



BIROn - Birkbeck Institutional Research Online

Thome, I. and Hohmann, D.M. and Zimmermann, K.M. and Smith, Marie L. and Kessler, R. and Jansen, A. (2022) "I spy with my little eye, something that is a face ...": a brain network for illusory face detection. *Cerebral Cortex* 32 (1), pp. 137-157. ISSN 1047-3211.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/44626/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

“I Spy with my Little Eye, Something that is a Face...”:

A Brain Network for Illusory Face Detection

Ina Thome^{1,2*}, Daniela M. Hohmann^{1,2}, Kristin M. Zimmermann^{1,2},

Marie L. Smith³, Roman Kessler^{1,2}, Andreas Jansen^{1,2,4*}

AFFILIATIONS:

¹ Laboratory for Multimodal Neuroimaging, Department of Psychiatry, University of Marburg, Germany

² Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Germany

³ Department of Psychological Sciences, Birkbeck College, University of London, London, UK

⁴ Core-Facility BrainImaging, Faculty of Medicine, University of Marburg, Germany

ADDRESS OF CORRESPONDENCE

Ina Thome // Andreas Jansen

Department of Psychiatry, University of Marburg

Rudolf-Bultmann-Straße 8, 35039 Marburg, Germany

Phone: +49 (0)6421 5864801 // -65273

E-Mail: ina.thome@staff.uni-marburg.de // andreas.jansen@staff.uni-marburg.de

RUNNING TITLE: A Brain Network for Illusory Face Detection

ABSTRACT

The most basic aspect of face perception is simply detecting the presence of a face, which requires the extraction of features that it has in common with other faces. Putatively, it is caused by matching high-dimensional sensory input with internal face templates, achieved through a top-down mediated coupling between prefrontal regions and brain areas in the ventral temporal cortex (“core system of face perception”). Illusory face detection tasks can be used to study these top-down influences. In the present functional magnetic resonance imaging study, we showed that illusory face perception activated just as real faces the core system, albeit with atypical left-lateralisation of the occipital face area (OFA). The core system was coupled with two distinct brain regions in the lateral prefrontal (inferior frontal gyrus, IFG) and orbitofrontal cortex (OFC). A Dynamic Causal Modelling (DCM) analysis revealed that activity in the core system during illusory face detection was upregulated by a modulatory face-specific influence of the IFG, not as previously assumed by the OFC. Based on these findings, we were able to develop the most comprehensive neuroanatomical framework of illusory face detection until now.

KEYWORDS: Dynamic causal modelling, face detection, functional magnetic resonance imaging, illusion, multivariate pattern analysis

1. INTRODUCTION

Faces are one of the most important stimuli in our lives since they convey a wealth of information, such as a person's identity, emotions or intentions. As a result, the human brain has developed highly efficient mechanisms to both detect and recognize faces. At the neuroanatomical level, faces processing is associated with a distributed hierarchical cortical network comprised of regions in the occipito-temporal cortex ("core system") as well as limbic, parietal and prefrontal regions ("extended system") (Haxby et al. 2000; Fairhall and Ishai 2007; Ishai 2008; Haxby and Gobbini 2011). The core system consists of at least three bilateral, typically right-lateralised (Corballis 1991; Kanwisher et al. 1997; Kanwisher and Yovel 2006; Rossion et al. 2012) regions: the *occipital face area* (OFA) in the inferior occipital gyrus, the *fusiform face area* (FFA) in the lateral fusiform gyrus and the *posterior superior temporal sulcus* (pSTS). The OFA is considered as the hierarchically first stage of the face processing network (Haxby et al. 2000; Rossion et al. 2012). It is typically associated with the analysis of invariant facial features like eyes or mouth and the decision whether an object is a face or not (Fox et al., 2009; Pitcher, 2011; Tsantani mariatsantani et al., 2020). The FFA predominantly processes identity (Kanwisher et al. 1997; Kanwisher and Yovel 2006; Fox, Moon, et al. 2009; Tsantani mariatsantani et al. 2020), while the pSTS is involved in perception of changeable features, such as expression and eye gaze (Hasselmo et al. 1989). The extended system consists of non-face specific brain regions outside the occipito-temporal cortex. It is recruited to extract further information from faces, e.g. emotions, trustworthiness or attractiveness of a person (Haxby et al. 2000; Haxby and Gobbini 2011; Duchaine and Yovel 2015). The anterior temporal lobe plays for instance a key role in the processing of semantic and biographic information, while the insula and in particular the amygdala have been associated with the processing of facial expressions (Haxby et al. 2000; Haxby and Gobbini 2011; Duchaine and Yovel 2015).

