shown at p(unc.) < 0.001, overlaid on the average structural T1 image in Montreal Neurological Institute (MNI) template space. **c–f** Representational similarity analysis (RSA) for both hemispheres: Hypothesis RDMs were based on representations extracted from two layers (pool4 and pool5) from two deep convolutional neural networks (DCNN) (VGG-Face, blue; VGG-16, green). The correlations between the hypothesis and neural RDMs were used to test the three hypothesis models: Prediction Error (PE), Sharpening, and a pure Sensory Input model. Grey error bars indicate the between-subject standard error of the mean (SEM), and black error bars the within-subject SEM[135] (N = 43 participants). For (**c**)–(**j**), asterisks indicate significance for each hypothesis model against zero (one-sided Wilcoxon signed rank tests), black Bonferroni-corrected for the number of tests per ROI (N = 6 (3 models × 2 DCNNs)), grey for p(unc.) < 0.05; horizontal lines indicate the significance of model comparisons (two-sided paired Wilcoxon signed rank tests) within DCNNSs (blue: VGG-Face, green: VGG-16), and black horizontal lines indicate significance of model comparisons between VGG-16 and VGG-Face, FDR-corrected[131] for the model comparisons per ROI. Grey rectangles display the lower and upper boundary of the noise ceiling for each ROI as an estimation of how well any model could perform given the noise in the data[65]. **g–j** RSA split up by hemisphere: Display of the corresponding RSA results for the three hypothesis models (i.e., P (Prediction Error), S (Sharpening), and I (Sensory Input)) split by hemisphere in the four ROIs (OFA, pFFA, aTL, and ITG/MTG) for the DCNNs VGG-Face (blue) and VGG-16 (green). Dots represent means, error bars the within-subject SEM (N = 43 participants). Source data are provided as a Source Data file.

'mismatch > match' might be due to a reduced PE or Sharpening processing and found more evidence for PE processing in the ITG/MTG cluster compared to sharpened representations (PE: M = 0.04, SEM_{ws} = 0.02; Sharpening: M = −0.01, SEM_{ws} = 0.01, p = 0.0366) and

pure sensory input processing (M = 0.0005, SEM_{ws} = 0.01, p = 0.0366; Fig. 4b, c, Supplementary Tables 3 and 4).

Secondly, we tested the correlations of the object-trained DCNNs with the neural dissimilarity structure. VGG-16 revealed evidence for PE
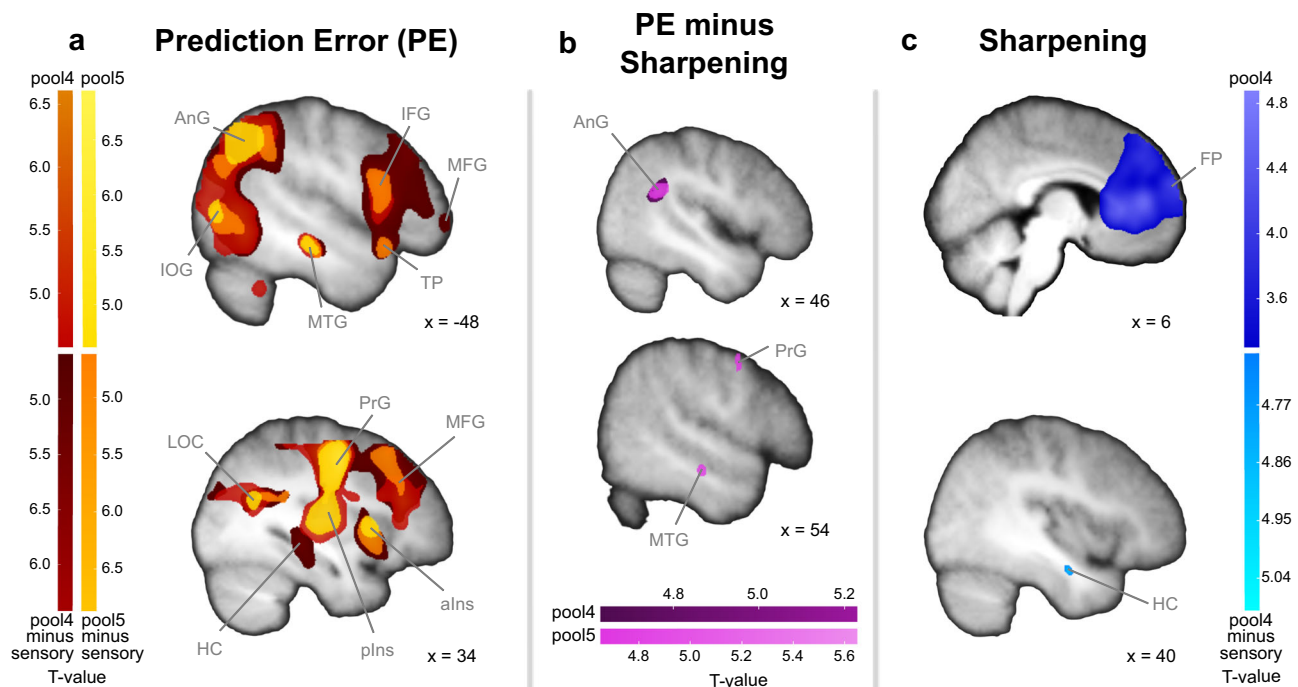
Fig. 5 | Whole-brain searchlight analyses for the hypothesis models Prediction Error (PE) and Sharpening based on VGG-Face. Results for the comparison of the neural and hypothesised dissimilarity structures based on pool4 and pool5 layers from VGG-Face are displayed against zero and as difference maps against a sensory input searchlight without prior influence based on the second convolutional layer[59], respectively. **a** Searchlight analyses results for PE: Clusters were identified in angular gyrus (AnG), inferior occipital gyrus (IOG), inferior frontal gyrus (IFG), middle frontal gyrus (MFG), temporal pole (TP), middle temporal gyrus (MTG), inferior occipital gyrus (IOG), lateral occipital cortex (LOC), precentral gyrus (PrG), anterior insula (aIns), posterior insula (pIns), and hippocampus (HC). **b** Comparison of the 'PE > Sharpening' searchlight results: Stronger correlations for PE than Sharpening were evident in the right AnG, bilateral SMG, PrG, and MTG. **c** Searchlight analysis results for Sharpening: Clusters were identified in the frontal pole (FP) and HC. All maps in (**a**)–(**c**) are displayed at p(FWE) < 0.05 (peak-corrected), except for Sharpening pool4 in (**c**), displayed at p(FWE) < 0.05 (cluster-corrected), with a cluster-inducing threshold of p(unc.) < 0.001. Maps are overlaid on the average structural T1 image in the Montreal Neurological Institute (MNI) template space. Source data are provided as a Source Data file.

processing along the whole face-processing hierarchy in line with VGG-Face (OFA: $M = 0.06$, $SEM_{ws} = 0.02$, $p = 0.0025$; pFFA: $M = 0.10$, $SEM_{ws} = 0.02$, $p = 0.0002$; aTL: $M = 0.10$, $SEM_{ws} = 0.02$, $p = 0.0001$; MTG: $M = 0.08$, $SEM_{ws} = 0.02$, $p = 0.0038$), with significantly higher correlations in pFFA and aTL than VGG-Face (pFFA: $Z = -2.40$, $p = .0315$, Wilcoxon's $r = 0.37$; aTL: $Z = -2.58$, $p = 0.0176$, Wilcoxon's $r = 0.39$; Fig. 4c–f, Supplementary Tables 5–7). ResNet50 showed consistent evidence for PE in pFFA, aTL, and MTG (pFFA: $M = 0.05$, $SEM_{ws} = 0.02$, $p = 0.0155$; aTL: $M = 0.05$, $SEM_{ws} = 0.02$, $p = 0.0151$; MTG: $M = 0.05$, $SEM_{ws} = 0.02$, $p = 0.0064$), and a trend in OFA ($M = 0.04$, $SEM_{ws} = 0.03$, $p = 0.10$; Supplementary Fig. 1, Supplementary Tables 8 and 9). Contrary to VGG-Face, both networks did not reveal any correlations with the hypothesis RDMs based on Sharpening in OFA (VGG-16: $M = -0.01$, $SEM_{ws} = 0.02$, $p = 0.70$; ResNet50: $M = -0.03$, $SEM_{ws} = 0.02$, $p = 0.95$), nor in any of the other ROIs.

Additionally, we conducted multivariate ROI analyses split up by hemisphere to explore lateralized representations (Fig. 4g–j, Supplementary Fig. 1, Supplementary Results). As in the bilateral analyses, we identified PE processing along the whole face-processing hierarchy for all three DCNNs (VGG-Face, VGG-16, ResNet50), evident by stronger correlations with the PE model compared to the Sensory Input and/or Sharpening model in all ROIs (Supplementary Results, Supplementary Tables 10–13). There was no main effect of hemisphere in any of the networks.

Prediction error and sharpened representations of expected face information in the whole brain

Furthermore, we conducted searchlight analyses to investigate how expected faces are represented in the whole brain beyond the predefined face-sensitive regions along the ventral stream (Fig. 5). All searchlight analyses were conducted with hypothesis models for PE and Sharpening both based on pool4 or pool5 VGG-Face[63] activations to test for lower-level as well as higher-level face image representations. With the lower-level hypothesis model for PE based on pool4, we identified correlations between the neural and hypothesised similarity structure in occipital and temporal regions, as well as in the right fusiform gyrus including the parahippocampal gyrus (p(FWE) < 0.05; Fig. 5a, Supplementary Table 14). Further large clusters were found in parietal and frontal regions. With the hypothesis model based on pool5, testing for higher-level PE representations, we found a similar pattern with additional correlations in the right insula (Fig. 5a, Supplementary Table 15). In contrast to a sensory input searchlight based on the second convolutional layer of VGG-Face[59], there was additional evidence for PE representations based on pool4 in the right hippocampus (p(FWE) < 0.05; Fig. 5a, Supplementary Tables 16 and 17). To further investigate the spatial overlap between the ROIs and the whole-brain searchlight approach, we conducted small-volume corrected analyses by applying our ROI masks. The ROIs overlapped with the PE searchlight maps in the OFA (pool4), as well as in the pFFA, aTL, and MTG (pool5), respectively (p(FWEsvc) < 0.001 in lOFA and rOFA; lpFFA: p(FWEsvc) = 0.001; rpFFA: p(FWEsvc) = 0.015; laTL: p(FWEsvc) = 0.015; raTL: p(FWEsvc) = 0.035; lMTG: p(FWEsvc) = 0.001; rMTG: p(FWEsvc) = 0.009).

The searchlight analysis testing for Sharpening based on pool4 revealed further evidence for enhanced representations of expected faces in the frontal lobe and postcentral gyrus (Fig. 5c, Supplementary Table 18). In contrast to a sensory input searchlight, there was evidence for sharpened representations in the right hippocampus (p(FWE) = 0.017, [34 −6 −22], $k = 64$; Fig. 5c), which was located more anterior than the hippocampal PE representations. There were no significant

correlations between the neural RDMs and the Sharpening hypothesis RDMs based on pool5 activations. Additionally, the correlation maps for Sharpening revealed overall concordance with our ROI analysis results (lOFA: p(FWEsvc) = 0.039 and trend in rOFA: p(FWEsvc) = 0.058, and p(FWEsvc) > 0.1 for all other ROIs (pFFA, aTL, MTG), based on pool4 for OFA and pool5 for all other ROIs). While the ROI and searchlight results overlapped, the strongest effects in the whole-brain searchlight analyses were observed in the angular gyrus, insula, and precentral gyrus for PE, and in the frontal pole for Sharpening (similar to refs. [74–76]).

Next, we compared the searchlight results based on PE and Sharpening. There were higher correlations between the neural RDMs and the hypothesis PE model than the Sharpening model (pool4) in the right AnG (p(FWE) = 0.01, [44 −46 22], k = 205), MFG (p(FWE) = 0.015, [28 30 28], k = 142), and putamen (p(FWE) = 0.048, [30 −16 2], k = 2). The difference correlation map based on pool5 revealed additional evidence for PE representations in the MTG as well as in frontal regions and bilateral SMG (Fig. 5b, Supplementary Table 19). There were no significant differences in correlations for the reverse contrasts 'Sharpening > PE' (pool4, pool5).

For comparison to the face-trained DCNN VGG-Face, we additionally investigated the searchlight results of the object-trained networks VGG-16 and ResNet50. Overall, searchlight analyses based on these networks showed a comparable distributed representation of PEs across the whole brain, mainly in parietal and frontal regions (Supplementary Figs. 2, 3, Supplementary Tables 20–28). None of these two networks revealed significant clusters for Sharpening.

Discussion

In this fMRI study, we provide evidence that prior context shapes the neural representation of presented faces. By means of our paradigm, in which participants learned to associate images of scenes with images of four face identities, we controlled for stimulus-related differences by presenting face morphs between two identities. On the behavioural level, we observed a facilitation effect, i.e., expected faces were identified faster compared to unexpected faces, and an assimilation effect, i.e., face morphs were more often classified as the expected identity. With univariate fMRI, we found reduced activation for expected compared to unexpected faces in the posterior FFA as well as in a more lateral cluster in the ITG/MTG. Crucially, multivariate fMRI RSA in combination with DCNNs revealed PE representations of presented faces along the whole face-processing hierarchy from OFA, pFFA, to the face-sensitive aTL, as well as in ITG/MTG. We found additional indications for sharpened representations at an earlier stage of the hierarchy in the OFA. Our results provide insights into the computational mechanisms underlying context-dependent stimulus representations along the face-processing hierarchy.

In our fMRI study, expectations for the upcoming faces were induced by images of indoor scenes that participants learned to associate. Faster identification of expected faces replicated and extended previous research on face perception that showed a facilitation effect for other context cues, such as names[8,13,16,77,78], identity cues[14], voice[9], or face primes[10–12,15,79]. When presented face information matches with prior expectations, judgements or identification is accelerated independent of the specific cue modality.

In addition, participants showed an assimilation effect and identified the presented ambiguous face morph more often as the expected than the unexpected face. This finding is in line with the facilitation of expected unmorphed face images and consistent with previous studies showing priming effects with non-face cues in the context of priming and associative learning[8,13,14,16,77,78], whereas contrastive perception[80] is typically observed in adaptation experiments that use long exposure to faces[17–19,81,82] (for review see refs. [83,84]). Hence, the observed assimilation effect in our study is based on the combination

of a short prior duration, a prior cue that is qualitatively different from face images, and a short target duration[85].

Through univariate fMRI, we observed reduced activation for expected compared to unexpected faces in a network involving parietal regions, midbrain regions, as well as bilateral anterior insula (for the contrasts 'mismatch > match' and 'unexpected > expected', respectively, see Supplementary Tables 1 and 2). This network has been repeatedly shown to be involved in the processing of surprise and error awareness[39,70,71,86]. In our study, in addition to surprise related to the unexpected face, this activation is also likely related to attention shifting and motor inhibition[87,88] as well as the internal verbalisation of the names associated with the faces[89]. Response times for recognising unexpected faces were notably longer than those for expected ones, evident in both the comparisons involving clear faces in the 'match vs. mismatch' and face morphs in the 'unexpected vs. expected' context. These prolonged response times suggest that the processing involved in identifying unexpected faces demands more time and effort. Consequently, the observed differences in univariate fMRI activity for the corresponding contrasts (as depicted in Fig. 4b) may be attributed to variations in effort or task difficulty rather than discrepancies in PEs or enhanced neural signals. Crucially, our multivariate analysis approach remains unaffected by this potential confounding factor. This is because we assessed expectation-dependent representations of faces in the face identification task, where participants were required to press one of four buttons with their right hand (corresponding to the index, middle, ring, or pinky finger) to identify the faces, and compared them with face representations from the neutral condition, where participants simply pressed the thumb of their left hand after viewing any face. This ensured that motor responses were controlled and did not introduce confounding influences into the observed patterns of similarity.

We specifically investigated the expectation-dependent univariate effect along the ventral face-processing hierarchy (OFA, pFFA, aTL) and observed reduced activation for expected faces in the pFFA, replicating previous reports of expectation suppression in FFA[23–26], and in more lateral clusters in the ITG and MTG. We used multivariate RSA to test whether these prior-dependent face activations along the ventral face-processing hierarchy are computationally explained by PE processing or sharpening of expected face information[4,7,30–33]. To do this, we compared the multivariate similarity of face morphs to neutral face images with the corresponding hypothetical similarity derived from PE and sharpened representations based on activations derived from DCNNs (Fig. 2a, Supplementary Fig. 4). We found evidence for PE processing at every stage of the processing hierarchy (OFA, pFFA, aTL), and in ITG/MTG (Fig. 4, Supplementary Fig. 1). The finding of PE representations in pFFA may explain the univariate expectation suppression effect in this area and rule out a predominant role of sharpened representations of the expected face. This extends previous univariate studies showing evidence for PE processing in FFA that did, however, not rule out Sharpening as an alternative model[23,90,91].

Furthermore, the PE searchlight analyses revealed a distributed network of occipital, parietal, temporal, and frontal brain regions highly similar to the activation networks observed for the univariate contrasts testing for increased signal for unexpected face information which is involved in the processing of surprise[39,70,71]. PEs in the IFG[69,92] and MFG are also in line with previous studies suggesting their involvement in face recognition[38,93] and the processing of facial features[94,95]. PEs in the fusiform gyrus extending into the parahippocampal gyrus may reflect the contextual association of linking the scene cues with the associated faces and names[96,97].

With the hypothesis model based on VGG-Face activations, we found evidence for sharpened representations of expected face information in OFA, an earlier stage of the face-processing hierarchy. This is in line with our recent finding of enhanced face representations for highly expected faces in OFA[39], suggesting that scene priors

sharpened the low-level facial features of associated faces in OFA, and with a study showing sharpening of prior information in Mooney face images along the whole ventral processing stream, already starting in early visual areas[31].

While in the ROI approach, we did not observe any evidence for Sharpening based on pool5 in the higher face-processing regions (pFFA, aTL), the searchlight analysis based on pool4 revealed further evidence for enhanced face representations in frontal areas, extending across the frontal pole, AC, and superior frontal gyrus. The sharpened representations in these frontal areas are in agreement with previous work showing top-down predictive face information in this region[74], but this searchlight cluster was not significantly stronger than a purely input-driven face representation. In contrast to expectation-independent sensory input, there were stronger sharpened face representations in the right hippocampus. The involvement of the hippocampus in expectation-dependent representations for both PEs as well as for prior confirming inputs have also been repeatedly observed during association learning[98] and application of these predictive associations[99,100].

Overall, the observed searchlight patterns were more extensive and stronger for PE than Sharpening and do not predominantly reflect the ventral face system, but extend to frontal and parietal regions. This network of regions has also previously been observed in studies investigating familiar face recognition[38,75,76]. Potentially, these dorsal regions may play a crucial role in representing familiar faces. As participants in our study acquired the association between face images and semantically distinct scene images (e.g., a library or fitness court), it is plausible that they attributed semantic meaning to these face images beyond mere visual representations.

Interestingly, although overall participants more often identified the presented face morph as the expected face, we observed more evidence for PEs than sharpened representations of expected facial features. This observation stands in contrast to the intuition that sharpened representations likely occur when facial input aligns with expectations, whereas PE becomes more prominent when facial input deviates significantly from expectations. In our paradigm, participants most likely noticed the deviation as indicated by slower RTs for 'expected' responses to face morphs than to clear faces in the match condition. Future work may help to determine how universal the dominance of PE over sharpened representations for partially expected faces is. For example, observations of PEs may be reduced in a paradigm in which the presented ambiguous face deviates less from the expected face.

Our initial computational models underlying the hypothesis RDMs used for testing whether representations of face morphs can be explained by the reduction or enhancement of the expected information were based on the DCNN VGG-Face[63] since the convolutional layers of this face-recognition network correspond to the hierarchically organised regions of face processing in the human brain[59,60] and also predicted face dissimilarity judgements[101]. We decided to use pool5 to test face representations in the high-level aTL because these layer activations correspond to the highest level of sensory face processing in the VGG-Face network, in contrast to activations from the highest connected layer (fc8) which activations rather reflect decision about choosing one of the 2.622 identities that this network was trained on. In addition, we used two DCNNs, i.e., VGG-16 and ResNet50, that were originally trained on object recognition. Therefore, our study contributes to the growing research investigating the correspondence of neural network activations to neural activations in the human brain. Additionally, as considering inter-individual differences in RSA analyses is crucial[30,68,69], we used individually weighted hypothesis RDMs by incorporating the behavioural responses. Specifically, by contrasting prior-confirming with prior-discarding responses for a face morph, we were able to capture individual perceptual dominance of one identity in a face morph that remained despite the individual face-

morph calibration on the first experimental day. By including these individual prior weights, we took into account that individuals may differ in their usage of prior information.

Our decision to leverage DCNNs as sophisticated hierarchical computational models for studying expectation-dependent face representations in the human brain was motivated by growing evidence supporting their alignment with the neural face-recognition systems observed in both humans and nonhuman primates[59–62,102]. Specifically, a recent intracranial electroencephalography study successfully related the layers pool4 and pool5 of VGG-Face to single neuronal recordings from OFA and FFA, respectively[60]. While also other methodological approaches, such as MEG[59] and fMRI[61], successfully related DCNN layer activations to brain activations, further research is needed to test whether the relationship between DCNNs and brain representations is readily applicable to more coarse neuronal representations such as the voxel-level resolution obtained with fMRI. In addition, there are limits in correspondence and fundamental differences in how the brain and DCNNs represent visual information[103]. Biological face recognition is far more complex than image labelling and involves objectives beyond physical properties and, likely, DCNNs do not capture several functional properties of face recognition (for a review see ref. [62]). However, applying RSA based on DCNNs revealed stronger evidence than a model-free classification approach (Supplementary Fig. 5, Supplementary Methods, Supplementary Results).