Interactions within the face perception network can be divided into bottom-up and top-down processing. On the one hand, sensory information enters the primary visual cortex and is further passed to the core system and subsequently to more anterior regions of the extended system in a bottom-up and feed-forward fashion. On the other hand, brain regions in the prefrontal cortex interact with the core system in the opposite direction, taking a higher-level representation (based on experience and expectations) as input to modify lower-level representations (Rauss and Pourtois 2013). Contextual information is used, for instance, for the interpretation of ambiguous sensory input (Smith et al. 2012). The interactions between bottom-up and top-down processes already occur during the first stages of face processing, in particular during the face detection stage. Face detection is an early process in face perception, describing the categorization whether or not a received stimulus is a face. It precedes more detailed face-specific analyses, for instance the recognition of identity, age or sex (Haxby et al. 2000; Haxby and Gobbini 2011; Duchaine and Yovel 2015).

It has been proposed that when a stimulus is presented to the retina, low spatial frequencies (LSFs) in the image are directly and rapidly projected from early visual areas (Bar 2003) or subcortical regions (Johnson 2005) to the prefrontal cortex (PFC) via anatomical “shortcuts” like the dorsal magnocellular pathway (Rempel-Clower and Barbas 2000; Kveraga et al. 2007). LSFs convey global information about the shape, like general orientation and structure. Thus, LSFs are essential to extract gist information from scenes, as they provide the possibility to grasp a global impression. Besides, it has been suggested that the rapid and specialised perception of faces heavily relies on LSFs. High spatial frequencies, on the other hand, represent abrupt spatial changes in the image, such as edges, and convey configural information and fine detail (Bar, 2004). Based on the LSFs information, the PFC generates a limited number of expectations about the most likely interpretations of the input image. These so-called “initial guesses” are then projected back to the occipito-temporal cortex to match the bottom-up input

with internal templates (also referred to as internal representations) of possible objects such as faces or houses and to endorse object recognition (Bar, 2003). The advantage of this top-down process is, that it limits the number of possible object representations that need to be taken into account and thus facilitates recognition. In this way, a first cue can be provided when a quick response is required (Bar, 2003). Mainly two candidate regions in the PFC have been discussed as the starting point for top-down object detection. One region is the ventromedial prefrontal cortex, also termed as the orbitofrontal cortex (OFC) (Bar et al., 2006; Kveraga et al., 2007; Summerfield et al., 2006). The other region is the ventrolateral prefrontal cortex (VLPFC), in which the inferior frontal gyrus (IFG) is located (Bar, 2003; Zhang et al., 2008). These two regions are adjacent to each other and are highly interconnected.

Over the past years, bottom-up sensory processing of faces has been studied extensively, whereas the neural mechanisms of top-down face processing are still poorly understood. In neuroimaging studies, different paradigms were developed to investigate top-down modulations in face perception. In these paradigms, subjects had to interpret ambiguous or noisy face stimuli (Summerfield et al. 2006; Meng et al. 2012), to imagine faces (Ishai et al. 2000; Mechelli 2004) or to identify illusory faces in pure noise images (“illusory face detection”, Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014)¹. Illusory face detection paradigms have the particular advantage that top-down influences on face perception are pronounced as stimuli do not include physical face information, but random

¹ Some remarks on terminology: Throughout the manuscript, we will use the term “illusory face detection task” for a paradigm in which subjects are instructed to detect faces in pure noise images. It is important to keep in mind that this task is different from typical “illusion paradigms”. In typical illusions (as for instance the illusory contours in a Kanizsa triangle) the misinterpretation of sensory input is primarily caused by specific aspects of the presented stimulus, while in the illusory face detection task face perception is primarily driven by the instructions. Further, in some studies the term “pareidolia task” is used synonymously with “illusory face detection task”. Pareidolia is defined as the “phenomenon to misperceive the sensory input as another unrelated object, pattern or meaning”. The most common example is face pareidolia, that is the illusory perception of faces in everyday objects. Using this definition, an “illusory face detection” paradigm might be classified as “face pareidolia” paradigm. In the neuroimaging literature, however, face pareidolia is often assessed with different paradigms using stimuli that have face-like characteristics (e.g. Dolan et al. 1997; Kanwisher et al. 1998; Wardle et al. 2020). To use a consistent terminology, we use the term “illusory face detection” instead of “face pareidolia”.

noise. A number of recent functional imaging studies have begun to explore the neural network involved in illusory face detection (Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014). These studies used “noise paradigms” in which subjects were presented with pure noise stimuli, but were instructed that 50% of them contained a face. With this approach, top-down face detection could be analysed while bottom-up face processing was largely excluded (since no face was present in the perceived stimulus). The previous findings yielded three main conclusions. First, illusory face perception was associated with brain activity in the core system of face processing, more specifically in the OFA and FFA (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014). The reported results, however, differed with regard to the relative contribution of OFA and FFA. Some studies particularly focused on the contribution of the FFA (Zhang et al. 2008; Li et al. 2009; Liu et al. 2014), while others implied that the OFA plays a central role as mediator in illusory face perception (Li et al. 2010). Additionally, the relative contribution of left- and right-hemispheric homologues of the core system were weighted differently. Zhang and colleagues (2008), for instance, found FFA activity only in the left hemisphere, while other studies highlighted the pivotal role of right-hemispheric areas (Li et al. 2009, 2010; Liu et al. 2014). Second, regions of the core system were strongly connected to several regions within the prefrontal cortex during illusory face detection, more specifically to the OFC and the IFG. The studies, however, differed both with regard to the prefrontal regions involved and their hemispheric dominance. Zhang and colleagues (2008) for example reported a significant correlation between the FFA and bilateral IFG. On the other hand, Li and colleagues (Li et al. 2009) focused on the right FFA and reported increased functional connectivity to bilateral OFC and left IFG during the illusory detection of faces. Yet another study ascribed a central role to the left OFC as top-down driver of illusory face detection, reporting that it was associated with a top-down influence from the left OFC to the right OFA (Li et al. 2010). These inconsistencies are all the more alarming since all studies were performed

by the same research group using an identical paradigm (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014). Furthermore, the functional roles of OFC and IFG, and in particular their functional differences, were not discussed. Third, as shown by an electroencephalography (EEG) study, prefrontal activity preceded activity in occipito-temporal regions (e.g. FFA and OFA) during the illusory perception of faces, and thus might be the starting point of illusory face perception (Smith et al. 2012). So far however, it was not specified which prefrontal region, OFC or IFG, contributed to the prefrontal EEG signal.

In summary, studies on illusory face detection suggested that the perception of illusory faces triggers a top-down mechanism starting with activity in the prefrontal cortex followed by activation of well-established face-specific brain areas in the occipito-temporal cortex (i.e. FFA and OFA). Many issues are however still unresolved, such as the exact involvement of OFA and/or FFA, the hemispheric lateralisation of the core system, the involvement of prefrontal regions, their functional roles and the interaction mechanisms/functional integration between PFC and core system.