By comparing VGG-Face to VGG-16, a DCNN with the identical architecture that was however trained on object recognition instead of face images[66], as well as to ResNet50, a more complex convolutional neural net with deeper architecture and skip connections[67], we observed commonalities as well as differences between these networks. Across all networks, the correlations between voxel- and network-based face representations were low, similar to other studies reporting significant but small correlations between face-selective brain areas and face-identification models based on their representational similarity[47,62]. Notably, PE was more dominant than sharpened representations in both ROI as well as searchlight analyses across all three DCNNs. Sharpened face representations in OFA were only observed based on VGG-Face and not based on the object-trained networks. However, consistent with prior findings that DCNN models trained on ImageNet demonstrate comparable or superior performance compared to models specifically trained for faces in predicting human neural responses to facial stimuli (see supplementary material of ref. [60] and the work of ref. [61]), our study revealed higher correlations between voxel-based similarity and PE similarity patterns when using VGG-16 compared to VGG-Face. Thus, our results suggest that the features extracted from VGG-16 can effectively form a representational space suitable for capturing the static facial images employed in our study. In sum, the incorporation of different DCNNs substantiates the PE hypothesis across all face-sensitive regions, with the superior performance of object-trained models, but raises uncertainties regarding Sharpening that was only observed based on VGG-Face. This incongruity across DCNNs underscores the critical importance of a careful model selection and comparison, as the choice of DCNN can significantly impact the interpretation of underlying neural representations in the human brain and may lead to different conclusions. Further research is needed to establish whether the observed pattern, wherein a face-trained DCNN also exhibits alignment with neural representations of expected facial features, while object-trained DCNNs align more strongly with neural representations of unexpected facial features in the human brain, can be extrapolated to other datasets.

Our additional ROI analyses based on VGG-Face investigating potential hemispheric differences in face representations suggested higher PE-based face representations in the left compared to the right pFFA and aTL (see Fig. 4h, i; although no main effect of hemisphere;

see Supplementary Results). This left lateralisation is in concordance with a previous meta-analysis and study showing left hemispheric aTL activation for familiar individuals, while right aTL was mainly involved in novel faces[104]. However, other studies pointed towards face processing as a predominantly right hemispheric process[105,106]. Indications for left lateralisation in our study may be related to the computational modelling based on the VGG-Face network that previously captured dissimilarity representations only in left hemispheric OFA and FFA[60]. However, in the respective paper, due to the smaller number of right hemispheric intracranial electrodes, analyses were solely based on left hemispheric electrodes. The left lateralisation was not prominently evident in the whole-brain searchlight analyses based on VGG-Face and VGG-16 (Fig. 5, Supplementary Fig. 2). Also, our additional ROI analysis based on VGG-16 did not show this left hemispheric dominance (Fig. 4g–j), whereas the overall weaker results based on ResNet50 indicated stronger effects in the left hemisphere (Supplementary Fig. 1). Therefore, we do not draw strong conclusions about any hemispheric differences in expectation-dependent face representations. Future research is needed to investigate whether other layers of VGG-Face or other neural network architectures would have a higher correspondence to right hemispheric face representations.

Our study exhibits typical characteristics of multivariate fMRI analyses focused on individual stimuli, including a relatively low noise ceiling and modest effect sizes. The maximum possible correlation values that could be observed in our fMRI data from the face-sensitive ROIs are all considerably smaller than 1 (Fig. 4, Supplementary Fig. 1), underscoring inherent constraints in our experimental data. These constraints may arise from factors such as limited spatial resolution, substantial measurement noise, or a shortage of data. In addition, these small effect sizes could be attributed to the noise added to the presented face images, potentially impeding clarity. Although this was intended to encourage the use of the prior, it might have inadvertently reduced neural responses. However, it is important to note that these limitations do not introduce differential effects among our experimental conditions. Consequently, measurement noise and other extraneous variables cannot account for the observed similarity effects within the multivariate analyses. For RSA, similar noise ceilings and correlation values between fMRI-response-based and hypothesis RDMs have been observed previously[4,39,47,102]. Analogously, low classification accuracies are also common in decoding task events using multivariate classification of fMRI data[52,69,75,76,107]. Despite these inherent limitations, distinctions in the observed correlations, particularly variations in the degree of similarity between expected and unexpected facial stimuli, provide evidence for the presence of expectation-dependent multivoxel representations.

Our findings from both the ROI and the searchlight approach point to the co-existence of representations of the unexpected as well as the expected information contained in images of morphed faces across the face-processing hierarchy, suggesting that different computational mechanisms may be simultaneously applied to combine priors with sensory input. Within the predictive coding framework[1,2,6], this could be interpreted as evidence for the co-existence of error units as well as representational units containing the updated face prior. Previous research has also suggested the co-existence of both unit types by identifying voxels that showed prediction or error processing consistently over time[22,108]. Future research using a higher spatial resolution (e.g., 7 T) will enable us to differentiate whether the co-existence of PE and Sharpening in OFA is linked to different types of cortical layers[109], with superficial layers containing bottom-up and deeper layers top-down information[110,111].

In conclusion, we used multivariate fMRI analysis combined with computational modelling based on the activations of DCNNs to investigate prior-dependent face representations along the ventral face-processing hierarchy. These analyses revealed PE processing throughout the entire face-processing hierarchy, as well as sharpened representations of expected faces based on a face-trained network at an early stage of processing. The observed PE and sharpened representations provide evidence for predictive processing, through which the brain combines prior knowledge with sensory input, thereby influencing our perception of faces.

Methods
This study was preregistered at the Open Science Framework (OSF) ([https://osf.io/sd54e]).

Participants
We preregistered to schedule 50 participants for this fMRI study. Seven participants were excluded from final data analyses: one due to technical issues, one due to anatomical anomalies, one due to extensive head movements, three did not take part in all study appointments, and one was identified as an outlier in the behavioural experiment analysis (see Supplementary Methods). In the final sample, 43 right-handed participants (22 females, self-reported gender) with a mean age of 24.37 years ($SD = 3.61$) and no current or past neurological or psychiatric disorders were included. Compensation for participation was 55 €. All experimental procedures were approved by the Ethics Committee of the Chamber of Physicians in Hamburg and participants provided written informed consent.

Stimuli
In this study, we used images of faces and scenes.

Specifically, we used images of four male faces that were created with FaceGen (FaceGen Modeller Core 3.22, Singular Inversion). The four face identities were created so that they differed in the facial features that are important for face discrimination[112]: shape, colour, and positioning of the eyes, eyebrows, nose, and mouth. Images were normalised for their general face shape so that they only differed in their central facial features. To ensure that the four faces were equally distinct and well-distinguishable, the activations of layer pool4 of the DCNN VGG-Face were used to evaluate their dissimilarity structure (Supplementary Fig. 6)[60]. All face images were normalised by independently equalising the mean luminance and standard deviation of the RGB channels. Noise was added to the face images to decrease the clarity of the sensory input and hence increase the usage of the prior information. The noise was added by applying Fourier transformation and adding a random phase structure to its original phase spectrum. After combining it with the original amplitude spectrum, an inverse Fourier transformation was performed. For each face image presentation (e.g., for each repetition of the image of Ari), a new random phase structure was applied, i.e., all presented face images had a unique noise pattern.

We used nine scene images to provide prior context. For the training and the main experiment, five indoor scenes were chosen as context primes for the four face images: four images were taken from the SUN database[113] and the fifth scene from the indoor scene database[114]. For the functional localiser, four additional indoor scenes were selected[113]. Scene images were converted to grey-scale and luminance-matched using the SHINE toolbox' histMatch-function[115]. We used grey-scaled scene images to avoid any colour confounds on the perception of the following face image. For further image specifications, please refer to the Supplementary Methods.

Experimental procedure
Participants came to the lab on two consecutive days. On the first day, they completed the individual face-morph calibration to identify each individual's personal morphs that equalled their 50/50 perceptual threshold so that both identities were equally likely to be seen in a morph (Supplementary Fig. 7, Supplementary Methods). Afterwards, participants took part in a training session in which they learned to associate each face with a scene. For a complete

experimental protocol of the training sessions, please refer to the Supplementary Methods.

On the second day, participants completed the fMRI experiment which was divided into four blocks. Each block was conceptually identical to the last part of the association training session and consisted of 107 experimental trials (16 match, 48 partial, 12 mismatch, 12 catch, 16 neutral, 3 neutral catch) and 36 null events. In match trials, the presented face was preceded by the associated scene. In mismatch trials, the presented face differed from the expected face. In partial trials, face morphs of two identities were presented. These face morphs always contained the expected face identity (that matched the preceding scene) as well as an unexpected face identity. The task was to indicate the face identity (Ari, Bob, Cid, Dan) by pressing one of four buttons with the right hand (index, middle, ring, pinky finger). In catch trials (question mark instead of face), participants were required to indicate which face they expected based on the preceding scene. In neutral trials (indicated by a fifth scene), there was an equal probability for each of the four face identities to occur. If a face appeared after the neutral scene, participants had to press a button with their left thumb for any face. In neutral catch trials (question mark instead of face), participants had to press a button with their left index finger to indicate that they had anticipated all faces with equal probabilities. An exemplary trial can be seen in Fig. 1b. In null event trials, a fixation cross was presented for the duration of a whole trial (5300 ms). The ratio of trials per condition (match, partial, mismatch, catch, neutral, neutral catch, null events) was identical in all four blocks and identical to the last part of the training sessions. The order of the trials was pseudo-randomised such that the same face or face morph was allowed to consecutively appear four times at maximum. This randomisation limitation was selected so that participants could not easily foresee which face was likely (or not likely) to appear next. Only two null event trials could appear consecutively after each other to avoid too long periods of fixation crosses. After each block (~12 min), short verbal feedback was given to keep the motivational and attentional level high for the whole duration of the experiment (~53 min, more details in Supplementary Methods).

Functional localiser

A functional localiser experiment was run to identify individual ROIs along the ventral face-processing hierarchy, i.e., the OFA, the FFA, and the higher-level face-sensitive region in the aTL[40–42,44]. The design was similar to established localiser paradigms[9,116,117]. Alternating blocks of face and scene images and neutral blocks with a fixation cross were shown. In the face blocks, the images of the known four faces (Ari, Bob, Cid, Dan) were presented. In the scene blocks, four unknown scenes were displayed. New scenes were chosen because participants had learned to associate each scene with one of the four faces. Therefore, the presentation of these scene images could have automatically triggered unwanted activation due to the recall of the associated faces. In each block, 44 images each with a duration of 500 ms were presented. There was no ISI between the images. Each block had a duration of 22 s. The task was to look at the fixation cross in the centre of the screen, no buttons had to be pressed. The order of the images within a block was pseudo-randomised such that the same image could not appear twice after each other. Due to the missing ISI, multiple consecutive presentations of the same image would have led to seemingly prolonged presentation durations. The starting block (faces or scenes) was counterbalanced across participants.

Behavioural data analysis

Analyses were performed as preregistered and additional analyses are described below. Since values of perceived face identity in face morphs, RTs, as well as accuracies were not normally distributed (Kolmogorov–Smirnov tests, all $p < 0.001$), non-parametric tests were used for the analyses instead of the preregistered parametric tests.

Our first variable of interest was the perceived face identity. In partial trials, participants answered which person they mostly recognised in a face morph. To investigate whether participants identified face morphs more often as the expected or the unexpected identity, a difference score was calculated for each face pair to indicate how likely the participant answered in favour of the prior. The mean of the difference scores of all scene and morph combinations was calculated to obtain an individual index for an assimilation and/or contrastive effect. Values above 50% indicated that a participant responded more often in favour of the expected face identity in a face morph (assimilation effect). Values below 50% were indicative of a contrastive effect. We tested whether the participants' scores significantly differed from 50% (no prior effect) using a two-sided Wilcoxon signed rank test and calculating Wilcoxon's r as a measure of effect size.

RTs were measured for the time point of a button press after face onset. Additionally to the preregistered conditions mismatch, match, and neutral, we included the partial condition because we were also interested in how fast participants processed face morphs. We calculated a non-parametric Friedman test and Kendall's W as effect size. Post-hoc paired tests between the average ranks of the different conditions were performed using Tukey's honestly significant difference (HSD) test for multiple comparisons. In an exploratory analysis, we investigated whether the RTs to the morphed faces in partial trials depended on the response given by the participants. Therefore, partial trials were split into trials with prior-confirming responses (assimilation effect) and trials with responses favouring the other identity contained in a morph (contrastive effect) and tested with a two-sided paired Wilcoxon signed rank test. Wilcoxon's r was calculated as a measure of effect size. Lastly, we tested whether RTs in trials with prior-confirming responses differed from RTs in the match condition using a two-sided paired Wilcoxon signed rank test, calculating Wilcoxon's r as a measurement of effect size.

fMRI data acquisition and preprocessing

All imaging data were acquired on a Siemens 3T scanner at the University Medical Center Hamburg-Eppendorf (Hamburg, Germany) with a 64-channel head coil. Functional data were obtained using a multi-band echo-planar imaging sequence (repetition time (TR) = 0.961 s, echo time (TE) = 30 ms, flip angle = 55°, field of view (FoV) = 224 mm, multi-band mode, number of bands: 3). Each volume of the experimental data contained 45 slices (voxel size $2 \times 2 \times 2$ mm plus 0.5 mm gap) and were obtained in descending order.

The parameters for the functional data were chosen to maximise the signal strength in the aTL. Due to its location near the sphenoidal sinuses (i.e., near air/tissue and bone/tissue interfaces), susceptibility artefacts can lead to a poor signal-to-noise ratio (SNR)[118,119]. We followed the proposed guidelines[119] to maximise our SNR in the aTL by choosing a short TR (<1000 ms), a voxel size of $2 \times 2 \times 2$ mm, and covering additional 'no-brain' space below the temporal lobe with our FoV (so that the aTL was not at the edge of the FoV).

An additional structural image (magnetisation prepared rapid acquisition gradient echo (MPRAGE)) was acquired for functional preprocessing and anatomical overlay (TR = 7.1 ms, TE = 2.98 ms, flip angle = 9°, FoV = 256 mm, 240 slices, voxel size $1 \times 1 \times 1$ mm, ascending order).

A fieldmap was acquired for field inhomogeneity corrections (TR = 495 ms, TE1 = 5.51 ms, TE2 = 7.97 ms, flip angle = 40°, FoV = 224 mm, 45 slices (voxel size $3 \times 3 \times 2$ mm plus 0.5 mm gap)). The slices were obtained in an interleaved order. The protocols with scanning parameters are available here: [https://osf.io/765jx/].

Structural and functional data were analysed using SPM12 and custom scripts in MATLAB. First, the functional images of all functional runs were realigned to the mean functional image. We then applied field mapping distortion correction to the functional volumes to correct for geometric distortions in EPI caused by magnetic field

inhomogeneity (with the FieldMap toolbox). The individual structural T1 image was co-registered to the mean, distortion-corrected functional image. The functional images were spatially normalised to MNI space. For the univariate analysis, the functional images were additionally smoothed with an 8-mm full-width at half maximum isotropic Gaussian kernel.

Univariate fMRI analysis

Data from the four functional runs were analysed using the general linear model (GLM) with a 128 s high pass filter. We applied SPM's alternative pre-whitening method to account for autocorrelation, FAST, which has been suggested to perform better than SPM's default[120]. Raw motion parameters (three translations and three rotations) were included as regressors of nuisance. This approach was also used for the multivariate analyses (see below).

For the four runs of the main experiment, onsets of ten events were modelled as separate regressors in the GLM, each convolved with the canonical SPM haemodynamic response. The first regressor was for the scenes that were presented at the start of each trial. We further specified six face regressors for the different conditions: neutral, match, mismatch, and partial. While neutral, match, and mismatch had one regressor each, the partial face onsets were divided into three regressors: in the first partial regressor, we included trials in which participants answered to have perceived the expected face identity (expected), in the second regressor we included trials in which they answered to have perceived the unexpected face identity within the morph (unexpected), and the third regressor consisted of onsets of partial trials in which participant either answered to have perceived an identity which was not within a morph or answered too slowly. We included three regressors of no interest, one for catch trials, one for button responses, and one for feedback. If a participant did not receive any feedback and/or never incorrectly identified a partial trial as an identity not contained in a morph or answered it too slowly, dummy onsets were defined. At the end of each run, we presented a fixation cross for 10 s to capture the haemodynamic response function of the last trial.

On the second level, we computed the 'mismatch > match' and the 'unexpected > expected' contrasts. For the whole-brain analyses, we report cluster activations (p(FWE) < 0.05, with cluster-inducing threshold of $p < 0.001$). For the small-volume corrected analyses of our ROIs (OFA, pFFA, aTL), we report peak activations (p(FWEsvc < 0.05)).

A functional localiser was run at the end of the experiment to define ROIs along the face-processing hierarchy. The GLM included two event types, each convolved with the canonical hemodynamic response function. The event types were the onsets of the face and the scene blocks. For the first-level analyses of the main experiment and the functional localiser, individual whole-brain masks were used (see Supplementary Methods). On the second level, we computed the contrast 'faces > scenes' to obtain the ROIs (see below for further details).

Regions of interest (ROI) extraction

We defined ROIs along the ventral face-processing stream (OFA, FFA, and aTL). As previous studies on face perception and/or face identification in humans and macaques revealed a contribution of right hemispheric[105,106,121,122] as well as bilateral[51,56,123] brain areas, we defined bilateral ROIs. We extracted the ROIs from SPM12 using MarsBaR[124]. The functional localiser, using the contrast 'faces > scenes', yielded bilateral activation clusters spanning from the inferior occipital gyrus (IOG) to the fusiform gyrus (Supplementary Table 29). Since a clear separation of these clusters into OFA and FFA was not possible, we overlaid our activation clusters with the OFA and pFFA clusters from an atlas map[125]. We obtained OFA ROIs in the right ($k = 892$, peak at [54 −70 −4]) and left hemisphere ($k = 483$, peak at [−50 −76 −8]) as well as pFFA ROIs in the right ($k = 848$, peak at [44 −46 −18]) and left hemisphere ($k = 477$, peak at [−44 −52 −20]). Previous literature

suggested a differentiation into a posterior and anterior part of the FFA[41,126,127]. When comparing our activation cluster with pFFA and aFFA clusters[125], we only found an overlay with the posterior part. The peak activations of our pFFA clusters are also comparable to the area FFA-2[126].

We obtained face-sensitive ROIs in the aTL from the functional localiser 'faces > scenes' in the right ($k = 192$, peak at [34 −8 −38]) and left hemisphere ($k = 153$, peak at [−40 −20 −38]) at p(unc.) < 0.01, as the clusters at $p < 0.001$ were too small with $k = 54$ and $k = 5$, respectively. These peak activations are close to previously reported face-selective regions in the temporal pole[39,43,126,128].

In addition to our main ROIs (OFA, pFFA, aTL), we extracted ROIs along the ventral face-processing hierarchy from our univariate contrast 'mismatch > match' to investigate with our multivariate analyses whether this expectation suppression effect might be due to PE processing or sharpened representations. The contrast revealed a lateral cluster in the left ITG and MTG ($k = 312$, peak at [−56 −42 −18]; at p(FWE) < 0.05 (cluster-corrected), based on a cluster-inducing threshold p(unc.) < 0.001; Supplementary Table 1) and in the right MTG with a comparable size to the left hemisphere ($k = 332$, peak at [58 −34 −16]; at p(unc.) < 0.01). These clusters identified based on the contrast 'mismatch > match' are independent of the RSA which is based on neutral and partial trials.

All bilateral ROIs (OFA, pFFA, aTL, and ITG/MTG) were transformed from MNI space into individual native spaces using the inversion matrices from SPM12's normalise-function.

Face-trained deep neural network

The DCNN VGG-Face, available at [www.robots.ox.ac.uk/~vgg/software/vgg_face/][63], was pre-trained to recognise 2.622 different face identities using a database containing 2.6 million face images. This model achieved a state-of-the-art performance level while using less data compared to other advanced models like DeepFace and FaceNet. This network performs best compared to numerous other neural networks in predicting humans' face dissimilarity judgements[101]. The network architecture of VGG-Face includes a total of 16 layers, consisting of 13 convolutional layers and 3 fully connected layers. A rectification linear unit follows each of these 16 layers. The 13 convolutional layers are organised into five blocks, with the first two blocks containing two consecutive convolutional layers followed by max pooling. The latter three blocks consist of three consecutive layers followed by max pooling. In DCNNs like VGG-Face, layers closer to the input layer capture lower-level facial features such as edges, textures, and local facial details, while higher layers in the network learn more complex and informative facial representations such as gender, age, and identity information[59]. We extracted face activations from the last two max pooling steps, i.e., pool4 and pool5, to design our hypothesis RDMs (see more detail below). These layers can be described as intermediate to higher layers in VGG-Face (see Fig. 2a for hierarchical model architecture), with pool5 located directly before the final three fully connected layers that lead to a classification of the input image as one of the face identities it was trained on[63]. The representational space of pool4 and pool5 activations is robust against low-level manipulations such as luminance and colour and has been previously related to our brain regions of interest, i.e., pool4 to lower-level inferior occipital gyrus and pool5 to higher-level face processing in fusiform gyrus[60].

Object-trained deep neural networks

Previous literature has shown that even though DCNNs like VGG-Face show correspondences to the single-cell and voxel-level representational space of face processing[60,61], DCNNs trained on object recognition can perform similarly (Supplementary Material of ref. 60) or even outperform them in the context of face processing[61]. Therefore, in addition to our preregistered approach to employ VGG-Face, we tested

two object-trained DCNNs for comparison with VGG-Face: firstly, we chose VGG-16[66], a convolutional network with the identical architecture as VGG-Face, i.e., consisting of 16 layers, but pre-trained on the ImageNet dataset[129]. From VGG-16, in agreement with our approach based on VGG-Face, we chose activations of layer pool4 for the hypothesis RDMs for OFA and layer pool5 for all higher ROIs. Secondly, we selected the DCNN ResNet50[67] because this network performed best across a large variety of tested networks in predicting neural responses to faces in a recent fMRI study[61]. For our hypothesis RDMs to test representations in all our ROIs (OFA, pFFA, aTL, MTG), we extracted face activations from the convolutional layer res5b_branch2b (MATLAB) because this layer best predicted neural responses in FFA[61].

Representational similarity analysis: computational modelling based on VGG-Face activations

To investigate whether PE or Sharpening mechanisms underlie the integration of expected and presented face information, we used RSA[64,65].

RSA involves defining theoretical dissimilarity matrices (i.e., hypothesis RDMs) between experimental conditions and comparing them to neural dissimilarity matrices (i.e., neural RDMs) based on the measured brain activation. By defining different theoretical models and comparing their correlation values with the neural data, we can test which of the hypothetical models fits the data best. The multivariate analyses were performed on realigned data in the individual's native space. A first-level analysis using a whole-brain mask was performed for each participant. Onsets of 25 events were modelled as separate regressors in the GLM, each convolved with the canonical SPM haemodynamic response. Four regressors were for the neutral trials differing by which face was presented after the neutral scene (neutral$_A$, neutral$_B$, neutral$_C$, neutral$_D$). Four regressors were for the match trials in which the presented face matched the expected face (match$_A$, match$_B$, match$_C$, match$_D$). Twelve regressors were for the partial trials, each for one combination of prior and presented face morph (e.g., A$_{prior}$AB$_{input}$, A$_{prior}$AC$_{input}$, A$_{prior}$AD$_{input}$, …, D$_{prior}$CD$_{input}$). Five regressors of no interest were for scenes, mismatch trials, catch trials, button responses, and presented feedback. In case no feedback was given in a run, a dummy onset was defined. For the multivariate analyses, we used T-images instead of beta estimates as in our previous studies[4,39] because due to the additional division of the beta values by the standard error estimates the influence of noisy single voxels can be reduced[130].

In this study, we defined three hypothesis RDMs to test how presented faces are represented depending on prior context. The two main hypothesis models were a (1) PE and a (2) Sharpening model (Fig. 2b). These models differ in how the prior and the input are mathematically combined. The third hypothesis model tested was a (3) pure Sensory Input model that only takes the visual properties of the face image into account without considering any influence of the prior (Fig. 2b).

To test our main research question about how the information of the prior is combined with the incoming face information, we used the partial trials in which the presented face contained the expected as well as an unexpected face part. By comparing the activation patterns of the partial trials with the 'pure' face representations measured in the neutral trials, we aimed at differentiating whether the representation observed for a face morph was more similar to the unexpected face part (i.e., PE processing) or more similar to the expected face part (i.e., Sharpening). We designed and chose the neutral trials to extract the pure face representations instead of the match trials for two important reasons: firstly, the neutral scene was not predictive of the upcoming face, therefore, the measured activation for the face was independent of prior information while in match trials the face expectation was confirmed. Secondly, the motor response required in neutral and

partial trials was different and therefore did not confound the RSA. While in partial trials participants were required to indicate which person they mostly recognised in a face by pressing one of four buttons with the right index, middle, ring, and pinky finger, their task in neutral trials indicated by the fifth scene was to press a button with the left thumb for whichever face appeared.

All three hypothesis RDMs, i.e., the PE, Sharpening, and Sensory Input hypothesis model, were based on the neural network activations of the DCNN VGG-Face[63] for both the expected as well as the presented faces. The dissimilarity structure of activations of the network's layers pool4 and pool5 for different face images significantly correlates with the representational dissimilarity structure of neural activations measured from electrodes in the human OFA and FFA, respectively[60]. To measure neural representations in the face-processing hierarchy (OFA, pFFA, aTL), we created the hypothesis RDMs based on the network activation extracted from lower-level layer pool4 for bilateral OFA and from higher-layer pool5 for all higher face-sensitive areas (pFFA, aTL, ITG/MTG clusters of 'mismatch > match'). Searchlight analyses were performed with both pool4 and pool5 activations.

The neural network activations were read out for the RDM creation as follows: in the main experiment, each participant saw each of the four faces (i.e., Ari, Bob, Cid, Dan) in the neutral condition. These four images were fed into the VGG-Face network to extract their activation vectors from layers pool4 and pool5. For the 12 partial conditions, we combined the network activations for the priors and the face morphs. For the prior activations, we used the corresponding four unmorphed face images weighted with the individual behaviour to account for the prior usage during the perceptual decision about which face was identified in a face morph (see below for further explanation). These weighted face images were fed into the VGG-Face network and read out at layers pool4 and pool5 to obtain the prior activation vectors. To obtain the morph activations, the six 50/50% morph images (AB, AC, AD, BC, BD, CD) were fed into the network and their activations were extracted at layers pool4 and pool5. This procedure resulted in prior activation vectors (pool4, pool5) and face morph activation vectors (pool4, pool5) which were differentially combined for the PE and Sharpening hypothesis RDMs (described below).

Prediction error model. For the calculation of the PE model, the individually weighted prior representation (i.e., precision) was subtracted from the input representation[7]:

$$\mathbf{PE} = \mathbf{morph} - (\mathbf{prior}.^* precision_{prior\,for\,morph}) \qquad (1)$$

For example, in trials in which the scene predictive for Ari preceded a face morph between Ari and Bob, this equation would translate into:

$$\mathbf{PE}(\mathbf{A}_{prior}\mathbf{AB}_{input}) = \mathbf{AB}_{input} - (\mathbf{Aprior}.^* precision_{prior\,for\,AB}) \qquad (2)$$

The precision of the prior was used to account for the individual prior usage during the perceptual decision which face participants identified in a face morph. The prior precision was calculated as follows:

$$precision_{prior\,for\,morph} = (n_{prior} - n_{otherpart})/n \qquad (3)$$

In detail, n_{prior} refers to the number of responses in favour of the expected face, while $n_{otherpart}$ refers to the number of responses in favour of the unexpected face in a face morph. N refers to the total number of trials in which the participant answered to have perceived the expected or the unexpected part in a face morph ($n_{prior} + n_{otherpart}$), i.e., we did not include trials in which participants identified a face that was not contained in a morph or were too slow. This calculation can result in precision values in the range of [−1 1]. If a participant always answered to have perceived the expected face in the morph AB

(irrespective of whether the prior was A or B), this would translate into a value of 1, therefore, giving a high weight to the prior. If a participant always answered to have perceived the unexpected face in the morph AB, this would translate into a value of −1, therefore, giving a highly negative weight to the prior that could lead to contrastive effects. The distance of all of the experimental conditions (neutral, partial) for the hypothesis RDM was calculated using '1 - Pearson Correlation'[64,65]. The RDM was rescaled to dissimilarity values between 0 and 1 while considering shared ranks (equal ranks stayed equal) (Fig. 2b). The same correlation metric and ranking were used for the Sharpening and Sensory Input RDMs.

Sharpening model. An alternative approach for how the Bayesian brain may combine priors/expectations with incoming sensory information is the multiplication of predictions and inputs[4,7]. We translated this sharpening of expected information into the following equation:

$$\text{Sharpening} = \log(\textbf{morph}.^{*}(\textbf{1} + \textbf{prior}.^{*}\ precision_{prior\,for\,morph})) \quad (4)$$

For example, in trials in which the scene predictive of Ari preceded a face morph between Ari and Bob, this would translate into:

$$\text{Sharpening}(\text{A}_{prior}\text{AB}_{input}) = \log(\textbf{AB}.^{*}(\textbf{1} + \text{A}_{prior}.^{*}\ precision_{prior\,for\,AB})) \quad (5)$$

Furthermore, '1+' was added to the prior to account for the case in which the prior had no effect on the perception of a morph, so that the face morph is treated as the sole basis of the measured information. Since DCNNs can have positive or negative activations, we extend the traditional Sharpening model to deal with cases of negative priors or inputs: when the layer activations of both input and prior have the identical sign, i.e., both are positive or both are negative, the sign of the input activations is preserved after sharpening, i.e., expected positive activations are sharpened to be "more positive" and negative activations are sharpened to be "more negative". On the other hand, when the activations of input and prior have opposite signs (i.e., one is positive and the other is negative), the input activations are dampened rather than sharpened while keeping the sign of the input activation. Dampening is achieved by multiplying the input activation with a number between 0 and 1. Specifically, the input activation is multiplied with (**1** - abs(**prior**.* $precision_{prior\,for\,morph}$)) for these cases where the prior is rescaled to be in the range −1 to 1 (which is necessary so that **1** - **prior** does not get negative). Finally, we applied a log transformation on the combined prior and morph activation to account for extraordinarily high values inherent to the multiplication of large activation numbers.

Sensory input model. We created pure Sensory Input hypothesis models to test whether a model without the combination of prior and input information would perform better than the PE or Sharpening model (Fig. 2b). For the neutral trials, the pool4 and pool5 activation vectors were created as for the other hypothesis models. For the partial trials, the activation vectors for the face morph images were taken without combining them with the prior. For instance, the activation of the morph image between Ari and Bob was extracted from the network, irrespective of the preceding scene.

The RDMs for our hypothesis models (PE, Sharpening, Sensory Input) for the object-trained DCNNs (VGG-16, ResNet50) can be found in the Supplementary Fig. 4.

Representational similarity analysis: ROI analyses
The multivariate ROI analyses were performed using the RSAtoolbox[65] in Python 3.9.12. Individual grey matter masks in native space with a threshold of zero were applied. We calculated the neural RDM for each ROI and averaged their right and left hemispheric neural RDMs to get an estimate of the mean neural representational space across hemispheres.

For each ROI, we obtained one Kendall's Tau A correlation coefficient for each participant and hypothesis RDM. We chose Kendall's Tau A as the appropriate correlation measurement for tied ranks[65]. Since correlation values for the different models (PE, Sharpening, Sensory Input) and ROIs were not normally distributed (Kolmogorov–Smirnov tests, $p < 0.001$), we used non-parametric tests to test for significance. For each model, we tested against zero using a one-sided Wilcoxon signed rank test, Bonferroni-corrected for the number of tests per ROI (for VGG-Face vs. VGG-16: $N = 6$ (3 models × 2 DCNNs), see Fig. 4; for ResNet50: $N = 3$, see Supplementary Fig. 1). For model comparisons, we used two-sided paired Wilcoxon signed rank tests, FDR-corrected[131] for the model comparisons per ROI (all model comparisons within each DCNN and within model comparisons across the DCNNs). We additionally calculated the lower and upper boundary of the noise ceiling for each ROI with the RSAtoolbox[62] to obtain an estimate of how well any model could perform given the noise in the data. For the calculation of the noise ceiling, we made sure to only consider the relevant dissimilarities in the neural RDMs corresponding to the hypothesis models (4 neutral conditions × 12 partial conditions). Additional analyses for the left and right hemispheres can be found in the Supplementary Results (Fig. 4g–j, Supplementary Fig. 1, Supplementary Tables 3–13).

Representational similarity analysis: searchlight analyses
To explore the representations beyond the prespecified ROIs, a multivariate searchlight was applied within the whole brain and the same analyses as in the ROI approach were computed. The searchlight analyses were performed in native space using a grey matter mask and a sphere with a radius of 6 mm, containing a maximum of 90 voxels and a minimum of 10% valid voxels. The resulting correlation maps were Fisher's z-transformed, normalised, and smoothed with an 8-mm full-width at half maximum isotropic Gaussian kernel. These maps were tested in a one-sample t-test on the second-level and significant results are reported at $p(\text{FWE}) < 0.05$, except for the ResNet50 searchlight results which are reported at $p(\text{FWE}) < 0.05$ (cluster-corrected), with a cluster-inducing threshold of $p(\text{unc.}) < 0.001$. Additionally, to compare the results of the searchlight analyses for the hypothesis PE and the Sharpening models, we calculated difference correlation maps on the individual participant level and conducted second-level one-sample t-tests across the participants[65].

Lastly, to specifically investigate expectation-dependent face information and potentially control for the representation of low-level visual information, we computed difference correlation maps between the PE and Sharpening searchlight maps and sensory input searchlight maps. For VGG-Face and VGG-16, the sensory input searchlight maps were based on the second convolutional layer (conv1_2)[59] of the respective network. For ResNet50, the sensory input map was based on the same layer activations as the PE and Sharpening RDMs (res5b_branch2b). The resulting searchlight maps may be indicative of expectation-dependent face information, extending beyond mere visual representations.

Multivariate classification analysis
In addition to the preregistered RSA, we conducted a simpler multivariate classification approach without model-based hypothesis RDMs (Supplementary Fig. 5, Supplementary Methods, Supplementary Results).

Statistics and reproducibility
Behavioural and fMRI data of 43 subjects were analysed with non-parametric tests.

Data of perceived face identity in face morphs were tested against a chance level of 50%, i.e., that the prior had no influence on the perception of the face morph, using a two-sided Wilcoxon signed rank test. RTs of the different conditions (match, mismatch, neutral, partial) were compared using a Friedman test and post-hoc tests (Tukey–Kramer).

For the RT analysis of the partial trials split up into trials in which participants had answered to have either perceived the expected or the unexpected identity, we used a two-sided paired Wilcoxon signed rank test. Similarly, for comparing the RTs of trials with prior-confirming responses to the match condition, we calculated a two-sided paired Wilcoxon signed rank test. Accuracy data were analysed using a Friedman test and post-hoc tests with Tukey–Kramer's critical value.

For the univariate whole-brain analyses of unexpected compared to expected faces ('mismatch > match' and 'unexpected > expected'), we report cluster activations (p(FWE) < 0.05) with a cluster-inducing threshold of $p < 0.001$. For the small-volume corrected ROI analyses (OFA, pFFA, aTL), we report peak activations (p(FWEsvc < 0.05)).

Using RSA[64,65], we calculated Kendall's Tau A correlations between the hypothesised and the neural dissimilarity structures. In our multivariate ROI analyses, the correlations for each hypothesis model (PE, Sharpening, Sensory Input) were tested against zero using one-sided Wilcoxon signed rank tests. Significance was evaluated by Bonferroni-correcting for the number of tests per ROI. For the model comparisons, we used two-sided paired Wilcoxon signed rank tests. Significance was inferred by FDR-correcting[131] the p-values for all model comparisons per ROI. For comparing left to right hemispheric correlations, we performed analyses of variances using ARTool[132]. For main effects and interactions, significance was evaluated by p(unc.) < 0.05, for post-hoc pairwise comparisons by $p < 0.05$, Tukey-corrected[133]. Furthermore, we report whole-brain searchlight analysis results for our different hypothesis models based on individual Fisher's z-transformed correlation maps using one-sample t-tests (p(FWE) < 0.05). Finally, we report exploratory classification ROI analyses of the face morphs as the expected or unexpected face identity at p(unc.) < 0.05.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The face stimuli used in this study were created with FaceGen Modeller Core 3.22 (Singular Inversion; [https://facegen.com]) and are available at the OSF ([https://osf.io/765jx/]). The scene images used in this study were taken from the SUN database[113] ([https://groups.csail.mit.edu/vision/SUN/hierarchy.html]) and the indoor scene database[114] ([https://web.mit.edu/torralba/www/indoor.html]) and are available at [https://osf.io/765jx/]. The exemplary scene image in Fig. 1 is in public domain and available at [https://commons.wikimedia.org]. The VGG-Face model[63] used in this study is available at [www.robots.ox.ac.uk/~vgg/software/vgg_face/]. The VGG-16[66] and ResNet50[67] models used in this study, pre-trained on the ImageNet dataset, are available via MATLAB ([https://de.mathworks.com/help/deeplearning/ref/vgg16.html]; [https://de.mathworks.com/help/deeplearning/ref/resnet50.html]). The raw behavioural and fMRI data generated in this study are available from the authors upon reasonable request. Source data are provided with this paper.

Code availability

Custom code for behavioural (MATLAB, R) and multivariate analyses (MATLAB, Python) is available at the OSF ([https://osf.io/765jx/]). We programmed our experiments using MATLAB R2020b ([https://de.mathworks.com]) and Psychtoolbox (v3.0.18; [www.psychtoolbox.org]). For stimulus presentation and data collection, we used different MATLAB and Psychtoolbox versions (MATLAB: R2016b, R2020b; Psychtoolbox: v3.014, v3.0.17, v3.0.18). For our behavioural data analyses, we used MATLAB R2020b and R/RStudio (R v4.2.0, [https://www.r-project.org]; Rstudio v2022.02.2; [https://posit.co]). For our univariate fMRI data analyses, we used SPM12 ([https://www.fil.ion.ucl.ac.uk/spm/software/spm12]). For our multivariate RSA, we used the RSAtoolbox[65] (v0.0.4; [https://github.com/rsagroup/rsatoolbox]) in

Python 3.9.12. For neuroanatomical labelling, we used the Neuromorphmetrics atlas (Neuromorphometrics, Inc.) implemented in SPM12 as well as the Harvard-Oxford Cortical Structural Atlas and the Harvard-Oxford Subcortical Structural Atlas in FSLeyes (v0.24.3). For visualisation, we used MRIcroGL (v1.2.20220720; [https://www.nitrc.org/projects/mricrogl]). For our non-parametric analysis of variance for the RSA split up by hemisphere, we used the ARTool-package[132,133] (v0.11.1; [https://cran.r-project.org/web/packages/ARTool/index.html]) in RStudio. For our multivariate classification analyses, we used The Decoding Toolbox[134] (v3.999F; [https://sites.google.com/site/tdtdecodingtoolbox/]) in MATLAB R2020b.

References

1. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
2. Friston, K. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).
3. Von Helmholtz, H. & Nagel, W. A. *Handbuch Der Physiologischen Optik* (L. Voss, 1909).
4. Blank, H. & Davis, M. H. Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLoS Biol.* **14**, e1002577 (2016).
5. Mumford, D. On the computational architecture of the neocortex. II The role of cortico-cortical loops. *Biol. Cybern.* **66**, 241–251 (1992).
6. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
7. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* **46**, 219–227 (2017).
8. Amado, C. et al. Neuroimaging results suggest the role of prediction in cross-domain priming. *Sci. Rep.* **8**, 10356 (2018).
9. Blank, H., Kiebel, S. J. & von Kriegstein, K. How the human brain exchanges information across sensory modalities to recognize other people: information across sensory modalities. *Hum. Brain Mapp.* **36**, 324–339 (2015).
10. Bruce, V. & Valentine, T. Semantic priming of familiar faces. *Q. J. Exp. Psychol. Sect. A* **38**, 125–150 (1986).
11. Rieth, C. A. & Huber, D. E. Priming and habituation for faces: Individual differences and inversion effects. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 596–618 (2010).
12. Schweinberger, S. R., Pfütze, E.-M. & Sommer, W. Repetition priming and associative priming of face recognition: evidence from event-related potentials. *J. Exp. Psychol. Learn. Mem. Cogn.* **21**, 722–736 (1995).
13. Shehzad, Z. & McCarthy, G. Perceptual and semantic phases of face identification processing: a multivariate electroencephalography study. *J. Cogn. Neurosci.* **31**, 1827–1839 (2019).
14. Todorova, L. & Neville, D. A. Associative and identity words promote the speed of visual categorization: a hierarchical drift diffusion account. *Front. Psychol.* **11**, 466396 (2020).
15. Vladeanu, M., Lewis, M. & Ellis, H. Associative priming in faces: Semantic relatedness or simple co-occurrence? *Mem. Cogn.* **34**, 1091–1101 (2006).
16. Wiese, H. & Schweinberger, S. R. Accessing semantic person knowledge: temporal dynamics of nonstrategic categorical and associative priming. *J. Cogn. Neurosci.* **23**, 447–459 (2011).
17. Gao, Y. & Wang, X. A proportionally suppressed and prolonged LPP acts as a neurophysiological correlate of face identity aftereffect. *Brain Res* **1746**, 146969 (2020).
18. Hills, P. J., Elward, R. L. & Lewis, M. B. Cross-modal face identity aftereffects and their relation to priming. *J. Exp. Psychol. Hum. Percept. Perform.* **36**, 876–891 (2010).

19. Walther, C., Schweinberger, S. R., Kaiser, D. & Kovács, G. Neural correlates of priming and adaptation in familiar face perception. *Cortex* **49**, 1963–1977 (2013).

20. Summerfield, C., Wyart, V., Mareike Johnen, V. & de Gardelle, V. Human scalp electroencephalography reveals that repetition suppression varies with expectation. *Front. Hum. Neurosci.* **5**, 67 (2011).

21. Johnston, P. et al. Temporal and spatial localization of prediction-error signals in the visual brain. *Biol. Psychol.* **125**, 45–57 (2017).

22. de Gardelle, V., Stokes, M., Johnen, V. M., Wyart, V. & Summerfield, C. Overlapping multivoxel patterns for two levels of visual expectation. *Front. Hum. Neurosci.* **7**, 158 (2013).

23. Egner, T., Monti, J. M. & Summerfield, C. Expectation and surprise determine neural population responses in the ventral visual stream. *J. Neurosci.* **30**, 16601–16608 (2010).