In the present functional magnetic resonance imaging (fMRI) study, we aimed to further investigate the top-down mechanisms involved in face detection. We used an illusory face detection paradigm in which subjects were shown pure noise images, but were told that half of them contained a face. Our main goal was to explore how the interaction of the prefrontal cortex with the core system leads to the illusory perception of faces. More specifically, we assessed how brain activity in the core system is upregulated by the prefrontal cortex. The fMRI data analysis was divided into five steps. Analyses 1-3 assessed the *brain activity pattern* during illusory face detection by comparing trials in which a face has been detected and in which no face has been detected. We hypothesised that illusory face perception activates, just as real face processing, the core network of face perception (Li et al., 2009, 2010; Liu et al., 2014; Zhang et al., 2008; Zimmermann et al., 2019) (hypothesis 1), but with a less right-lateralised

hemispheric dominance (Harris & Aguirre, 2010; Meng et al., 2012; Rossion et al., 2000) (hypothesis 2). We further hypothesized that illusory face perception elicits activity in prefrontal regions, in particular in the OFC and the IFG (Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014) (hypothesis 3). Analysis 4 assessed the *functional connectivity pattern* between the core system and the prefrontal cortex using a psychophysiological interaction (PPI) analysis. We hypothesised that the core system, in particular the OFA, is functionally stronger coupled with the prefrontal cortex during illusory perception of faces (Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014) (hypothesis 4). Analysis 5 assessed how – and which – regions in the prefrontal cortex upregulate brain activity in the core system during illusory face detection using a Dynamic Causal Modelling (DCM) analysis. More specifically we tested two hypotheses. On the one hand, we assessed whether OFA activity is driven either by the OFC or the IFG. On the other hand, we assessed whether this interaction is mediated by a modulatory, region-unspecific influence on the connections from PFC to OFA (e.g. OFC→OFA connection) or whether connections to the OFA are gated by prefrontal activity (either in the OFC or the IFG).

2. MATERIALS AND METHODS

2.1 Subjects

Ten subjects (five men, mean age 23 +/- 2 years) participated in the study. All participants were healthy, with no history of neurological or psychiatric diseases. They were right-handed as assessed by the Edinburgh Handedness Inventory (EHI) (Oldfield 1971) with a cut-off at +30. They were naïve to the purpose of the study and had normal or corrected-to-normal vision. All subjects gave written informed consent prior to the experiment, which conformed to the Declaration of Helsinki and was approved by the local ethics committee of the Medical Faculty of the University of Marburg.

The number of subjects included in this study was relatively low ($n = 10$). This might lead some readers to categorize the study out of reflex as underpowered. Therefore, we would like to elaborate the study design in more detail. The design was motivated by a previous pilot study conducted in our research group. In this study, we assessed brain activity associated with illusory face detection using a similar task as in the present study (Zimmermann et al. 2019). Unexpectedly, we did not find significant fMRI activity at the group level in the core system of face perception when comparing events in which a face was detected against events in which no face was detected, being unable to replicate previous findings (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014). However, when we additionally analysed activity also at the single subject level, we could show that about half of the subjects had highly significant ($p < 0.05$, corrected for multiple comparisons) activity in bilateral OFA and FFA. In contrast, the other half did not show any activity in these regions, even at statistical thresholds as liberal as $p = 0.05$, uncorrected. We had argued that these differences were related to individual differences in task execution. Only some participants really detected faces in the noise images (and thus showed face-related activity), while the other subjects simply responded in the desired way (as later was confirmed by some subjects in a follow-up survey).

We therefore decided to design the present experiment under the following requirements. First, we put more effort in creating a realistic cover story to convince subjects that faces will be present in about half of the noise stimuli². Second, we explicitly queried after the fMRI experiment whether the subjects detected faces or not. This would have given us the possibility to exclude subjects who did not believe the cover story and thus did not expect to see any faces in the presented stimuli. After we explained the rationale of the study, however, every participant was completely astonished that no face was ever presented. Third, we decided to present results not only at the group level, but explicitly at the single subject level. For each group result, we will show every individual data point, both for behavioural and for fMRI data. This data presentation will show the variation across the individual participants and can reassure that significant group results are not driven by few outliers. Fourth, we decided to increase the power of the experiment in particular at the single subject level. Instead of measuring 30 subjects for ~15-20 min (the standard duration of illusory face detection experiments, (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014)), we decided to measure each subject for almost two hours (distributed over three sessions). Last, we implemented additional multivariate pattern analyses. These analyses independently showed, using a complementary approach, that activity specifically in the neural network underlying illusory face detection (as shown by standard analyses) is highly predictive for the decision whether or not a stimulus is detected as face also at the single subject level (see also the recommendations for methodological decisions in fMRI data analysis pipelines in high-level visual cortex (Weiner and Grill-Spector 2013)).