24. Grotheer, M. & Kovács, G. The relationship between stimulus repetitions and fulfilled expectations. *Neuropsychologia* **67**, 175–182 (2015).

25. Pajani, A., Kouider, S., Roux, P. & de Gardelle, V. Unsuppressible repetition suppression and exemplar-specific expectation suppression in the fusiform face area. *Sci. Rep.* **7**, 160 (2017).

26. Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M. & Egner, T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* **11**, 1004–1006 (2008).

27. Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15**, 745–756 (2014).

28. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).

29. Knill, D. C. & Pouget, A. The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends Neurosci.* **27**, 712–719 (2004).

30. Blank, H., Spangenberg, M. & Davis, M. H. Neural prediction errors distinguish perception and misperception of speech. *J. Neurosci.* **38**, 6076–6089 (2018).

31. González-García, C. & He, B. J. A gradient of sharpening effects by perceptual prior across the human cortical hierarchy. *J. Neurosci.* **41**, 167–178 (2021).

32. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* **75**, 265–270 (2012).

33. Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A* **20**, 1434 (2003).

34. de Lange, F. P., Heilbron, M. & Kok, P. How do expectations shape perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).

35. Yon, D., Gilbert, S. J., de Lange, F. P. & Press, C. Action sharpens sensory representations of expected outcomes. *Nat. Commun.* **9**, 4288 (2018).

36. Alink, A. & Blank, H. Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage* **231**, 117824 (2021).

37. Ufer, C. & Blank, H. Multivariate analysis of brain activity patterns as a tool to understand predictive processes in speech perception. *Lang. Cogn. Neurosci.* **0**, 1–17 (2023).

38. Blank, H., Wieland, N. & von Kriegstein, K. Person recognition and the brain: merging evidence from patients and healthy individuals. *Neurosci. Biobehav. Rev.* **47**, 717–734 (2014).

39. Blank, H., Alink, A. & Büchel, C. Multivariate functional neuroimaging analyses reveal that strength-dependent face expectations are represented in higher-level face-identity areas. *Commun. Biol.* **6**, 1–10 (2023).

40. Collins, J. A. & Olson, I. R. Beyond the FFA: the role of the ventral anterior temporal lobes in face processing. *Neuropsychologia* **61**, 65–79 (2014).

41. Grill-Spector, K., Weiner, K. S., Kay, K. & Gomez, J. The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* **3**, 167–196 (2017).

42. Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. The distributed human neural system for face perception. *Trends Cogn. Sci.* **4**, 223–233 (2000).

43. Rajimehr, R., Young, J. C. & Tootell, R. B. H. An anterior temporal face patch in human cortex, predicted by macaque maps. *Proc. Natl Acad. Sci. USA* **106**, 1995–2000 (2009).

44. Tsao, D. Y., Moeller, S. & Freiwald, W. A. Comparing face patch systems in macaques and humans. *Proc. Natl Acad. Sci. USA* **105**, 19514–19519 (2008).

45. Anzellotti, S. & Caramazza, A. From parts to identity: invariance and sensitivity of face representations to different face halves. *Cereb. Cortex* **26**, 1900–1909 (2016).

46. Pitcher, D., Walsh, V. & Duchaine, B. The role of the occipital face area in the cortical face perception network. *Exp. Brain Res.* **209**, 481–493 (2011).

47. Tsantani, M. et al. FFA and OFA encode distinct types of face identity information. *J. Neurosci.* **41**, 1952–1969 (2021).

48. Liu, J., Harris, A. & Kanwisher, N. Perception of face parts and face configurations: an fMRI study. *J. Cogn. Neurosci.* **22**, 203–211 (2010).

49. Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J. & Tootell, R. B. H. Lower-level stimulus features strongly influence responses in the fusiform face area. *Cereb. Cortex* **21**, 35–47 (2011).

50. Axelrod, V. & Yovel, G. Successful decoding of famous faces in the fusiform face area. *PLoS ONE* **10**, e0117126 (2015).

51. Nestor, A., Plaut, D. C. & Behrmann, M. Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc. Natl Acad. Sci. USA* **108**, 9998–10003 (2011).

52. Anzellotti, S., Fairhall, S. L. & Caramazza, A. Decoding representations of face identity that are tolerant to rotation. *Cereb. Cortex* **24**, 1988–1995 (2014).

53. Yang, H., Susilo, T. & Duchaine, B. The anterior temporal face area contains invariant representations of face identity that can persist despite the loss of right FFA and OFA. *Cereb. Cortex* **26**, 1096–1107 (2016).

54. Larsson, J. & Smith, A. T. fMRI repetition suppression: neuronal adaptation or stimulus expectation? *Cereb. Cortex* **22**, 567–576 (2012).

55. Ouden, H. E. M., den, Daunizeau, J., Roiser, J., Friston, K. J. & Stephan, K. E. Striatal prediction error modulates cortical coupling. *J. Neurosci.* **30**, 3210–3219 (2010).

56. Schwiedrzik, C. M. & Freiwald, W. A. High-level prediction signals in a low-level area of the macaque face-processing hierarchy. *Neuron* **96**, 89–97.e4 (2017).

57. Kaliukhovich, D. A. & Vogels, R. Stimulus repetition probability does not affect repetition suppression in macaque inferior temporal cortex. *Cereb. Cortex* **21**, 1547–1558 (2011).

58. Vinken, K., Beeck, H. P. Ode & Vogels, R. Face repetition probability does not affect repetition suppression in macaque inferotemporal cortex. *J. Neurosci.* **38**, 7492–7504 (2018).

59. Dobs, K., Isik, L., Pantazis, D. & Kanwisher, N. How face perception unfolds over time. *Nat. Commun.* **10**, 1–10 (2019).

60. Grossman, S. et al. Convergent evolution of face spaces across human face-selective neuronal groups and deep convolutional networks. *Nat. Commun.* **10**, 4934 (2019).

61. Ratan Murty, N. A., Bashivan, P., Abate, A., DiCarlo, J. J. & Kanwisher, N. Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nat. Commun.* **12**, 5540 (2021).

62. van Dyck, L. E. & Gruber, W. R. Modeling biological face recognition with deep convolutional neural networks. *J. Cogn. Neurosci.* **35**, 1521–1537 (2023).

63. Parkhi, O. M., Vedaldi, A. & Zisserman, A. Deep face recognition. In *Proc. British Machine Vision Conference 2015* 41.1–41.12 (British Machine Vision Association, Swansea, 2015).

64. Kriegeskorte, N. Representational similarity analysis – connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* https://doi.org/10.3389/neuro.06.004.2008 (2008).

65. Nili, H. et al. A toolbox for representational similarity analysis. *PLoS Comput. Biol.* **10**, e1003553 (2014).

66. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. In 1–14 https://doi.org/10.48550/arXiv.1409.1556 (2015).

67. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* 770–778 (IEEE, Las Vegas, NV, USA, 2016).

68. Levine, S. M. & Schwarzbach, J. V. Individualizing representational similarity analysis. *Front. Psychiatry* **12**, 729457 (2021).

69. Lee, J. & Geng, J. J. Idiosyncratic patterns of representational similarity in prefrontal cortex predict attentional performance. *J. Neurosci.* **37**, 1257–1268 (2017).

70. Fouragnan, E., Retzler, C. & Philiastides, M. G. Separate neural representations of prediction error valence and surprise: evidence from an fMRI meta-analysis. *Hum. Brain Mapp.* **39**, 2887–2906 (2018).

71. Ham, T. E. et al. Distinct frontal networks are involved in adapting to internally and externally signaled errors. *Cereb. Cortex* **23**, 703–713 (2013).

72. Cohen, M. X., Heller, A. S. & Ranganath, C. Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Cogn. Brain Res.* **23**, 61–70 (2005).

73. Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X. & Milham, M. P. Distinct neural mechanisms of risk and ambiguity: a meta-analysis of decision-making. *NeuroImage* **32**, 477–484 (2006).

74. Summerfield, C. et al. Predictive codes for forthcoming perception in the frontal cortex. *Science* **314**, 1311–1314 (2006).

75. Muukkonen, I., Ölander, K., Numminen, J. & Salmela, V. R. Spatio-temporal dynamics of face perception. *NeuroImage* **209**, 116531 (2020).

76. Visconti di Oleggio Castello, M. et al. The neural representation of personally familiar and unfamiliar faces in the distributed system for face perception. *Sci. Rep.* **7**, 12237 (2017).

77. Ambrus, G. G., Amado, C., Krohn, L. & Kovács, G. TMS of the occipital face area modulates cross-domain identity priming. *Brain Struct. Funct.* **224**, 149–157 (2019).

78. Wiese, H. & Schweinberger, S. R. Event-related potentials indicate different processes to mediate categorical and associative priming in person recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* **34**, 1246–1263 (2008).

79. Young, A. W., Flude, B. M., Hellawell, D. J. & Ellis, A. W. The nature of semantic priming effects in the recognition of familiar people. *Br. J. Psychol.* **85**, 393–411 (1994).

80. Leopold, D., O'Toole, A., Vetter, T. & Blanz, V. Prototype-referenced shape encoding revealed by high-level aftereffects. *Nat. Neurosci.* **4**, 89–94 (2001).

81. Walther, C., Schweinberger, S. R. & Kovács, G. Decision-dependent aftereffects for faces. *Vis. Res* **100**, 47–55 (2014).

82. Walther, C., Schweinberger, S. R. & Kovács, G. Adaptor identity modulates adaptation effects in familiar face identification and their neural correlates. *PLoS ONE* **8**, e70525 (2013).

83. Mueller, R., Utz, S., Carbon, C.-C. & Strobach, T. Face adaptation and face priming as tools for getting insights into the quality of face space. *Front. Psychol.* **11**, 501653 (2020).

84. Snyder, J. S., Schwiedrzik, C. M., Vitela, A. D. & Melloni, L. How previous experience shapes perception in different sensory modalities. *Front. Hum. Neurosci.* **9**, 594 (2015).

85. Leopold, D. A., Rhodes, G., Müller, K.-M. & Jeffery, L. The dynamics of visual adaptation to faces. *Proc. R. Soc. B Biol. Sci.* **272**, 897–904 (2005).

86. Loued-Khenissi, L., Pfeuffer, A., Einhäuser, W. & Preuschoff, K. Anterior insula reflects surprise in value-based decision-making and perception. *NeuroImage* **210**, 116549 (2020).

87. Pelgrims, B., Andres, M. & Olivier, E. Double dissociation between motor and visual imagery in the posterior parietal cortex. *Cereb. Cortex* **19**, 2298–2307 (2009).

88. Sylvester, C.-Y. C. et al. Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia* **41**, 357–370 (2003).

89. Sperling, R. et al. Putting names to faces. *NeuroImage* **20**, 1400–1410 (2003).

90. Apps, M. A. J. & Tsakiris, M. Predictive codes of familiarity and context during the perceptual learning of facial identities. *Nat. Commun.* **4**, 2698 (2013).

91. Zaragoza-Jimenez, N. et al. Modeling face recognition in the predictive coding framework: a combined computational modeling and functional imaging study. *Cortex* https://doi.org/10.1016/j.cortex.2023.05.021 (2023).

92. Guntupalli, J. S., Wheeler, K. G. & Gobbini, M. I. Disentangling the representation of identity from head view along the human face processing pathway. *Cereb. Cortex* **27**, 46–53 (2017).

93. Taylor, M. J. et al. Neural correlates of personally familiar faces: parents, partner and own faces. *Hum. Brain Mapp.* **30**, 2008–2020 (2009).

94. Maurer, D. et al. Neural correlates of processing facial identity based on features versus their spacing. *Neuropsychologia* **45**, 1438–1451 (2007).

95. Renzi, C. et al. Processing of featural and configural aspects of faces is lateralized in dorsolateral prefrontal cortex: a TMS study. *NeuroImage* **74**, 45–51 (2013).

96. Aminoff, E. M., Kveraga, K. & Bar, M. The role of the parahippocampal cortex in cognition. *Trends Cogn. Sci.* **17**, 379–390 (2013).

97. Bar, M., Aminoff, E. & Ishai, A. Famous faces activate contextual associations in the parahippocampal cortex. *Cereb. Cortex* **18**, 1233–1238 (2008).

98. Aitken, F. & Kok, P. Hippocampal representations switch from errors to predictions during acquisition of predictive associations. *Nat. Commun.* **13**, 3294 (2022).

99. Stachenfeld, K. L., Botvinick, M. M. & Gershman, S. J. The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).

100. Barron, H. C., Auksztulewicz, R. & Friston, K. Prediction and memory: a predictive coding account. *Prog. Neurobiol.* **192**, 101821 (2020).

101. Jozwik, K. M. et al. Face dissimilarity judgments are predicted by representational distance in morphable and image-computable models. *Proc. Natl Acad. Sci. USA* **119**, e2115047119 (2022).

102. Jiahui, G. et al. Modeling naturalistic face processing in humans with deep convolutional neural networks. *Proc. Natl Acad. Sci. USA* **120**, e2304085120 (2023).

103. Xu, Y. & Vaziri-Pashkam, M. Limits to visual representational correspondence between convolutional neural networks and the human brain. *Nat. Commun.* **12**, 2065 (2021).

104. Von Der Heide, R., Skipper, L. & Olson, I. Anterior temporal face patches: a meta-analysis and empirical study. *Front. Hum. Neurosci.* **7**, 17 (2013).

105. Jonas, J. et al. A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials. *Proc. Natl Acad. Sci. USA* **113**, E4088–E4097 (2016).

106. Kriegeskorte, N., Formisano, E., Sorger, B. & Goebel, R. Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc. Natl Acad. Sci. USA* **104**, 20600–20605 (2007).

107. Erez, Y. & Duncan, J. Discrimination of visual categories based on behavioral relevance in widespread regions of frontoparietal cortex. *J. Neurosci.* **35**, 12383–12393 (2015).

108. de Gardelle, V., Waszczuk, M., Egner, T. & Summerfield, C. Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cereb. Cortex* **23**, 2235–2244 (2013).

109. Walsh, K. S., McGovern, D. P., Clark, A. & O'Connell, R. G. Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Ann. N. Y. Acad. Sci.* **1464**, 242–268 (2020).

110. Felleman, D. J & Van Essen, D. C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).

111. Friston, K. Hierarchical models in the brain. *PLoS Comput. Biol.* **4**, e1000211 (2008).

112. Abudarham, N. & Yovel, G. Reverse engineering the face space: Discovering the critical features for face identification. *J. Vis.* **16**, 40 (2016).

113. Xiao, J., Hays, J., Ehinger, K. A., Oliva, A. & Torralba, A. SUN database: large-scale scene recognition from abbey to zoo. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 3485–3492 (IEEE, San Francisco, CA, USA, 2010).

114. Quattoni, A. & Torralba, A. Recognizing indoor scenes. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 413–420 (IEEE, Miami, FL, USA, 2009).

115. Willenbockel, V. et al. Controlling low-level image properties: The SHINE toolbox. *Behav. Res. Methods* **42**, 671–684 (2010).

116. Fox, C. J., Iaria, G. & Barton, J. J. S. Defining the face processing network: optimization of the functional localizer in fMRI. *Hum. Brain Mapp.* **30**, 1637–1651 (2009).

117. Kanwisher, N., McDermott, J. & Chun, M. M. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* **17**, 4302–4311 (1997).

118. Buck, S. & Sidhu, M. K. A Guide to designing a memory fMRI paradigm for pre-surgical evaluation in temporal lobe epilepsy. *Front. Neurol.* **10**, 488840 (2020).

119. Nau, M. Functional imaging of the human medial temporal lobe: a neuroscientist's guide to fMRI pulse sequence optimization https://doi.org/10.17605/OSF.IO/CQN4Z (2019).

120. Olszowy, W., Aston, J., Rua, C. & Williams, G. B. Accurate autocorrelation modeling substantially improves fMRI reliability. *Nat. Commun.* **10**, 1220 (2019).

121. Tsantani, M., Kriegeskorte, N., McGettigan, C. & Garrido, L. Faces and voices in the brain: a modality-general person-identity representation in superior temporal sulcus. *NeuroImage* **201**, 116004 (2019).

122. Volfart, A. et al. Intracerebral electrical stimulation of the right anterior fusiform gyrus impairs human face identity recognition. *NeuroImage* **250**, 118932 (2022).

123. Lee, S.-M., Tibon, R., Zeidman, P., Yadav, P. S. & Henson, R. Effects of face repetition on ventral visual stream connectivity using dynamic causal modelling of fMRI data. *NeuroImage* **264**, 119708 (2022).

124. Brett, M., Anton, J.-L., Valabregue, R. & Poline, J.-B. Region of interest analysis using an SPM toolbox. In *8th international conference on functional mapping of the human brain*, Vol. 16, 497 (2002).

125. Zhen, Z. et al. Quantifying interindividual variability and asymmetry of face-selective regions: a probabilistic functional atlas. *NeuroImage* **113**, 13–25 (2015).

126. Pinsk, M. A. et al. Neural representations of faces and body parts in macaque and human cortex: a comparative fMRI study. *J. Neurophysiol.* **101**, 2581–2600 (2009).

127. Weiner, K. S. & Grill-Spector, K. Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage* **52**, 1559–1573 (2010).

128. Garrido, L. et al. Voxel-based morphometry reveals reduced grey matter volume in the temporal cortex of developmental prosopagnosics. *Brain* **132**, 3443–3455 (2009).

129. Deng, J. et al. ImageNet: a large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, Miami, FL, USA, 2009).

130. Misaki, M., Kim, Y., Bandettini, P. A. & Kriegeskorte, N. Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage* **53**, 103–118 (2010).

131. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

132. Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 143–146 (ACM, Vancouver, BC, Canada, 2011).

133. Elkin, L. A., Kay, M., Higgins, J. J. & Wobbrock, J. O. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology* 754–768 (ACM, Virtual Event USA, 2021).

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

135. Morey, R. D. Confidence intervals from normalized data: a correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* **4**, 61–64 (2008).

Author contributions

A.G. and H.B. designed the project; A.G. performed experiments and analysed data; A.G. and H.B. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-47749-9.

Correspondence and requests for materials should be addressed to Annika Garlichs or Helen Blank.

Peer review information *Nature Communications* thanks Gyula Kovacs and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary Information: Prediction Error Processing and Sharpening of Expected Information Across the Face-Processing Hierarchy

Authors: Annika Garlichs[1]✉ & Helen Blank[1]✉

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf,

20246 Hamburg, Germany. ✉emails: a.garlichs@uke.de, h.blank@uke.de

Supplementary Figures

Supplementary Methods

Supplementary Results

Supplementary Tables

Supplementary Figures

Supplementary Figure 1. RSA based on the network ResNet50[1]. a-d) We performed Representational Similarity Analysis (RSA) in our four ROIs (occipital face area, OFA; posterior fusiform face area, pFFA; anterior temporal lobe, aTL; inferior/middle temporal gyrus, ITG/MTG) and applied hypothesis models based on Prediction Error (PE), a Sharpening, and a pure Sensory Input model. For the creation of the hypothesis representational dissimilarity matrices (RDM), we extracted activations from layer res5b_branch2b in MATLAB (based on[2]). Grey bars indicate the between-subject standard error of the mean (SEM), black bars indicate the within-subject SEM[3] (N = 43 participants). Asterisks indicate the tests of each hypothesis model against zero (one-sided Wilcoxon signed rank test), black asterisks showing significance Bonferroni-corrected for the number of models per ROI (N = 3), grey asterisks showing uncorrected significance $p < .05$, and horizontal lines indicate model comparison results (two-sided paired Wilcoxon signed rank tests), FDR-corrected[4] per ROI. Grey rectangles display the lower and upper boundary of the noise ceiling for each ROI[5].
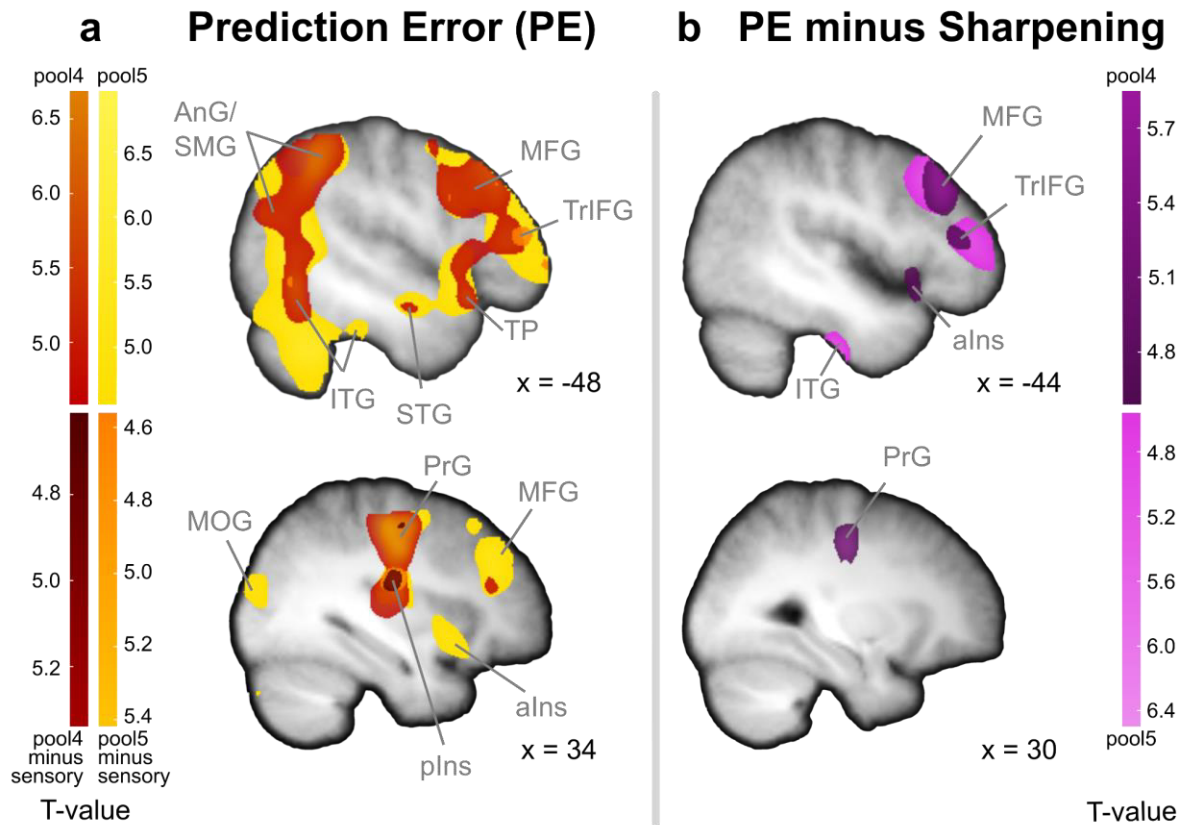
Supplementary Figure 2. Whole-brain searchlight analyses for the hypothesis models Prediction Error (PE) and Sharpening (VGG-16). Results for the comparison of the neural and hypothesised dissimilarity structures based on pool4 and pool5 layers from VGG-16 are displayed against zero and as difference maps against a sensory searchlight without prior influence based on the second convolutional layer[6], respectively. **a) Searchlight analyses results for PE**: Clusters were identified in angular gyrus (AnG), inferior occipital gyrus (IOG), triangular part of the inferior frontal gyrus (TrIFG), temporal pole (TP), superior temporal gyrus (STG), inferior temporal gyrus (ITG), middle occipital gyrus (MOG), precentral gyrus (PrG), middle frontal gyrus (MFG), anterior insula (aIns), and posterior insula (pIns). **b) Comparison of the PE > Sharpening searchlight results**: Stronger correlations for PE than Sharpening were evident in the left MFG, TrIFG, and ITG, and in the right PrG. All maps are displayed at *p*(FWE) < .05. Maps are overlaid on the average structural T1 image in Montreal Neurological Institute (MNI) template space.
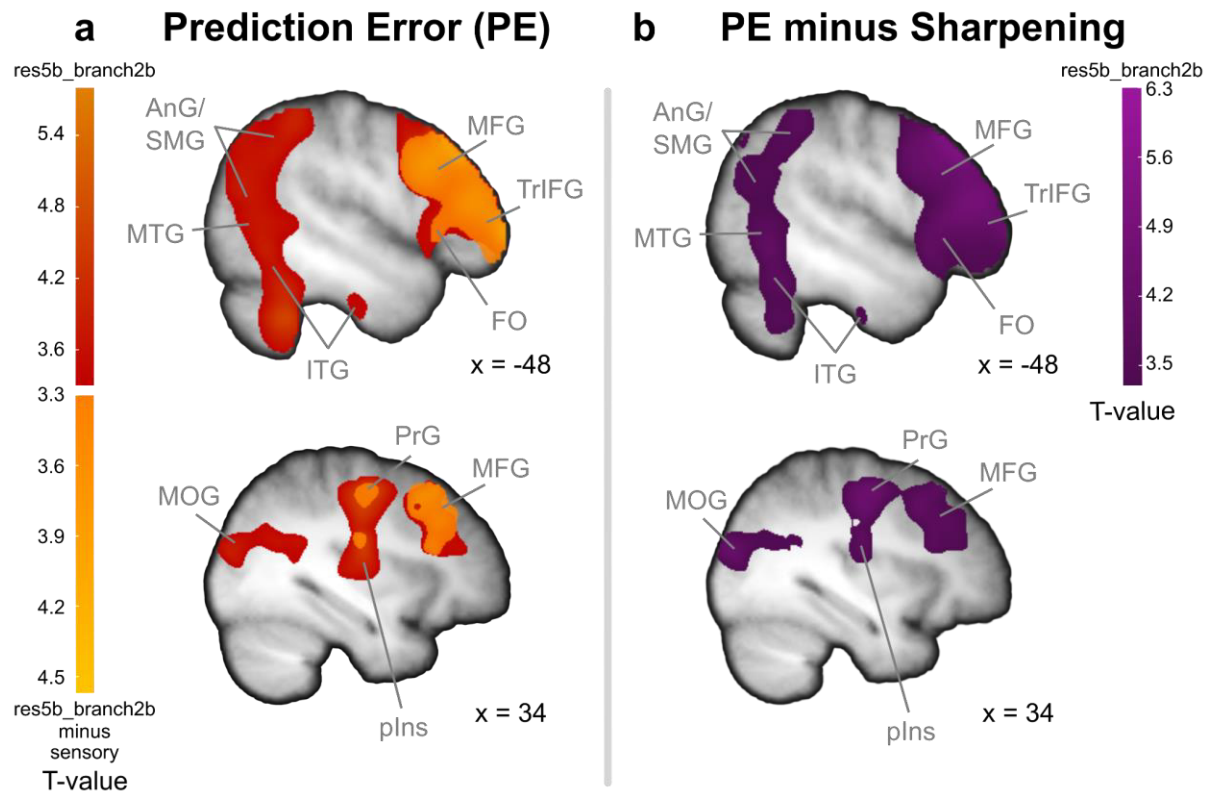
Supplementary Figure 3. Whole-brain searchlight analyses for the hypothesis models Prediction Error (PE) and Sharpening (ResNet50). Results for the comparison of the neural and hypothesised dissimilarity structures based on layer res5b_branch2b from ResNet50 are displayed against zero and as difference maps against a sensory searchlight without prior influence also based on res5b_branch2b activations. **a) Searchlight analyses results for PE**: Clusters were identified in angular gyrus (AnG), supramarginal gyrus (SMG), middle frontal gyrus (MFG), triangular part of the inferior frontal gyrus (TrIFG), frontal operculum (FO), inferior temporal gyrus (ITG), middle temporal gyrus (MTG), middle occipital gyrus (MOG), precentral gyrus (PrG), and posterior insula (pIns). **b) Comparison of the PE > Sharpening searchlight results**: Stronger correlations for PE than Sharpening were evident in the areas identified in **a)**. All maps are displayed at $p < .001$, uncorrected. Maps are overlaid on the average structural T1 image in Montreal Neurological Institute (MNI) template space.

Hypothesis Representational Dissimilarity Matrices (RDM)
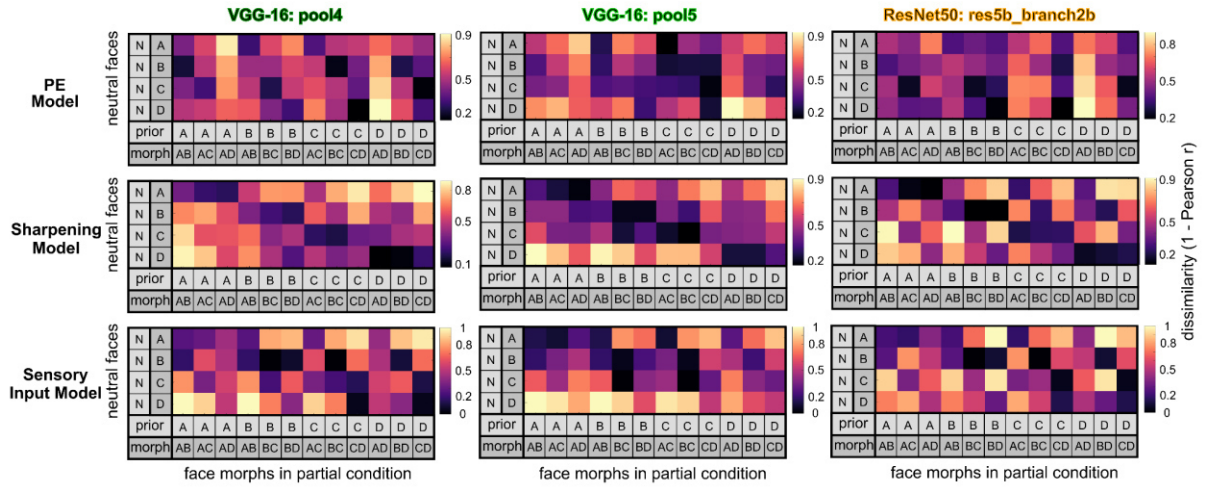
Supplementary Figure 4. Hypothesis Representational Dissimilarity Matrices (RDMs) for the object-trained neural networks VGG-16 and ResNet50. For VGG-Face, the RDMs for our hypothesis models (Prediction Error, PE; Sharpening; Sensory Input) were based on pool4 and pool5 activations. For ResNet50, the RDMs were based on activations extracted from the layer res5b_branch2b in MATLAB. For visualisation, the displayed PE and Sharpening RDMs were averaged across individual ($N = 43$) RDMs.
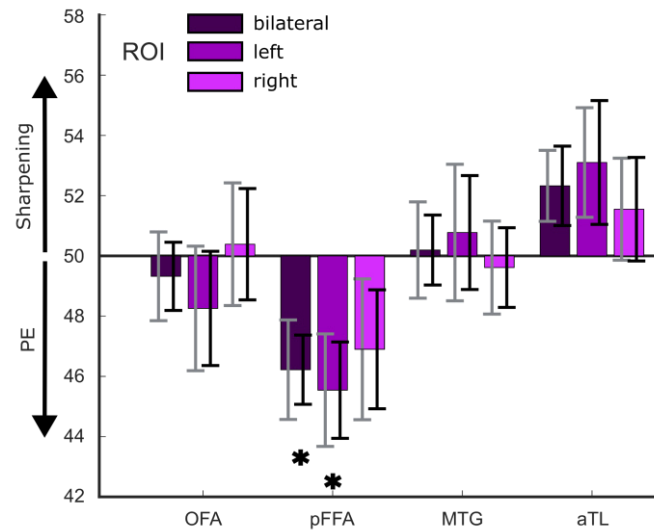
Supplementary Figure 5. Classification analyses of the face morph images. Images of face morphs were classified as either the expected or unexpected identity. Classification values above 50% indicate classification as the expected identity (i.e., Sharpening) while values below 50% indicate classification as the unexpected identity (i.e., Prediction Error (PE)). The darkest purple colour shows the mean classification values across hemispheres, the middle purple colour shows the left hemispheric values, and the light purple shows the right hemispheric values. Grey bars indicate the between-subject standard error of the mean (SEM), black bars indicate the within-subject SEM[3] (N = 43 participants). Asterisks show significance ($p < .05$, uncorrected). OFA = occipital face area, pFFA = posterior fusiform face area, MTG = middle temporal gyrus, aTL = anterior temporal lobe.
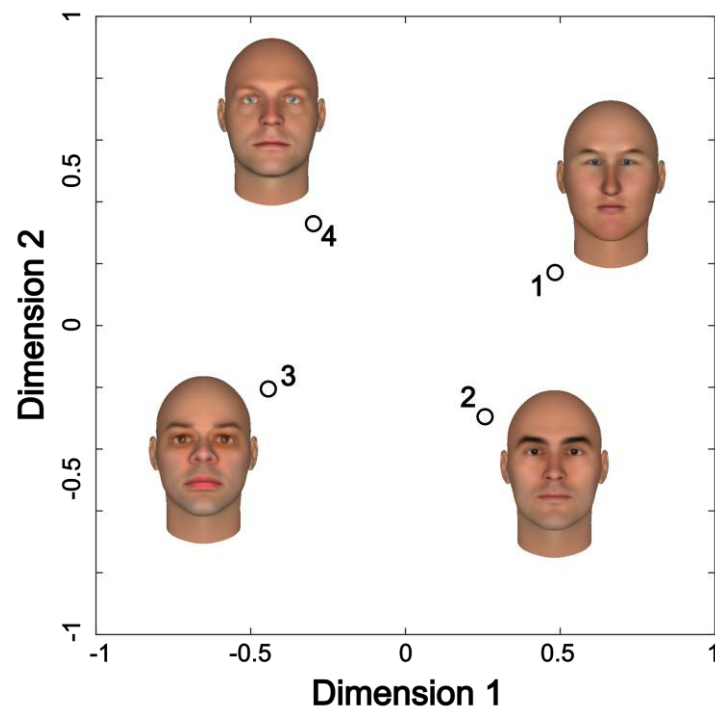
Supplementary Figure 6. Multidimensional scaling (MDS) (pool4) of the four face images. Four male identities were created using FaceGen Modeller Core 3.22. To ensure that the four selected faces were equally distinct from each other, a classical MDS was performed based on the activations of layer pool4 of the deep neural network VGG-Face[7]. The distance measure was '1 - Pearson correlation' and the dissimilarities were rescaled to values between 0 and 1. The dissimilarity structure based on two dimensions showed that the images were equally dissimilar to each other.
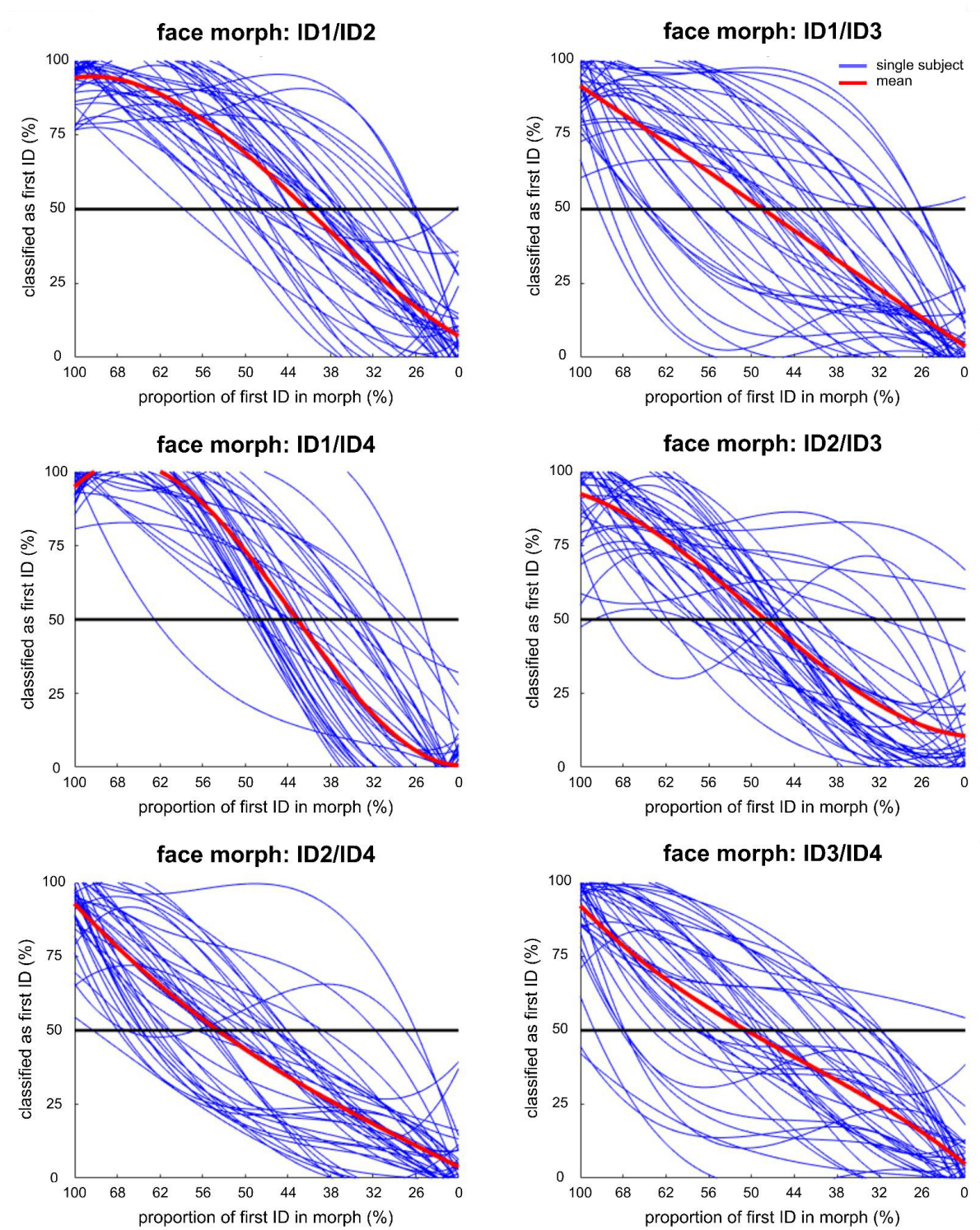
Supplementary Figure 7. Calibration results of the morph selection experiment. Participants performed a face morph calibration experiment on the day before the fMRI experiment took place (see Supplementary Methods: Morph Calibration Experiment and Selection of Morph Levels in Face Pairs for Individual Participants). The goal was to select for each participant the individual morph levels for each face pair that equaled the point of ambiguous perception (i.e., the morph level that they individually identified as the first and the

9

second identity in 50% of the cases). A polynomial function of 3rd order was fitted to each participant's data to grasp the typical nonlinear classification of face morphs between two identities[8]. Single participants' fitted functions are shown as blue lines ($N = 43$). The mean response function across all participants is shown in red. The black line indicates the threshold for identifying a morph level as the first and the second identity equally often, i.e., ambiguous perception.

Supplementary Methods

Image Specifications

All face images and all morphs were created in two sizes: 400 x 400 and 531 x 531 pixels. The smaller images were used for the behavioural training sessions outside of the scanner on monitors (20") with a spatial resolution of 1920 x 1200. The distance between the retina and the monitor was approximately 72 cm. The larger images were used for the experiment in the fMRI scanner where images were displayed via a NordicNeuroLab InroomViewing Device 40" LCD monitor with a spatial resolution of 1920 x 1080. The distance from the participant's eyes to the head coil mirror was 13 cm and from the mirror to the monitor 150 cm. The image sizes were calculated so that they had the same visual angle inside and outside of the scanner (8.59° x 8.59°).

From each original scene image, an excerpt was taken and rescaled to 531 x 531 and 400 x 400 pixels. Again, the larger images were used in the scanner for the main experiment and the functional localizer and the smaller images were used for training sessions outside of the scanner.

Morph Calibration Experiment

On the first day, participants started with the morph calibration experiment. The goal of the calibration was to identify each individual's morph level that corresponded to the 50/50 perceptual threshold at which both identities were equally often identified. For the calibration experiment, morphs between each pair of the four faces ranging from 26/74 to 68/32 (faceID1/faceID2) in steps of six percent were created. Morph windows were slightly shifted towards the less dominant face identity within each pair based on a behavioural pilot with an independent set of participants ($N = 9$).

In the first part of the calibration, participants learned to associate each of the four faces with their respective names. They got accustomed to the duration of 100 ms per face presentation and the additional noise applied to the faces. In the second part of the calibration, morph faces for all face pairs were presented. The task was to indicate via pressing one of four buttons with the right hand (index, middle, ring, and pinky finger) which person they saw: Ari, Bob, Cid, or Dan. If the presented face was not one of the four original faces, their task was to indicate which of the four faces they saw predominantly. Feedback was provided if the participants were too slow or if they wrongly answered a person that was not contained in a morph. The calibration was divided into three sub-blocks for five minutes: In each block, each morph level of a face pair (eight morph levels per pair) was presented twice. Additionally, each of the original faces was presented twice in each block to reinforce the originally learned face representations. This resulted in 112 trials per block. Between each block, there was a break

screen showing the original four faces and their names. The order of faces was pseudo-randomized so that the same face or face morph was consecutively presented twice at maximum. After the calibration experiment, the most ambiguous face morph for each face pair was selected (see Supplementary Methods: Selection of Morph Levels in Face Pairs for Individual Participants) and noise was added to the images for the consecutive training sessions and the main experiment.

Selection of Morph Levels in Face Pairs for Individual Participants

Based on the calibration experiment, the individual morph levels for each face pair that equaled the point of ambiguous perception were selected (see Supplementary Figure 7). We tested six morph levels per face pair. For each morph level, the number of responses in favour of the first and the second identity within the morph was counted across the three blocks of the experiment. A polynomial function of 3rd order was fitted to the data to grasp the typical nonlinear classification of face morphs between two identities[8]. Two additional morph levels (100/0 and 0/100) with perfect classification scores were added for each morph pair to obtain a better fit. The fitting of the polynomial function resulted in a vector of 1010 interpolated values representing probabilities of choosing the first identity in a face morph. All interpolated values close to 0.5 (i.e., 50% probability of identifying a morph as the first and the second identity) with a tolerance of 0.001 were extracted to identify the most ambiguous morph level. In case the polynomial function did not cross the threshold line, a morph level was chosen to counteract the perceptual dominance of one face identity in a pair. For the first three participants, the morph levels were chosen based on visual inspection of the psychometric curves.

Training Sessions

Participants came to the lab on two consecutive days to take part in the experiment. On the first day, they completed the individual face morph calibration (25 min), a training session for learning the associations between scenes and faces (25 min), and a repetition of the last part of the association training which was a short form of the final experiment (5 min). On the second day, participants started with a shortened version of the association training (15 min), followed by a short repetition of the final task (5 min). Afterwards, they completed a short part (5 min) in which the inter-trial intervals (ITI) were as long as they would be in the scanner. For all previous training sessions, ITIs were shortened to reduce overall duration. All training sessions were performed outside of the scanner. Finally, the experiment in the scanner took place (52 min) followed by a functional localizer (4 min).