² Marburg, the city in which the subjects were recruited, is famous for its infrastructure for blind and visually impaired people. It has, for instance, the oldest secondary school dedicated to the education of blind people and still the only secondary school in which it is possible to attend classes from 5th grade until the final exams. In the cover story, subjects were instructed that the experiment – recognition of faces in noisy images – was motivated by a cooperation project with several educational institutions that are dedicated to the special needs of blind and visually impaired people. This cover story is highly plausible in a city in which blind people are an everyday sight.

2.2 Construction of stimulus material

For the illusory face detection task, we generated with MATLAB (R2008a, MathWorks; Natick, MA; <http://www.mathworks.com/>) random black-and-white noise fields with the aspect ratio of a face (516 x 384 pixels). An example stimulus is presented in Fig. 1. These stimuli are comparable to those used in Smith et al. (2012). An individual set of stimuli was generated for each subject. In order to verify that these random noise images can trigger illusory face detection, we conducted a behavioural pilot study and calculated classification images, which indeed revealed facial structures, supporting the validity of the paradigm (Fig. 1, also see supplementary material S1).

2.3 Experimental procedure

To induce illusory face detection, subjects were presented with the random black-and-white noise stimuli. We instructed them that faces would be present on half of the trials, filling the images' space and facing towards them like in a passport photograph. Each subject observed 999 random noise images equally distributed over three fMRI sessions. The experiment was split up into three sessions to gain a reasonable number of trials and to ensure participants' alertness over the entire experiment. Sufficing alertness was determined with a questionnaire three times during each session. The three sessions were conducted within a fortnight and at nearly the same time of the day. No reference or target face was ever presented. After each stimulus, the participants were instructed to indicate whether they had perceived a face or not by pressing either both index (face) or middle fingers (no-face) on two response boxes. Each run lasted ~35 minutes, resulting in a total duration of ~1 hour, 45 minutes.

All stimuli were presented with a vertical visual angle of 7 degrees using Presentation 18.1 (Neurobehavioral Systems, Albany, CA, USA, <http://www.neurobs.com/>). Each stimulus image was presented for 2500 ms followed by a randomly varied interstimulus interval of 3000

ms \pm 100, 200, 300 or 400 ms.

After the third run of the illusory face detection task, all subjects in the fMRI experiment additionally participated in a “functional face localizer task” (Frässle et al., 2016b, 2016c; Rossion et al., 2012) to visualise individual face-sensitive brain areas (OFA, FFA, STS). Subjects perceived pictures of faces, houses and Fourier-randomised (scrambled) versions of both. Faces were full-frontal photographs taken from the Center for Vital Longevity Face Database (Ebner 2008). House stimuli were kindly provided by Joshua Goh (Goh et al. 2010). Nine blocks of each stimulus category were presented. Each block consisted of 20 images of one category (300 ms stimulus followed by 425 ms of fixation cross) and lasted for 14.5 seconds. Between stimulus blocks, a baseline block of 14.5 seconds was inserted in which only the fixation cross was shown. The localizer task lasted approximately 13 minutes. Participants were instructed to view the images and press both index fingers when the same image appeared twice in a row (1-back task). This procedure was introduced to ensure that participants were paying attention during this passive viewing task.