After the completion of the calibration experiment (see Supplementary Methods: Morph Calibration Experiment), participants took part in an association training session in

which they learned to associate each face with a scene. The training started by having definite transition probabilities, i.e., that the faces always correctly appeared after the scene to which they belonged. The task was to correctly identify each face as Ari, Bob, Cid, or Dan. No morphs were shown during this learning phase. Later, a neutral scene was introduced after which each face was equally likely to appear. The task was still to identify the faces, however, participants should press a button with the left thumb for *whichever* face appeared after the fifth, neutral scene. The allocation of which scene was predictive of which identity and which scene was neutral was counterbalanced across participants. There were five different allocation versions. As a next step of the association training session, catch trials were introduced: In some trials, a question mark followed a scene instead of a face. If a question mark appeared, the task for the participant was to indicate which person they had anticipated after the previous scene (catch trial). If the question mark appeared after the neutral scene, they should press a button with the left index finger to signal that all faces had been equally likely to appear (neutral catch trial). Afterwards, mismatch trials were introduced. Transition probabilities for the four scenes were manipulated such that they were followed by the correct face in ~66.66 % of the trials, in ~20 % by a face which had not been anticipated, and in ~13.33 % by a question mark. In the last part, face morphs were added as partial trials. This part had the same transition probabilities and number of trials as one block in the final fMRI experiment. The part consisted of 107 trials. There were 16 match trials, 48 partial trials, 12 mismatch trials, 12 catch trials, 16 neutral trials, and 3 neutral catch trials. The transition probabilities were as follows: The four main scenes (88 trials) were followed by the expected face or a morph containing it in ~72,73 % of the trials (~18.18 % match trials, ~54.54 % partial trials)., as well as mismatch (~13.64 %) and catch trials (~13.64 %). The neutral scene (19 trials) was followed by one of the four faces in ~84.21% of the trials and neutral catch trials with a question mark (~15.79 %). An exemplary trial can be seen in Figure 1b.

For the last part of the association training, participants were given feedback (see Supplementary Methods: Feedback) by the experimental leader when there were misunderstandings concerning the task and brief feedback about single mistakes was also provided if participants performed well. Finally, participants repeated the last part of the association training session and received a second feedback with improvements or still occurring errors. Overall, participants had learned and understood the task (accuracy: $M = 85.50$ %, $SD = 7.54$ %).

On the second day, participants completed a shortened version of the association training session and obtained feedback. The task and condition distribution of the final part was identical to one block in the fMRI experiment. The participants were informed that the timing in the scanner was different due to the specificities of the MRI scanner. They completed a four-minute experiment with longer ITIs (1500-2500 ms, mean: 2000 ms) and null events

(fixation cross) with a trial duration of 5300 ms so that they would not be surprised in the scanner.

Feedback

Participants received verbal feedback after completion of the training sessions and after each block of the fMRI experiment. They were given an overall score of their performance which was the mean of the accuracy scores for the match, partial, mismatch, catch, neutral, and neutral catch trials, as well as individual feedback for the different conditions. Partial trials were classified as correct if the participant answered in time to have perceived one of the two identities that were present in the morph. Feedback for specific conditions was provided if participants had not understood the task.

Behavioural Outlier

Although in the preregistration we had not defined an exclusion criterion, we decided to exclude one participant who performed worse than 99.9% of the participants (i.e., z = -4.16 for the z-standardised mean accuracy based on the match, mismatch, partial, catch, neutral, and neutral catch conditions). Therefore, the reported results are based on 43 participants.

Whole-Brain Mask

For the first-level analyses of the main experiment and the functional localizer, individual whole-brain masks were created based on grey matter, white matter, and cerebrospinal fluid probability maps obtained from SPM12's segmentation-function. The threshold for all three tissue types was set to 0. The mask was additionally smoothed with a 2-mm full-width at half maximum isotropic Gaussian kernel and its resulting mask binarized using a threshold of 0. For the second-level analysis, a group-level whole-brain mask was created by averaging the individual skull-stripped normalised anatomies. The average anatomy was then thresholded at 0.6, smoothed at 8-mm full-width, and finally binarized.

Anatomical labels of brain regions reported in the Supplementary Tables were taken from the Neuromorphometrics atlas (Neuromorphometrics, Inc.) implemented in SPM12 and the Harvard-Oxford Cortical and Subcortical Structural Atlases implemented in FSLeyes. In the case of white matter localizations, the respective brain lobe was provided for anatomical orientation.

Multivariate Classification Analysis

In addition to the RSA, we conducted a simpler multivariate classification approach without model-based hypotheses RDMs by using The Decoding Toolbox[9]. We used L2-norm support vector machines (SVM) from the library LIBSVM[10] and performed a classification

14

analysis. To answer the question whether a 50/50 morphed face is classified as the expected or unexpected face identity based on voxel-based activation patterns, we trained individual classifiers on the T-images of pairs of the neutral faces and tested them for each corresponding morph combination (AB, AC, AD, BC, BD, CD) across the four runs. The classifier's performance, i.e., classification score, was evaluated using classification accuracy minus chance (50%). We labelled the T-images of the test set so that classification scores larger than .5 indicate that a morph was classified as the expected face and classification scores below .5 indicate that a morph was classified as the unexpected face. Classification analyses were separately conducted for each participant, for each morph combination and each ROI separated by hemisphere and averaged across these to obtain a total classification score per participant per ROI. As classification scores were not normally distributed (Kolmogorov-Smirnow tests, all $p < .001$), we used two-sided Wilcoxon signed rank tests for all classification analyses.

Supplementary Results

Control Indices

To control whether participants were attentive and correctly performed the task, a Friedman test with the within-subject factor condition (match, mismatch, catch, neutral, neutral catch) and accuracy (%) as the dependent variable was calculated. We performed a non-parametric test instead of the preregistered ANOVA because the accuracies were not normally distributed due to ceiling effects (Kolmogorow-Smirnow tests, all $p < .001$). Match and mismatch trials were classified as correct if the presented (unmorphed) face was correctly identified in time. In catch trials, participants had to correctly answer the question mark by indicating which person they expected based on the preceding scene. Neutral trials were correct if participants pressed the button with the left thumb irrespective of which face was presented and neutral catch trials were correctly answered if they pressed the button with the left index finger to indicate that all persons were equally likely to be expected. Note that for the conditions mentioned, no morphs were presented.

Participants performed well in all experimental control conditions, especially in the match condition. There was a significant difference in accuracy ($\chi^2(4) = 32.17$, $p < .001$, Kendall's W = 0.19). Tukey's HSD test for multiple comparisons showed that participants significantly more often correctly identified a face in the match ($M = 97.64\%$, $SD = 3.98\%$) compared to the mismatch ($M = 90.94$, $SD = 11.77$, $p < .001$, LB/UB [0.62, 2.33]), catch ($M = 95.11$, $SD = 5.12$, $p = .049$, [0.003, 1.72]), and neutral catch condition ($M = 91.28$, $SD = 9.08$;

p < .001, [0.44, 2.16]). Participants also showed higher accuracy in the neutral compared to the neutral catch condition (p = .01, [0.13, 1.85]), possibly due to the more frequent occurrence of neutral compared to neutral catch trials. No difference in accuracy was observed between the match and neutral condition (M = 96.22, SD = 5.48; p = .86, [-0.54, 1.17]) nor in any other condition comparison.

Reaction Time Analysis

A Friedman test with the within-subject factor condition (match, mismatch, partial, neutral) and RT as the dependent variable was conducted. Additionally to the results reported in the main text, in the neutral condition, participants responded faster (M = 547.7, SD = 55.84) compared to all other conditions (match: p = .047, LB/UB [0.01, 1.44]; mismatch: p < .001, [1.70, 3.13]; partial: p < .001, [-3.02, -1.59]), because they simply had to press a fifth button for whichever face appeared after the neutral scene.

Multivariate ROI Analyses: Left vs. Right Hemisphere

In our main multivariate ROI analyses, we investigated the correlations between the hypothesis models and bilateral ROIs along the ventral face-processing hierarchy. We performed additional analyses to test whether the correlations differed between the left and the right hemispheres (see Figure 4g-j, Supplementary Figure 1). We tested for main effects and interactions between the factors hemisphere (left, right) and model (PE, Sharpening, Sensory Input) for each ROI and each deep neural network with the non-parametric Aligned Rank Transform (ART)[11] and calculated η_p^2 as a measure of effect size.

There was no significant main effect of hemisphere for any of the three neural networks (VGG-Face, VGG-16, ResNet50) in any of the ROIs (OFA, pFFA, aTL, MTG), i.e., the correlations of the neural data with the three hypothesis models (PE, Sharpening, Sensory Input) did not differ between the left and right hemisphere (all p(unc.) < .05; see Supplementary Tables 10-13).

There was a significant main effect of model for every neural network in every ROI (all p(unc.) < .05; see Supplementary Tables 10-13). For VGG-Face, post-hoc pairwise comparisons[12], Tukey-corrected, revealed that these main effects were mainly driven by higher correlation values with the PE model compared to the Sensory Input model in OFA and pFFA as well as higher correlation values for the PE model compared to the Sharpening model in pFFA, aTL, and ITG/MTG (see Supplementary Table 10, 11). The object-trained network VGG-16 showed a similar pattern, with stronger correlations of the PE model compared to the Sensory Input model in pFFA and aTL as well as stronger PE correlations compared to Sharpening correlations in all ROIs (see Supplementary Tables 10, 12). Additionally, there were stronger correlations of the Sensory Input model compared to the Sharpening model in

pFFA. The neural network ResNet50 showed similar higher correlations for the PE model, mainly compared to the Sharpening model in all ROIs, but contrary to the other object-trained network VGG-16, it did not reveal any evidence for stronger correlations with the Sensory Input model compared to the other hypothesis models (see Supplementary Tables 10, 13).

Multivariate Classification Analysis

In the classification analysis, ambiguous morphed faces were more often classified as the unexpected face identity based on multivariate activation patterns in the posterior FFA ($M = 46.22$, $SD = 10.83$, $z = -2.17$, $p = .03$), especially in the left FFA ($M = 45.54$, $SD = 12.25$, $z = -2.29$, $p = .02$; see Supplementary Figure 5). Classification as the expected or unexpected face did not differ in OFA ($M = 49.32$, $SD = 9.66$, $p = .72$, $z = -0.35$), aTL ($M = 52.33$, $SD = 7.73$, $p = .055$, $z = 1.92$), and MTG ($M = 50.19$, $SD = 10.48$, $p = .93$, $z = 0.09$). Hence, the classification analysis confirms the RSA results for PE representations in the FFA, while classification did not reveal either PE or Sharpening in the other ROIs. Overall, both multivariate analyses approaches, i.e., RSA and classification analysis, suggest concurrent PE and sharpened face representations. The major difference between the two different multivariate analyses approaches is that RSA allows us to test for both PE and sharpened face representations simultaneously in one region, while the classification approach is forced to classify morphed faces as either the expected or unexpected face. In addition, the RSA approach also takes the similarity of a morphed face with all neutral faces into account (see hypothesis RDMs in Figure 2b and Supplementary Figure 4, which are not empty in the off-diagonal), while the classification approach only compares a morphed face with the two corresponding neutral face images. Furthermore, the classification analysis we implemented here was a model-free approach. In contrast, our RSA approach relied on hypothesis RDMs generated from various DCNN layer activations, derived from images based on combinations of expected and presented face images with PE or sharpening computations.

Supplementary Tables

Supplementary Table 1. Univariate analysis of the contrast 'mismatch > match'.

cluster p(FWE)	cluster equivk	peak p(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.000	7236	0.000	11.04	7.52	-36 -42 48	left superior parietal lobule
		0.000	9.93	7.08	42 -32 48	right supramarginal gyrus
		0.000	9.65	6.97	-44 -32 48	
0.000	6702	0.000	10.87	7.46	-4 12 50	left supplementary motor cortex
		0.000	10.58	7.34	-30 24 -6	left anterior insula
		0.000	10.42	7.29	2 20 46	
0.000	3052	0.000	10.35	7.25	32 26 -4	right anterior insula
		0.000	9.87	7.06	34 24 4	
		0.009	6.15	5.16	50 8 38	
0.000	716	0.002	6.72	5.51	38 -52 -32	right cerebellum exterior
		0.010	6.11	5.14	26 -52 -26	
		0.014	5.99	5.06	30 -62 -30	
0.000	1755	0.003	6.49	5.37	-6 -16 12	left thalamus proper
		0.012	6.06	5.11	10 10 2	right caudate
		0.016	5.95	5.04	-8 -16 -12	
0.005	309	0.010	6.12	5.15	32 -74 -52	right cerebellum exterior
		0.992	3.74	3.46	36 -64 -42	
0.000	647	0.012	6.04	5.09	12 -76 -24	right cerebellum exterior
		0.041	5.61	4.82	-8 -78 -30	
		0.098	5.28	4.60	8 -82 -30	
0.015	246	0.054	5.50	4.75	-12 -56 -46	left cerebellum exterior
		0.383	4.70	4.19	-12 -54 -38	
		0.412	4.66	4.16	-2 -54 -38	
0.000	499	0.088	5.32	4.63	16 -60 -48	right cerebellum exterior
		0.092	5.30	4.61	28 -54 -54	
		0.469	4.59	4.11	16 -66 -60	
0.005	312	0.104	5.25	4.58	-56 -42 -18	left inferior temporal gyrus
		0.856	4.14	3.77	-68 -38 -16	

0.000	479	0.115	5.21	4.55	-38 -56 -32	left cerebellum exterior
		0.210	4.97	4.38	-36 -68 -28	
		0.310	4.80	4.26	-28 -54 -32	
0.001	403	0.136	5.15	4.51	22 54 -14	right anterior orbital gyrus
		0.932	4.00	3.66	42 48 -16	
		0.976	3.86	3.55	34 50 -18	

Note. For the univariate contrast 'mismatch > match', a one-sample *t*-test was performed (*N* = 43 participants). Clusters at p(FWE) > .05, with a cluster-inducing threshold of p(unc.) < .001, are reported (k > 10 voxels).

Supplementary Table 2. Univariate analysis of the contrast 'unexpected > expected'.

cluster p(FWE)	cluster equivk	peak p(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.000	13699	0.000	9.49	6.90	-36 -36 44	left superior parietal lobule
		0.000	8.71	6.55	20 -52 -22	
		0.000	8.35	6.38	-8 -62 46	
0.000	1829	0.000	7.73	6.07	40 -44 48	right superior parietal lobule
		0.000	7.27	5.82	40 -34 48	
		0.190	4.97	4.39	38 -56 46	
0.000	4469	0.000	7.28	5.82	26 -6 52	right superior frontal gyrus
		0.001	6.93	5.63	-24 -6 52	
		0.002	6.61	5.44	-4 36 24	left anterior cingulate gyrus
0.000	5723	0.001	6.91	5.62	10 -6 8	right thalamus proper
		0.003	6.50	5.38	10 -14 6	
		0.003	6.49	5.37	34 20 -2	right anterior insula
0.000	2271	0.029	5.70	4.88	-42 48 16	left middle frontal gyrus
		0.033	5.65	4.85	-20 50 -18	
		0.060	5.43	4.70	-28 54 -12	left anterior orbital gyrus
0.000	635	0.092	5.26	4.59	30 46 -18	right anterior orbital gyrus
		0.150	5.07	4.45	34 60 -10	
		0.463	4.56	4.09	38 44 -4	
0.001	444	0.197	4.96	4.37	40 36 20	right frontal pole
		0.336	4.72	4.21	44 34 30	right middle frontal gyrus
		0.781	4.20	3.82	46 36 38	
0.031	223	0.214	4.92	4.35	54 -32 -12	right middle temporal gyrus
		0.227	4.90	4.33	56 -40 -14	
0.004	348	0.510	4.51	4.05	-56 -40 -14	left middle temporal gyrus
		0.706	4.29	3.88	-54 -26 -26	
		0.991	3.71	3.43	-58 -18 -30	

Note. For the univariate contrast 'unexpected > expected', a one-sample *t*-test was performed (N = 43 participants). Clusters at p(FWE) > .05, with a cluster-inducing threshold of p(unc.) < .001, are reported (k > 10 voxels).

Supplementary Table 3. Means and standard error of the means (SEM) of the multivariate ROI analyses (VGG-Face).

ROI	PE			Sharpening			Input		
	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)
bilateral									
OFA	0.0553	0.0178	0.0136	0.0283	0.0131	0.0110	0.0096	0.0133	0.0087
pFFA	0.0408	0.0158	0.0133	-0.0108	0.0138	0.0173	-0.0006	0.0121	0.0101
aTL	0.0341	0.0195	0.0144	-0.0172	0.0141	0.0150	-0.0131	0.0149	0.0120
ITG/ MTG	0.0418	0.0177	0.0167	-0.0074	0.0141	0.0142	0.0005	0.0145	0.0116
unilateral									
lOFA	0.0552	0.0172	0.0131	0.0236	0.0134	0.0103	0.0196	0.0148	0.0095
rOFA	0.0517	0.0178	0.0135	0.0241	0.0128	0.0112	0.0062	0.0115	0.0075
lpFFA	0.0426	0.0133	0.0131	-0.0213	0.0134	0.0158	-0.0016	0.0117	0.0096
rpFFA	0.0240	0.0197	0.0166	-0.0182	0.0146	0.0176	-0.0136	0.0136	0.0117
laTL	0.0481	0.0188	0.0139	-0.0086	0.0153	0.0154	0.0068	0.0165	0.0114
raTL	0.0146	0.0170	0.0156	-0.0053	0.0129	0.0140	-0.0123	0.0114	0.0101
lITG/ MTG	0.0212	0.0165	0.0156	-0.0067	0.0160	0.0169	-0.0080	0.0158	0.0112
rITG/ MTG	0.0353	0.0177	0.0151	-0.009	0.0134	0.0154	0.0002	0.0131	0.0112

Note. This table provides the means and SEMs (N = 43) for the DCNN VGG-Face in Figure 4. Individual correlation coefficients (Kendall's Tau A) between the representational dissimilarity matrices (RDM) of the three hypothesis models (PE, Sharpening, Sensory Input) and the neural RDM were calculated. Correlations for the occipital face area (OFA) are based on pool4 activations, the correlations for the other ROIs on pool5 activations. Within-subject (ws) SEMs were calculated considering the individuals' correlation coefficients across the two compared DCNNs (VGG-Face and VGG-16). For *p*-values of the model tests against zero and the model comparison tests, please refer to the Supplementary Table 4. pFFA = posterior fusiform face area; aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus; bw = between-subject; ws = within-subject.

Supplementary Table 4. P-values of the multivariate ROI analyses (VGG-Face).

ROI	PE vs. 0	Sharpening vs. 0	Input vs. 0	PE vs. Sharpening	PE vs. Input	Sharpening vs. Input
bilateral						
OFA	0.0030*	0.0232*	0.2913	0.2941	0.0396*	0.2941
pFFA	0.0044*	0.5528	0.4140	0.0315*	0.0315*	0.6311
aTL	0.0780	0.8462	0.7838	0.0460*	0.0154*	0.7766
ITG/ MTG	0.0168*	0.8641	0.5457	0.0366*	0.0366*	0.8468
unilateral						
lOFA	0.0013*	0.0319*	0.0961	0.1624	0.1624	0.7766
rOFA	0.0048*	0.0321*	0.3451	0.3042	0.0424*	0.3917
lpFFA	0.0016*	0.8729	0.5866	0.0044*	0.0192*	0.2004
rpFFA	0.2381	0.8841	0.8242	0.1460	0.0778	0.9278
laTL	0.0206*	0.7820	0.3145	0.0202*	0.0202*	0.2746
raTL	0.2322	0.6482	0.9432	0.4373	0.2529	0.9615
lITG/ MTG	0.1256	0.7620	0.7542	0.3443	0.3443	0.5232
rITG/ MTG	0.0263*	0.8847	0.4726	0.0928	0.0928	0.4826

Note. This table reports the precise *p*-values for the DCNN VGG-Face in Figure 4 and corresponds to the data in Supplementary Table 3. We tested for the significance of the hypothesis model correlations against a null correlation using one-sided Wilcoxon signed-rank tests. Model comparisons were tested using paired, two-sided Wilcoxon signed rank tests. For the tests against zero, a black asterisk indicates Bonferroni-corrected significance considering the number of tests per ROI ($N = 6$ (3 models x 2 DCNNs); $p < .0083$), a grey asterisk indicates uncorrected significance (p(unc.) $< .001$). For the model comparisons, an asterisk indicates FDR-corrected[4] significance for the model comparisons per ROI. OFA = occipital face area; pFFA = posterior fusiform face area aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus.

Supplementary Table 5. Means and standard error of the means (SEM) of the multivariate ROI analyses (VGG-16).

ROI	PE			Sharpening			Input		
	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)
bilateral									
OFA	0.0623	0.0191	0.0173	-0.0146	0.019	0.0216	0.0206	0.0142	0.0091
pFFA	0.101	0.0245	0.0223	-0.0099	0.0197	0.0193	0.0413	0.0207	0.0154
aTL	0.0963	0.0219	0.0205	-0.0396	0.0153	0.0163	0.0272	0.0183	0.0135
ITG/ MTG	0.0760	0.0225	0.0209	-0.0127	0.0193	0.0188	0.0357	0.0211	0.0144
unilateral									
lOFA	0.0523	0.0179	0.0161	-0.0124	0.0172	0.0191	0.0284	0.0147	0.0087
rOFA	0.0535	0.0202	0.0178	-0.0076	0.0186	0.0223	0.0172	0.0137	0.0096
lpFFA	0.0971	0.0233	0.0212	-0.0055	0.0181	0.0172	0.0460	0.0223	0.0159
rpFFA	0.0843	0.0261	0.0235	-0.0205	0.0197	0.0206	0.0212	0.0208	0.0152
laTL	0.0831	0.0206	0.0194	-0.0266	0.0163	0.0186	0.0254	0.0176	0.0129
raTL	0.0572	0.0176	0.0163	-0.0162	0.0175	0.0147	0.0236	0.0202	0.0143
lITG/ MTG	0.0531	0.0248	0.0243	-0.0146	0.0212	0.0203	0.0153	0.0222	0.0151
rITG/ MTG	0.0662	0.0226	0.0207	-0.0037	0.0181	0.0187	0.0373	0.0198	0.0137

Note. This table provides the means and SEMs (N = 43) for the DCNN VGG-16 in Figure 4. Individual correlation coefficients (Kendall's Tau A) between the representational dissimilarity matrices (RDM) of the three hypothesis models (PE, Sharpening, Sensory Input) and the neural RDM were calculated. Correlations for the occipital face area (OFA) are based on pool4 activations, the correlations for the other ROIs on pool5 activations. Within-subject (ws) SEMs were calculated considering the individuals' correlation coefficients across the two compared DCNNs (VGG-Face and VGG-16). For p-values of the model tests against zero and the model comparison tests, please refer to the Supplementary Table 6. pFFA = posterior fusiform face area; aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus; bw = between-subject; ws = within-subject.

Supplementary Table 6. P-values of the multivariate ROI analyses (VGG-16).

ROI	PE vs. 0	Sharpening vs. 0	Input vs. 0	PE vs. Sharpening	PE vs. Input	Sharpening vs. Input
bilateral						
OFA	0.0025*	0.7045	0.0611	0.1351	0.1471	0.1471
pFFA	0.0002*	0.6322	0.0379*	0.0086*	0.0315*	0.0433*
aTL	0.0001*	0.9871	0.0901	0.0010*	0.0047*	0.0026*
ITG/ MTG	0.0038*	0.6227	0.0797	0.0366*	0.1922	0.0366*
unilateral						
lOFA	0.0032*	0.8093	0.0374*	0.1410	0.3881	0.1379
rOFA	0.0099*	0.5096	0.1148	0.3339	0.3042	0.4826
lpFFA	0.0002*	0.6637	0.0249*	0.0050*	0.1206	0.0244*
rpFFA	0.0018*	0.7270	0.1509	0.0395*	0.0395*	0.2551
laTL	0.0003*	0.9221	0.0428*	0.0092*	0.0202*	0.0202*
raTL	0.0017*	0.8177	0.1956	0.0313*	0.1575	0.0630
lITG/ MTG	0.0206*	0.7506	0.3451	0.3443	0.3443	0.3443
rITG/ MTG	0.0051*	0.6459	0.0411*	0.0928	0.4066	0.0928

Note. This table reports the precise *p*-values for the DCNN VGG-16 in Figure 4 and corresponds to the data in Supplementary Table 5. We tested for the significance of the hypothesis model correlations against a null correlation using one-sided Wilcoxon signed-rank tests. Model comparisons were tested using paired, two-sided Wilcoxon signed rank tests. For the tests against zero, a black asterisk indicates Bonferroni-corrected significance considering the number of tests per ROI (N = 6 (3 models x 2 DCNNs); p < .0083), a grey asterisk indicates uncorrected significance (p(unc.) < .001). For the model comparisons, an asterisk indicates FDR-corrected[4] significance for the model comparisons per ROI. OFA = occipital face area; pFFA = posterior fusiform face area aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus.

Supplementary Table 7. P-values of the multivariate ROI analyses (VGG-Face vs. VGG-16).

ROI	PE vs. PE	Sharpening vs. Sharpening	Input vs. Input
bilateral			
OFA	0.9134	0.0396*	0.1471
pFFA	0.0315*	0.8185	0.0715
aTL	0.0176*	0.2970	0.0837
ITG/MTG	0.3105	0.8468	0.1212
unilateral			
lOFA	0.6743	0.1379	0.1624
rOFA	0.6799	0.3042	0.2506
lpFFA	0.0192*	0.3979	0.0192*
rpFFA	0.0778	0.9278	0.2178
laTL	0.1247	0.4762	0.3965
raTL	0.1556	0.9247	0.1575
lITG/ MTG	0.3702	0.8232	0.3443
rITG/ MTG	0.4066	0.6726	0.1029

Note. This table reports the precise *p*-values for the model comparisons of the DCNNs VGG-Face and VGG-16 in Figure 4 and corresponds to the data in Supplementary Table 3 and 5. An asterisk indicates FDR-corrected[4] significance for the model comparisons per ROI. OFA = occipital face area; pFFA = posterior fusiform face area aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus.

Supplementary Table 8. Means and standard error of the means (SEM) of the multivariate ROI analyses (ResNet50).

ROI	PE			Sharpening			Input		
	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)	mean	SEM (bw)	SEM (ws)
bilateral									
OFA	0.0355	0.0240	0.0253	-0.0264	0.0156	0.0196	0.0048	0.0132	0.0098
pFFA	0.0521	0.0214	0.0247	-0.0347	0.0156	0.0184	0.0131	0.0130	0.0098
aTL	0.0484	0.0204	0.0207	-0.0475	0.0146	0.0155	-0.0107	0.0153	0.0093
ITG/ MTG	0.0542	0.0187	0.0189	-0.0465	0.0133	0.0147	-0.0053	0.0143	0.0085
unilateral									
lOFA	0.0313	0.0221	0.0242	-0.0171	0.0160	0.0177	0.0107	0.0136	0.0095
rOFA	0.0292	0.0256	0.0255	-0.0295	0.0148	0.0203	0.0014	0.0123	0.0086
lpFFA	0.0478	0.0181	0.0199	-0.0213	0.0131	0.0148	0.0087	0.0120	0.0083
rpFFA	0.0500	0.0236	0.0263	-0.0491	0.0170	0.0204	-0.0012	0.0135	0.0096
laTL	0.0363	0.0218	0.0217	-0.0290	0.0161	0.0175	0.0089	0.0135	0.0077
raTL	0.0281	0.0181	0.0201	-0.0309	0.0138	0.0140	-0.0183	0.0140	0.0088
lITG/ MTG	0.0401	0.0196	0.0223	-0.0419	0.0164	0.0180	-0.0135	0.0146	0.0089
rITG/ MTG	0.0482	0.0197	0.0208	-0.0370	0.0148	0.0159	-0.0093	0.0137	0.0082

Note. This table provides the means and SEMs (*N* = 43) for the DCNN ResNet50 in Supplementary Figure 1. Individual correlation coefficients (Kendall's Tau A) between the representational dissimilarity matrices (RDM) of the three hypothesis models (PE, Sharpening, Sensory Input) and the neural RDM were calculated. Correlations were based on res5b_branch2b activations. For *p*-values of the model tests against zero and the model comparison tests, please refer to the Supplementary Table 9. OFA = occipital face area; pFFA = posterior fusiform face area; aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus; bw = between-subject; ws = within-subject.

Supplementary Table 9. P-values of the multivariate ROI analyses (ResNet50).

ROI	PE vs. 0	Sharpening vs. 0	Input vs. 0	PE vs. Sharpening	PE vs. Input	Sharpening vs. Input
bilateral						
OFA	0.1011	0.9522	0.3399	0.1726	0.3310	0.0420*
pFFA	0.0155*	0.9781	0.1171	0.0327*	0.1359	0.0038*
aTL	0.0151*	0.9981	0.7132	0.0109*	0.0204*	0.0110*
ITG/ MTG	0.0064*	0.9988	0.7350	0.0014*	0.0062*	0.0014*
unilateral						
lOFA	0.1003	0.8547	0.1716	0.3008	0.6420	0.0203*
rOFA	0.1625	0.9659	0.3496	0.2502	0.4116	0.1255
lpFFA	0.0111*	0.9506	0.2381	0.0342*	0.0902	0.0284*
rpFFA	0.0503	0.9935	0.4952	0.0366*	0.1490	0.0121*
laTL	0.0267*	0.9432	0.3124	0.0894	0.3556	0.0413*
raTL	0.1057	0.9873	0.9367	0.0496*	0.0496*	0.1413
lITG/ MTG	0.0245*	0.9863	0.8921	0.0621	0.0621	0.0621
rITG/ MTG	0.0129*	0.9797	0.6615	0.0199*	0.0199*	0.0201*

Note. This table reports the precise *p*-values for the DCNN ResNet50 in Supplementary Figure 1 and corresponds to the data in Supplementary Table 8. We tested for the significance of the hypothesis model correlations against a null correlation using one-sided Wilcoxon signed-rank tests. Model comparisons were tested using paired, two-sided Wilcoxon signed rank tests. For the tests against zero, a black asterisk indicates Bonferroni-corrected significance considering the number of tests per ROI ($N = 3$; $p < .017$), a grey asterisk indicates uncorrected significance (p(unc.) $< .001$). For the model comparisons, an asterisk indicates FDR-corrected[4] significance for the model comparisons per ROI. OFA = occipital face area; pFFA = posterior fusiform face area aTL = anterior temporal lobe; ITG = inferior temporal gyrus; MTG = middle temporal gyrus.

Supplementary Table 10. Means and standard deviations for the main effects of the ROI analyses split-up by hemisphere.

VGG-Face

ROI	PE		Sharpening		Input		left		right	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
OFA	0,0534	0.1143	0.0239	0.0852	0.0129	0.0866	0.0328	0.1001	0.0273	0.0949
pFFA	0.0333	0.1100	-0.0198	0.0915	-0.0076	0.0829	0.0065	0.0875	-0.0026	0.1072
aTL	0.0313	0.1182	-0.0069	0.0923	-0.0028	0.0929	0.0154	0.1126	-0.0010	0.0918
ITG/ MTG	0.0282	0.1118	-0.0079	0.0961	-0.0039	0.0944	0.0021	0.1054	0.0089	0.0987

VGG-16

ROI	PE		Sharpening		Input		left		right	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
OFA	0.0529	0.1243	-0.0100	0.1168	0.0228	0.0930	0.0228	0.1117	0.0210	0.1179
pFFA	0.0907	0.1614	-0.0130	0.1235	0.0336	0.1411	0.0459	0.1452	0.0283	0.1518
aTL	0.0702	0.1255	-0.0214	0.1101	0.0245	0.1237	0.0273	0.1269	0.0215	0.1239
ITG/ MTG	0.0597	0.1547	-0.0092	0.1285	0.0263	0.1377	0.0179	0.1507	0.0333	0.1349

ResNet50

ROI	PE		Sharpening		Input		left		right	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
OFA	0.0302	0.1558	-0.0233	0.1008	0.0061	0.0845	0.0083	0.1161	0.0003	0.1227
pFFA	0.0489	0.1370	-0.0352	0.1001	0.0037	0.0832	0.0117	0.0994	-0.0001	0.1271
aTL	0.0322	0.1307	-0.0299	0.0978	-0.0047	0.0907	0.0054	0.1170	-0.0071	0.1035
ITG/ MTG	0.0442	0.1280	-0.0395	0.1017	-0.0114	0.0921	-0.0051	0.1155	0.0006	0.1116

Note. PE = Prediction Error, ROI = region of interest, *M* = mean, *SD* = standard deviation, left = left hemisphere, right = right hemisphere, OFA = occipital face area, pFFA = posterior fusiform face area, aTL = anterior temporal lobe, ITG = inferior temporal gyrus, MTG = middle temporal gyrus.

Supplementary Table 11. Main effects and post-hoc tests for the ROI analyses split-up by hemisphere (VGG-Face).

Main effects and interaction of factors 'hemisphere' and 'model'

ROI	factor	df	SS	MS	F	p	$\eta_p{}^2$
OFA	hemisphere	1	3852.7481	3852.7481	0.9359	0.3344	0.0044
	model	2	39357.0233	19678.5116	4.9035	0.0083*	0.0446
	hemisphere:model	2	2557.5891	1278.7946	0.3097	0.7340	0.0029
pFFA	hemisphere	1	7370.7016	7370.7016	1.5714	0.2114	0.0074
	model	2	64890.3023	32445.1512	7.2968	0.0009*	0.0650
	hemisphere:model	2	4555.5891	2277.7946	0.4836	0.6173	0.0046
aTL	hemisphere	1	6763.7248	6763.7248	1.6550	0.1997	0.0078
	model	2	27796.4651	13898.2326	3.4311	0.0342*	0.0316
	hemisphere:model	2	8146.6589	4073.3295	1.0094	0.3662	0.0095
MTG	hemisphere	1	1937.3992	1937.3992	0.4857	0.4866	0.0023
	model	2	36569.1860	18284.5930	4.7405	0.0097*	0.0432
	hemisphere:model	2	4162.7674	2081.3837	0.5261	0.5917	0.0050

Post-hoc tests for main effect 'model'

ROI	contrast	df	SE	t-ratio	p
OFA	Input - PE	210	9.661	-3.0139	0.0081*
	Input - Sharpening	210	9.661	-0.7703	0.7216
	PE - Sharpening	210	9.661	2.2436	0.0664
pFFA	Input - PE	210	10.169	-2.8873	0.0119*
	Input - Sharpening	210	10.169	0.7227	0.7503
	PE - Sharpening	210	10.169	3.6100	0.0011*
aTL	Input - PE	210	9.706	-2.0127	0.1116
	Input - Sharpening	210	9.706	0.4457	0.8964
	PE - Sharpening	210	9.706	2.4584	0.0390*
MTG	Input - PE	210	9.471	-2.3511	0.0512
	Input - Sharpening	210	9.471	0.5463	0.8485
	PE - Sharpening	210	9.471	2.8975	0.0115*

Note. We tested the correlation values (Kendall's Tau A) for each region of interest (ROI) for main effects and interactions between the factors hemisphere (left, right) and model (PE, Sharpening, Sensory Input) using the non-parametric Aligned Rank Transform (ART)[11]. An asterisk indicates significance (p(unc.) < .05 for main effects and interaction, Tukey-corrected for post-hoc tests[12]). *Df* = degrees of freedom, *SS* = sum of squares, *MS* = mean square, *SE* = standard error of the mean, ROI = region of interest, OFA = occipital face area, pFFA = posterior fusiform face area, aTL = anterior temporal lobe, MTG = middle temporal lobe.

Supplementary Table 12. Main effects and post-hoc tests for the ROI analyses split-up by hemisphere (VGG-16).

Main effects and interaction of factors 'hemisphere' and 'model'

ROI	factor	df	SS	MS	F	p	η_p^2
OFA	hemisphere	1	240.3140	240.3140	0.0440	0.8340	0.0002
	model	2	56799.2326	28399.6163	5.4971	0.0047*	0.0497
	hemisphere:model	2	3831.9767	1915.9884	0.3514	0.7041	0.0033
pFFA	hemisphere	1	5242.5155	5242.5155	1.2246	0.2697	0.0058
	model	2	104916.3488	52458.1744	13.2052	0.0000*	0.1117
	hemisphere:model	2	2173.0465	1086.5233	0.2524	0.7772	0.0024
aTL	hemisphere	1	3132.5620	3132.5620	0.7089	0.4008	0.0034
	model	2	130361.0465	65180.5233	16.3647	0.0000*	0.1348
	hemisphere:model	2	5105.0930	2552.5465	0.5796	0.5610	0.0055
MTG	hemisphere	1	5152.7481	5152.7481	1.1194	0.2913	0.0053
	model	2	46720.4012	23360.2006	5.2176	0.0061*	0.0473
	hemisphere:model	2	1073.5814	536.7907	0.1162	0.8903	0.0011

Post-hoc tests for main effect 'model'

ROI	contrast	df	SE	t-ratio	p
OFA	Input - PE	210	10.9611	-1.5647	0.2633
	Input - Sharpening	210	10.9611	1.7493	0.1895
	PE - Sharpening	210	10.9611	3.3140	0.0031*
pFFA	Input - PE	210	9.6117	-2.5804	0.0283*
	Input - Sharpening	210	9.6117	2.5587	0.0300*
	PE - Sharpening	210	9.6117	5.1391	0.0000*
aTL	Input - PE	210	9.6243	-2.8151	0.0147*
	Input - Sharpening	210	9.6243	2.9057	0.0113*
	PE - Sharpening	210	9.6243	5.7207	0.0000*
MTG	Input - PE	210	10.2040	-1.6535	0.2257
	Input - Sharpening	210	10.2040	1.5766	0.2580
	PE - Sharpening	210	10.2040	3.2300	0.0041*

Note. We tested the correlation values (Kendall's Tau A) for each region of interest (ROI) for main effects and interactions between the factors hemisphere (left, right) and model (PE, Sharpening, Sensory Input) using the non-parametric Aligned Rank Transform (ART)[11]. An asterisk indicates significance (p(unc.) < .05 for main effects and interaction, Tukey-corrected for post-hoc tests[12]). *Df* = degrees of freedom, *SS* = sum of squares, *MS* = mean square, *SE* = standard error of the mean, ROI = region of interest, OFA = occipital face area, pFFA = posterior fusiform face area, aTL = anterior temporal lobe, MTG = middle temporal lobe.

Supplementary Table 13. Main effects and post-hoc tests for the ROI analyses split-up by hemisphere (ResNet50).

Main effects 'model' and 'hemisphere'							
ROI	**factor**	**df**	**SS**	**MS**	**F**	**p**	**η_P^2**
OFA	hemisphere	1	1344.6550	1344.6550	0.2529	0.6156	0.0012
	model	2	43267.9767	21633.9884	4.2061	0.0162*	0.0385
	hemisphere:model	2	192.1473	96.0736	0.0180	0.9822	0.0002
pFFA	hemisphere	1	5260.5620	5260.5620	0.9338	0.3350	0.0044
	model	2	99196.3488	49598.1744	9.5595	0.0001*	0.0834
	hemisphere:model	2	2333.4961	1166.7481	0.2075	0.8128	0.0020
aTL	hemisphere	1	12909.3992	12909.3992	2.6021	0.1082	0.0122
	model	2	71736.9302	35868.4651	7.6288	0.0006*	0.0677
	hemisphere:model	2	2907.1705	1453.5853	0.2915	0.7475	0.0028
MTG	hemisphere	1	3760.5620	3760.5620	0.7321	0.3932	0.0035
	model	2	111785.9535	55892.9767	11.8647	0.0000*	0.1015
	hemisphere:model	2	269.3256	134.6628	0.0261	0.9742	0.0002

Post-hoc tests for main effect 'model'					
ROI	**contrast**	**df**	**SE**	**t-ratio**	**p**
OFA	Input - PE	210	10.9369	-0.7995	0.7037
	Input - Sharpening	210	10.9369	2.0147	0.1112
	PE - Sharpening	210	10.9369	2.8142	0.0147*
pFFA	Input - PE	210	10.9845	-1.8493	0.1563
	Input - Sharpening	210	10.9845	2.5067	0.0344
	PE - Sharpening	210	10.9845	4.3560	0.0001*
aTL	Input - PE	210	10.4567	-2.2329	0.0681
	Input - Sharpening	210	10.4567	1.6591	0.2235
	PE - Sharpening	210	10.4567	3.8920	0.0004*
MTG	Input - PE	210	10.4669	-3.1972	0.0045*
	Input - Sharpening	210	10.4669	1.5842	0.2547
	PE - Sharpening	210	10.4669	4.7814	0.0000*

Note. We tested the correlation values (Kendall's Tau A) for each region of interest (ROI) for main effects and interactions between the factors hemisphere (left, right) and model (PE, Sharpening, Sensory Input) using the non-parametric Aligned Rank Transform (ART)[11]. An asterisk indicates significance (p(unc.) < .05 for main effects and interaction, Tukey-corrected for post-hoc tests[12]). Df = degrees of freedom, SS = sum of squares, MS = mean square, SE = standard error of the mean, ROI = region of interest, OFA = occipital face area, pFFA = posterior fusiform face area, aTL = anterior temporal lobe, MTG = middle temporal lobe.