2.4 MRI data acquisition

Functional and anatomical data were acquired with a 12-channel head matrix receive coil on a Siemens 3 Tesla TIM Trio MR scanner at the Department of Psychiatry, University of Marburg. Functional MRI data were collected using a T2*-weighted gradient-echo echo planar imaging sequence sensitive to the blood-oxygenation-level-dependent (BOLD) contrast (TR = 1450 ms, TE = 25 ms, voxel size = $3 \times 3 \times 4$ mm³, 30 slices, 4 mm thickness, flip angle = 90°, matrix size = 64×64 voxels, FoV = 192×192 mm²). Slices were acquired in interleaved descending order parallel to the intercommissural plane (anterior to posterior commissure). High-resolution T1-weighted anatomical scans, covering the whole brain, were acquired with a magnetization-prepared rapid gradient-echo (3D MP-RAGE) sequence in sagittal plane (TR = 1900 ms, TE =

2.26 ms, voxel size = $1 \times 1 \times 1 \text{ mm}^3$, 176 slices, 1 mm thickness, flip angle = 9° , matrix size = 256×256 voxels, FoV = $256 \times 256 \text{ mm}^2$).

2.5 MRI data analysis: Preprocessing

Functional images were analysed using SPM8 (v4191, Statistical Parametric Mapping, Wellcome Trust Center for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk>) and MATLAB (R2008b, MathWorks; Natick, MA, USA; <http://www.mathworks.com/>). *Illusory face detection task*: First, the first four scans of each session were excluded for signal saturation. Second, the remaining functional images were slice time corrected. Third, all three sessions were realigned by transforming the first scan from each session to the first scan of the first session. The images within each session were then aligned to the mean image. Fourth, the high-resolution anatomical image was coregistered with the mean functional image, then segmented and normalized to the MNI standard space. All functional images were then normalized using the spatial normalization parameters obtained from the unified segmentation-normalization approach to the anatomical image (Ashburner and Friston 2005), and resampled to a voxel size of $2 \times 2 \times 2 \text{ mm}^3$. Finally, all images were smoothed with an isotropic 6 mm full width at half maximum (FWHM) Gaussian kernel. *Functional localizer task*: First, the first four scans were excluded. Second, the remaining functional images were realigned to the mean image. Third the high-resolution anatomical image was coregistered with the mean functional image, then segmented and normalized to the MNI standard space. All functional images were then normalized using the spatial normalization parameters obtained from the unified segmentation-normalization approach to the anatomical image (Ashburner and Friston 2005), and resampled to a voxel size of $2 \times 2 \times 2 \text{ mm}^3$. Finally, all images were smoothed with an isotropic 6 mm FWHM Gaussian kernel.

2.6 MRI data analysis: Statistical analysis

The statistical analysis of the fMRI data was divided into five steps (see introduction). First, we analysed brain activity in the core system (analysis 1). Second, we assessed hemispheric dominance of brain activity in the core system (analysis 2). Third, we assessed brain activity in the prefrontal cortex (analysis 3). Fourth, we assessed the functional connectivity between core system and prefrontal cortex (analysis 4). Last, we assessed how – and which - regions in the prefrontal cortex upregulate brain activity in the core system (analysis 5). These analyses are described in the following.

2.6.1 Assessment of brain activity (analysis 1 and 3)

Statistical analysis of the preprocessed functional data was conducted using a General Linear Model (GLM) (Friston et al. 1995; Worsley and Friston 1995). At the subject level, we modelled each condition (illusory face detection task: “face detected”, “no-face detected”; functional localizer: “faces”, “houses”, “scrambled images”) as a regressor, convolved with the canonical hemodynamic response function implemented in SPM. Furthermore, the six realignment parameters were introduced as nuisance regressors to control for movement-related artefacts. Low-frequency noise in the data was accounted for by a high-pass filter (cut-off frequency: 1/128 Hz). The illusory face detection task consisted of three sessions. The “face detected” and “no-face detected” regressors were entered for each session separately in one GLM (i.e. were not concatenated). Parameter estimate (β -) images were computed for each subject. We calculated contrast images and t-statistic images for the contrasts “face detected > no-face detected” (illusory face detection task) and “2*faces > houses + scrambled images” (localizer task). At the group level, the individual contrast images were entered in a random-effects analysis (illusory face detection task: one-way ANOVA with “session” as within-subject factor; functional localizer: one-sample t-test). Anatomy labels for peak activations were assessed with the SPM Anatomy toolbox (version 2.1) and corresponding Brodmann areas (BA)

were identified with the MNI to Talairach Converter (Lacadie et al., 2008) embedded in BioImage Suite (Papademetris et al. 2006, <https://bioimagesuiteweb.github.io/webapp/mni2tal.html>). In analysis 1, we assessed brain activity in the core system. We expected that illusory face perception is associated with brain activity in particular in the bilateral OFA and FFA, just as for real face processing (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014; Zimmermann et al. 2019) (hypothesis 1). In analysis 3, we assessed brain activity in the prefrontal cortex. We expected activation in the bilateral OFC and IFG (Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014) (hypothesis 3).