Supplementary Table 14. Searchlight analysis for the hypothesis model Prediction Error (PE) (*p*(FWE) < .05) (VGG-Face, pool4).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	5841	6.61	5.44	52 -54 28	right angular gyrus
0.002		5.82	4.96	48 -58 -2	right middle temporal gyrus
0.004		5.50	4.75	32 -60 22	right lateral occipital cortex, superior division
0.000	10955	6.58	5.43	32 -8 48	right precentral gyrus
0.001		6.14	5.16	48 18 14	right inferior frontal gyrus, pars opercularis
0.002		5.83	4.96	12 22 40	right paracingulate gyrus
0.003	7916	5.67	4.86	-54 -54 46	left angular gyrus
0.003		5.61	4.82	-26 -86 26	left superior occipital gyrus
0.006		5.39	4.67	-44 -60 -4	left inferior temporal gyrus, temporooccipital part
0.010	157	5.17	4.52	-46 -24 -14	left middle temporal gyrus, posterior division
0.013	585	5.07	4.45	-44 14 8	left inferior frontal gyrus, pars opercularis
0.027		4.81	4.27	-34 26 26	left middle frontal gyrus
0.015	207	5.03	4.43	-42 26 48	left middle frontal gyrus
0.033		4.73	4.21	-30 10 56	
0.039		4.66	4.16	-30 16 50	
0.030	30	4.76	4.23	44 50 22	right middle frontal gyrus
0.032	65	4.74	4.22	56 -16 -16	right middle temporal gyrus, posterior division
0.042		4.63	4.14	60 -12 -30	
0.035	8	4.71	4.19	20 2 58	right superior frontal gyrus
0.037	14	4.68	4.17	-48 16 -14	left temporal pole
0.040	16	4.66	4.16	30 -28 -26	right fusiform gyrus
0.043	9	4.62	4.13	-38 58 6	left middle frontal gyrus
0.047	5	4.59	4.11	-4 56 12	left superior frontal gyrus medial segment
0.048	1	4.58	4.10	-18 10 -6	left putamen
0.048	3	4.58	4.10	30 14 2	right insular cortex
0.050	1	4.57	4.09	28 -22 -16	right hippocampus

Supplementary Table 15. Searchlight analysis for the hypothesis model Prediction Error (PE) (*p*(FWE) < .05) (VGG-Face, pool5).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	1498	6.95	5.64	30 -8 46	right precentral gyrus
0.004		5.54	4.77	34 -14 14	right posterior insula
0.031		4.79	4.26	20 -2 58	right superior frontal gyrus
0.002	1071	5.74	4.90	12 24 40	right paracingulate gyrus
0.011	463	5.19	4.54	-54 -54 44	left angular gyrus
0.031		4.79	4.26	-50 -64 32	
0.012	438	5.14	4.50	28 -66 28	right lateral occipital cortex, superior division
0.015		5.06	4.44	32 -60 22	
0.017		5.03	4.42	46 -54 28	right angular gyrus
0.014	264	5.08	4.46	48 16 14	right inferior frontal gyrus, pars opercularis
0.030		4.80	4.26	56 10 40	right precentral gyrus
0.040		4.69	4.18	50 6 24	
0.015	155	5.07	4.46	32 12 4	right insular cortex
0.016	54	5.04	4.43	-46 -24 -14	left middle temporal gyrus, posterior division
0.018	74	4.99	4.40	-28 8 54	left middle frontal gyrus
0.019	67	4.97	4.38	54 -16 -16	right middle temporal gyrus, posterior division
0.021	663	4.94	4.36	-28 -74 32	left lateral occipital cortex, superior division
0.029		4.82	4.28	-24 -90 34	left occipital pole
0.029		4.82	4.27	-34 -86 34	left middle occipital gyrus
0.028	39	4.83	4.28	-64 -58 -2	left middle temporal gyrus, temporooccipital part
0.034	7	4.76	4.23	-42 26 48	left middle frontal gyrus
0.037	28	4.73	4.21	-22 -62 -24	left cerebellum exterior
0.044	9	4.66	4.16	-40 14 12	left frontal operculum cortex
0.049	1	4.61	4.13	-10 -66 26	left precuneus
0.049	1	4.61	4.13	2 -14 36	right middle cingulate gyrus

Supplementary Table 16. Difference searchlight results for 'Prediction Error (PE) minus Sensory' (*p*(FWE) < .05) (VGG-Face, pool4 minus conv1_2).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	34547	6.92	5.62	50 4 42	right precentral gyrus
0.000		6.80	5.55	52 -46 36	right angular gyrus
0.000		6.73	5.51	-2 14 40	left middle cingulate gyrus
0.008	2080	5.30	4.61	0 -60 14	left precuneus
0.008		5.28	4.60	-6 -50 52	
0.009		5.23	4.56	-10 -48 22	left cingulate gyrus, posterior division
0.019	106	4.96	4.38	-46 -20 -14	left middle temporal gyrus, posterior division
0.024	93	4.88	4.32	-34 -54 -60	left cerebellum exterior
0.047	1	4.61	4.12	30 4 -32	right parahippocampal gyrus, anterior division
0.049	1	4.60	4.12	32 2 -30	right parahippocampal gyrus, anterior division
0.049	2	4.60	4.11	-38 54 4	left middle frontal gyrus

Supplementary Table 17. Difference searchlight results for 'Prediction Error (PE) minus Sensory' (*p*(FWE) < .05) (VGG-Face, pool5 minus conv1_2).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	2511	6.87	5.59	-2 12 42	left middle cingulate gyrus
0.009		5.27	4.59	2 -14 40	right middle cingulate gyrus
0.013		5.13	4.50	8 -16 46	right supplementary motor cortex
0.001	3221	6.18	5.18	32 -6 48	right precentral gyrus
0.001		5.94	5.03	42 10 -2	right anterior insula
0.002		5.84	4.97	52 4 40	
0.004	639	5.60	4.81	42 -46 22	right angular gyrus
0.045		4.66	4.16	42 -64 12	right lateral occipital cortex, inferior division
0.005	539	5.44	4.71	-24 -86 30	left superior occipital gyrus
0.008	349	5.33	4.63	-56 -52 46	left supramarginal gyrus, posterior division
0.012	466	5.17	4.52	34 24 48	right middle frontal gyrus
0.026		4.87	4.31	30 32 24	
0.013	1013	5.12	4.49	-46 14 18	left inferior frontal gyrus, pars opercularis
0.018		5.01	4.41	-34 24 50	left middle frontal gyrus
0.023		4.93	4.35	-38 26 32	
0.014	94	5.10	4.47	-48 16 -16	left temporal pole
0.015	176	5.08	4.46	54 -18 -16	right middle temporal gyrus, posterior division
0.016	112	5.04	4.43	-50 -22 -18	left middle temporal gyrus, posterior division
0.022	406	4.93	4.36	-44 -66 -2	left inferior occipital gyrus
0.032		4.80	4.26	-56 -72 6	left lateral occipital cortex, inferior division
0.037		4.74	4.22	-62 -64 -2	left middle temporal gyrus, temporooccipital part
0.024	135	4.90	4.34	-52 -68 30	left angular gyrus
0.046	2	4.65	4.15	-14 -62 -48	left cerebellum exterior

Supplementary Table 18. Searchlight analysis for the hypothesis model Sharpening (p(FWE) < .05) (VGG-Face, pool4).

peak p(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.027	20	4.88	4.32	-6 68 14	left frontal pole
0.027	35	4.88	4.32	-52 -22 50	left postcentral gyrus
0.037	23	4.77	4.24	0 26 2	left subcallosal cortex

Supplementary Table 19. Difference searchlight results for 'Prediction Error (PE) minus Sharpening' (*p*(FWE) < .05) (VGG-Face, pool5) .

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.003	380	5.65	4.84	12 22 42	right paracingulate gyrus
0.043		4.71	4.20	-4 12 44	left supplementary motor cortex
0.007	75	5.37	4.66	22 -6 56	right superior frontal gyrus
0.008	302	5.34	4.64	40 -46 22	right supramarginal gyrus, posterior division
0.016		5.09	4.47	32 -54 26	right parietal lobe
0.010	559	5.26	4.58	-16 -64 -22	left cerebellum exterior
0.013		5.15	4.51	-26 -66 -20	
0.015	148	5.12	4.48	-54 -50 46	left supramarginal gyrus
0.017	117	5.05	4.44	-48 10 8	left inferior frontal gyrus, pars opercularis
0.021	50	4.98	4.39	30 36 26	right middle frontal gyrus
0.025	37	4.91	4.34	-38 -56 -58	left cerebellum exterior
0.030	44	4.85	4.30	52 2 48	right precentral gyrus
0.040		4.74	4.22	48 -2 38	
0.033	19	4.81	4.27	54 -18 -16	right middle temporal gyrus
0.050	1	4.65	4.16	-60 8 8	left precentral gyrus

Supplementary Table 20. Searchlight analysis for the hypothesis model Prediction Error (PE) (*p*(FWE) < .05) (VGG-16, pool4).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	1988	6.69	5.49	34 -16 48	right precentral gyrus
0.004		5.53	4.77	34 -14 20	right insular cortex
0.001	2440	6.24	5.22	-40 28 48	left middle frontal gyrus
0.013		5.09	4.46	-48 16 -14	left temporal pole
0.017		4.98	4.39	-54 10 40	
0.001	2837	6.23	5.21	-50 -50 44	left supramarginal gyrus
0.006		5.36	4.65	-46 -58 -8	left inferior temporal gyrus, temporooccipital part
0.007		5.31	4.62	-56 -38 54	
0.001	1550	6.17	5.18	56 -48 48	right angular gyrus
0.001		5.90	5.01	50 -50 30	
0.002	1666	5.77	4.92	2 24 34	right paracingulate gyrus
0.003		5.66	4.85	2 18 40	
0.013		5.08	4.46	-2 38 52	left superior frontal gyrus medial segment
0.002	916	5.71	4.88	0 -50 18	left posterior cingulate gyrus
0.004	472	5.56	4.79	2 -14 42	right middle cingulate gyrus
0.031	26	4.76	4.23	34 36 18	right frontal pole
0.034	22	4.73	4.21	-4 -60 -20	cerebellar vermal lobules I-V
0.037	70	4.69	4.19	-22 -82 24	left superior occipital gyrus
0.040	18	4.67	4.17	-48 -8 -18	left superior temporal gyrus
0.048	2	4.60	4.11	36 -84 18	right middle occipital gyrus
0.049	3	4.59	4.11	50 8 42	right precentral gyrus
0.049	2	4.59	4.11	-24 -96 20	left occipital pole

Supplementary Table 21. Searchlight analysis for the hypothesis model Prediction Error (PE) (*p*(FWE) < .05) (VGG-16, pool5).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	10278	6.97	5.65	-36 28 52	left middle frontal gyrus
0.001		6.16	5.17	36 -8 48	right precentral gyrus
0.001		5.95	5.04	-6 40 52	left superior frontal gyrus
0.001	951	6.04	5.10	-4 -46 18	left posterior cingulate gyrus
0.041		4.64	4.15	-16 -56 34	left precuneus cortex
0.001	10506	5.91	5.01	-44 -58 16	left angular gyrus
0.002		5.81	4.95	-46 -58 4	left middle temporal gyrus
0.002		5.80	4.94	-50 -52 -38	left cerebellum exterior
0.002	1726	5.70	4.87	60 -54 38	right angular gyrus
0.004		5.52	4.76	58 -48 50	
0.045		4.60	4.11	54 -66 4	right inferior occipital gyrus
0.006	284	5.34	4.64	-2 -14 42	left middle cingulate gyrus
0.007	543	5.32	4.62	40 12 -6	right anterior insula
0.011	415	5.13	4.50	40 -84 12	right middle occipital gyrus
0.017	311	4.98	4.39	20 -40 -20	right cerebellum exterior
0.034		4.71	4.19	26 -28 -14	right hippocampus
0.017	151	4.98	4.39	6 -92 30	right occipital pole
0.020	82	4.92	4.35	32 26 52	right middle frontal gyrus
0.025	96	4.82	4.28	-68 -20 2	left superior temporal gyrus
0.029	80	4.77	4.24	-10 42 -10	left medial frontal cortex
0.033	36	4.72	4.20	30 -86 -36	right cerebellum exterior
0.042	22	4.63	4.14	38 -76 -18	right occipital fusiform gyrus
0.044	4	4.61	4.12	-18 10 -10	left putamen

Supplementary Table 22. Difference searchlight results for 'Prediction Error (PE) minus Sensory' (*p*(FWE) < .05) (VGG-16, pool4 minus conv1_2).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.007	523	5.34	4.64	0 24 36	paracingulate gyrus
0.007		5.33	4.63	0 14 42	
0.009	384	5.24	4.57	-36 30 40	left middle frontal gyrus
0.019	87	4.97	4.38	6 -14 54	right supplementary motor cortex
0.042		4.68	4.17	-2 -12 46	left middle cingulate gyrus
0.022	73	4.91	4.34	32 -16 22	right central opercular cortex
0.025	65	4.87	4.31	42 -10 52	right precentral gyrus
0.048	2	4.62	4.13	-44 18 26	left opercular part of the inferior frontal gyrus

**Supplementary Table 23. Difference searchlight results for 'Prediction Error (PE) minus Sensory' (*p*(FWE) < .05) (VGG-16, pool5 minus conv1_2*).*

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.005	172	5.42	4.69	32 -14 24	right parietal lobule
0.006	500	5.36	4.65	-32 30 52	left middle frontal gyrus
0.023		4.86	4.31	-38 34 36	
0.027		4.80	4.26	-44 40 14	
0.021	38	4.89	4.33	-6 40 56	left superior frontal gyrus
0.021	62	4.88	4.32	-28 20 -14	left posterior orbital gyrus
0.026	55	4.81	4.27	38 14 -4	right anterior insula
0.030	41	4.76	4.23	-44 -60 -4	left inferior temporal gyrus, temporooccipital part
0.030	18	4.75	4.23	-56 -60 -34	left cerebellum
0.032	10	4.73	4.21	14 66 24	right frontal pole
0.039	17	4.66	4.16	-46 50 0	left middle frontal gyrus
0.043	11	4.62	4.13	2 16 42	right paracingulate gyrus
0.044	12	4.61	4.13	40 -86 16	right middle occipital gyrus
0.045	7	4.60	4.12	-50 8 30	left precentral gyrus
0.047	1	4.59	4.11	-42 -58 20	left angular gyrus
0.047	3	4.58	4.10	36 32 34	right middle frontal gyrus

Supplementary Table 24. Difference searchlight results for 'Prediction Error (PE) minus Sharpening' (*p*(FWE) < .05) (VGG-16, pool4).

peak *p*(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.002	658	5.85	4.97	-38 28 50	left middle frontal gyrus
0.002	608	5.70	4.88	36 -14 46	right precentral gyrus
0.013		5.11	4.48	32 -16 30	right parietal lobule
0.034		4.74	4.21	34 -16 22	right insular cortex
0.007	541	5.30	4.61	12 22 38	right paracingulate gyrus
0.009		5.25	4.58	2 16 42	
0.010		5.21	4.55	2 24 34	
0.022	137	4.91	4.34	-42 16 -10	left insular cortex
0.033	34	4.75	4.23	8 -14 50	right precentral gyrus
0.034	118	4.74	4.22	-44 36 12	left triangular part of the inferior frontal gyrus
0.038	19	4.69	4.18	36 -86 18	right middle occipital gyrus
0.047	2	4.61	4.12	-52 -52 42	left supramarginal gyrus

Supplementary Table 25. Difference searchlight results for 'Prediction Error (PE) minus Sharpening' (p(FWE) < .05) (VGG-16, pool5).

peak p(FWE)	cluster equivk	peak T	peak equivZ	x,y,z {mm}	label
0.000	1115	6.50	5.38	-34 28 52	left middle frontal gyrus
0.006		5.35	4.64	-42 26 36	
0.008	87	5.24	4.57	-4 42 54	left superior frontal gyrus
0.009	103	5.22	4.56	-44 -20 -38	left inferior temporal gyrus
0.010	546	5.16	4.52	-46 40 10	left triangular part of the inferior frontal gyrus
0.016	94	4.99	4.40	18 64 24	right frontal pole
0.035	25	4.70	4.19	34 -86 16	right middle occipital gyrus
0.039	20	4.66	4.16	-2 28 36	left superior frontal gyrus medial segment
0.043	8	4.61	4.13	-2 -48 18	left posterior cingulate gyrus
0.047	3	4.58	4.10	-30 -60 30	left lateral occipital cortex, superior division
0.049	2	4.57	4.09	-50 -50 -38	left cerebellum exterior

Supplementary Table 26. Searchlight analysis for the hypothesis PE model, *p*(unc.) < .001 + cluster-corrected *p*(FWE) < .05, *k* > 10 voxel) (ResNet50, res5b_branch2b).

cluster *p*(FWE)	cluster equivk	peak *p*(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.000	15925	0.002	5.80	4.94	-38 28 50	left middle frontal gyrus
		0.023	4.87	4.31	0 26 34	paracingulate gyrus
		0.049	4.57	4.09	38 -12 50	right precentral gyrus
0.000	6163	0.033	4.73	4.21	-50 -48 -38	left cerebellum exterior
		0.119	4.21	3.82	-42 -58 -6	left inferior temporal gyrus
		0.155	4.09	3.73	-44 -46 52	left supramarginal gyrus
0.002	3876	0.035	4.70	4.19	56 -48 48	right angular gyrus
		0.045	4.61	4.12	50 -48 34	
		0.165	4.06	3.71	28 -66 28	right lateral occipital cortex, superior division
0.039	1392	0.059	4.50	4.04	8 -46 26	right cingulate gyrus, posterior division

Supplementary Table 27. Difference searchlight results for 'Prediction Error (PE) minus Sensory', *p*(unc.) < .001 + cluster-corrected *p*(FWE) < .05, *k* > 10 voxel (ResNet50, res5b_branch2b).

cluster *p*(FWE)	cluster equivk	peak *p*(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.001	4705	0.052	4.57	4.09	-38 30 36	left middle frontal gyrus
		0.069	4.46	4.01	-34 28 48	
		0.096	4.32	3.91	-48 50 0	left frontal pole
0.009	2495	0.126	4.21	3.82	42 26 36	right middle frontal gyrus
		0.190	4.03	3.68	38 18 44	
		0.215	3.97	3.64	48 -8 52	right precentral gyrus

Supplementary Table 28. Difference searchlight results for 'Prediction Error (PE) minus Sharpening', *p*(unc.) < .001 + cluster-corrected *p*(FWE) < .05, *k* > 10 voxel (ResNet50, res5b_branch2b).

cluster *p*(FWE)	cluster equivk	peak *p*(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.000	17100	0.000	6.30	5.25	-38 28 50	left middle frontal gyrus
		0.010	5.17	4.52	6 -14 50	right supplementary motor cortex
		0.014	5.06	4.45	-2 26 34	left superior frontal gyrus medial segment
0.004	3231	0.044	4.61	4.12	56 -44 52	right supramarginal gyrus
		0.092	4.32	3.91	34 -86 18	right middle occipital gyrus
		0.167	4.06	3.71	58 -62 26	right angular gyrus
0.002	3722	0.079	4.38	3.96	-44 -58 -6	left inferior temporal gyrus, temporooccipital part
		0.206	3.97	3.64	-28 -58 28	left angular gyrus
		0.238	3.90	3.58	-48 -54 -22	left inferior temporal gyrus

Supplementary Table 29. Localizer 'faces > scenes', *p*(unc.) < .001 + cluster-corrected

***p*(FWE) < .05, *k* > 10 voxel.**

cluster *p*(FWE)	cluster equivk	peak *p*(FWE)	peak T	peak equivZ	x,y,z {mm}	label
0.000	6300	0.000	14.14	Inf	48 -62 2	right middle temporal gyrus
		0.000	12.33	Inf	52 -70 -2	
		0.000	10.40	7.28	42 -70 -8	
0.000	2887	0.000	10.75	7.41	-46 -72 4	left inferior occipital gyrus
		0.000	9.75	7.01	-50 -76 -6	
		0.000	7.89	6.14	-44 -52 -20	
0.001	444	0.007	6.17	5.18	-8 -76 -44	left cerebellum exterior
		0.151	5.05	4.44	-22 -78 -56	
0.000	820	0.026	5.72	4.89	52 8 32	right precentral gyrus
		0.227	4.88	4.32	54 10 16	
		0.494	4.50	4.05	48 2 42	
0.000	730	0.034	5.62	4.82	-38 -36 40	left supramarginal gyrus
		0.205	4.93	4.35	-62 -26 46	
		0.842	4.10	3.74	-54 -26 34	
0.019	262	0.042	5.54	4.78	44 36 8	right triangular part of the inferior frontal gyrus

Supplementary References

1. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 770–778 (2016). doi:10.1109/CVPR.2016.90.

2. Ratan Murty, N. A., Bashivan, P., Abate, A., DiCarlo, J. J. & Kanwisher, N. Computational models of category-selective brain regions enable high-throughput tests of selectivity. *Nat. Commun.* **12**, 5540 (2021).

3. Morey, R. D. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutor. Quant. Methods Psychol.* **4**, 61–64 (2008).

4. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).

5. Nili, H. *et al.* A Toolbox for Representational Similarity Analysis. *PLOS Comput. Biol.* **10**, e1003553 (2014).

6. Dobs, K., Isik, L., Pantazis, D. & Kanwisher, N. How face perception unfolds over time. *Nat. Commun.* **10**, 1–10 (2019).

7. Parkhi, O. M., Vedaldi, A. & Zisserman, A. Deep Face Recognition. in *Procedings of the British Machine Vision Conference 2015* 41.1-41.12 (British Machine Vision Association, Swansea, 2015). doi:10.5244/C.29.41.

8. Blank, H., Kiebel, S. J. & von Kriegstein, K. How the human brain exchanges information across sensory modalities to recognize other people: Information Across Sensory Modalities. *Hum. Brain Mapp.* **36**, 324–339 (2015).

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

10. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 27:1-27:27 (2011).

11. Wobbrock, J. O., Findlater, L., Gergle, D. & Higgins, J. J. The aligned rank transform for nonparametric factorial analyses using only anova procedures. in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 143–146 (ACM, Vancouver BC Canada, 2011). doi:10.1145/1978942.1978963.

12. Elkin, L. A., Kay, M., Higgins, J. J. & Wobbrock, J. O. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. in *The 34th Annual ACM Symposium on User Interface Software and Technology* 754–768 (ACM, Virtual Event USA, 2021). doi:10.1145/3472749.3474784.

19.   Eidesstattliche Versicherung

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe, insbesondere ohne entgeltliche Hilfe von Vermittlungs- und Beratungsdiensten, verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe. Das gilt insbesondere auch für alle Informationen aus Internetquellen.

Soweit beim Verfassen der Dissertation KI-basierte Tools („Chatbots") verwendet wurden, versichere ich ausdrücklich, den daraus generierten Anteil deutlich kenntlich gemacht zu haben. Die „Stellungnahme des Präsidiums der Deutschen Forschungsgemeinschaft (DFG) zum Einfluss generativer Modelle für die Text- und Bilderstellung auf die Wissenschaften und das Förderhandeln der DFG" aus September 2023 wurde dabei beachtet.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich damit einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Datum   Unterschrift