We additionally assessed the neural network underlying illusory face detection using a multivariate pattern analysis (MVPA). This analysis assesses BOLD responses as patterns of activity, thus increasing the amount of information that can be decoded from brain activity. It can be used for instance to quantify to which degree BOLD activity in specific regions predicts whether a subject has classified a noise stimulus as face or not. Multivariate pattern analyses of fMRI data have often proven to be more sensitive and more informative about the functional organization of cortex than univariate GLM analyses, in particular at the single subject level (Haxby 2012). It can thus be used as complementary evidence to assess whether specific brain regions are associated with specific cognitive functions. Our multivariate pattern analysis confirmed the findings from the GLM analysis, showing that the brain activity in the core system of face perception as well as in the OFC and IFG is highly predictive for the behavioural responses. In order to simplify the reading flow, MVPA methods and results will be only shortly presented in the main text (see discussion, §4.3). They will be described, however, in detail in a separate section in the supplementary material (S8).

2.6.2 Assessment of hemispheric lateralisation (analysis 2)

Hemispheric lateralisation was described by a lateralisation index (LI) given by the following expression (Jansen et al. 2006):

$$LI = \frac{A_L - A_R}{A_L + A_R}, \quad (1)$$

where A_L and A_R refer to values of fMRI-measured activity for regions of interest (ROIs) within the left (L) and right (R) hemisphere. Positive values indicate left-hemispheric dominance, negative values right-hemispheric lateralisation. The LI typically ranges from -1 to 1. LIs were calculated using the bootstrap procedure (Wilke and Schmithorst 2006) implemented in the LI toolbox for SPM8 (Wilke and Lidzba 2007). This involved the initial thresholding and masking of an input image (t-map). Subsequently, 100 bootstrapped samples (resampling ratio $k = 0.25$) were drawn from all surviving voxels. From these 100 samples from each hemisphere, all possible LI combinations (10,000) were calculated and plotted in a histogram. Only the central 50% were kept. This procedure was repeated for 20 regularly spaced thresholding steps, resulting in a weighted mean LI. The weighting factor was the respective threshold of the t-image. LI values obtained at higher thresholds were weighted progressively stronger.

In analysis 2, LIs were computed for each subject for both the localizer task (contrast “2*faces > houses + scrambled images”) and the illusory face detection task (“face detected > no-face detected”) for both OFA and FFA. To determine their localization, we used box-shaped ROIs. They were designed with the WFU PickAtlas (Maldjian et al., 2004; Maldjian et al., 2003) (v 3.0.5) and were based on literature coordinates as reported by Berman et al. (2010), Fox et al. (2009), Pitcher et al. (2011), Rossion et al. (2012) and the automated meta-analysis platform Neurosynth.org (for a detailed list, see Table S1 in the supplementary material). The FFA ROI was centred around $x = \pm 40$, $y = -47$, $z = -16$ with dimensions ($x = 40$ mm, $y = 26$ mm, $z = 28$ mm). The OFA ROI was centred around $x = \pm 35$, $y = -79$, $z = -16$ with dimensions

($x = 30$ mm, $y = 38$ mm, $z = 28$ mm). The LIs for the face localizer task and the illusory face detection task were compared by means of a two-sided paired-sample t-test. We hypothesized that, contrary to the perception of real faces, illusory face detection will lead to a less right-lateralised activation pattern (Harris & Aguirre, 2010; Meng et al., 2012; Rossion et al., 2000) (hypothesis 2).

2.6.3 Assessment of functional connectivity (analysis 4)

To analyse the connectivity pattern between the core system of face perception and prefrontal regions during the illusory face detection task, we performed a PPI analysis (Friston et al. 1997). A PPI analysis investigates how the functional connectivity (i.e. the temporal correlation) between a seed region and target regions changes depending on the experimental context. We conducted separate PPI analyses for each of the four regions of the core system (i.e. left and right OFA, left and right FFA) using the generalized PPI (gPPI) toolbox (McLaren et al., 2012). The generalised approach potentially improves model fit, specificity to true negative findings, and sensitivity to true positive findings and is thus superior to the standard PPI implementation provided by SPM (McLaren et al. 2012). In order to define the seed regions, the individual coordinates derived from the first level analysis of the illusory face detection task were used (see Table S4, supplementary material). Seed regions were defined as an 8 mm radius sphere centred on the most significantly active voxel (peak voxel) in the main effect analysis (“face detected > no-face detected”) found in the right FFA, left FFA, right OFA and left OFA, respectively³. Time series for each ROI were extracted as the first eigenvariate at $p < 0.05$ (uncorrected for multiple comparisons) on the individual subject level. We created for each subject a new first level design matrix with the following regressors: task regressors (“face”

³ We were interested in the coupling between prefrontal regions and the core-system. Thus, seeds could have been placed either in the prefrontal regions or the core-system. As the localisation of the core system is well known, unlike the localisation of the prefrontal regions, we decided to use the four core system regions as seed for the PPI.

and “no-face”), PPI regressors (“face” and “no-face”), and the original BOLD time series as well as the realignment parameters as covariates of no interest. The illusory face detection task consisted of three sessions, the regressors for all sessions were entered as separate sessions in one GLM (i.e. were not concatenated). Parameter estimate (β -) images were calculated for each subject. We calculated a contrast image and a t-statistic image for the contrast “face detected > no-face detected” (i.e. a “1” on the PPI regressor “face detected” and “-1” on the PPI regressor “no-face detected”). The resulting contrast images were entered into a random-effects group analysis (one-way ANOVA within subject, with “session” as within-subject factor). We hypothesised that the core system, in particular the OFA, is functionally stronger coupled with both the bilateral OFC and the bilateral IFG during illusory face detection (Zhang et al. 2008; Li et al. 2009, 2010; Smith et al. 2012; Liu et al. 2014) (hypothesis 4).

2.6.4 Upregulation of brain activity in the core system by prefrontal regions (analysis 5)

The interaction between prefrontal cortex and core system was assessed using DCM (version 12) implemented in SPM12 (r6685, Statistical Parametric Mapping, Wellcome Trust Center for Neuroimaging, London, UK; <http://www.fil.ion.ucl.ac.uk>) and MATLAB (R2009b, MathWorks; Natick, MA, USA; <http://www.mathworks.com/>). DCM is a Bayesian framework for investigating the effective connectivity in a neural network based on neuroimaging data. In the present study, we used a non-linear extension of DCM, being able to also test whether the activity in a given region gates interactions between other regions (Stephan et al. 2008). The state equation in non-linear DCM is given by:

$$\frac{dz}{dt} = Az + \sum_{i=1}^m u_i B^{(i)} z + Cu + \sum_{j=1}^n z_j D^{(j)} z, \quad (2)$$

where z depicts the neuronal activities, u corresponds to the experimental input. A describes the endogenous (fixed or context-independent) connection strengths, B^i defines how the

experimental manipulation u_i affects the connections among the network regions (modulatory connectivity), C describes how the driving inputs directly influence the neuronal state of the network regions, and D^j codes how the j^{th} region gates connections in the system. The dynamics of the neuronal states in each region are translated into predictions of the measured BOLD signal by a hemodynamic forward model (Balloon-Windkessel model; (Buxton et al., 1998)). Using a Variational Laplace approach with Gaussian assumptions on the prior and posterior distributions, the posterior densities of the model parameters (i.e. conditional mean and covariance) can be estimated by maximising the negative free energy.

The starting point for a DCM analysis is the selection of a fixed set of regions, their possible connections, driving inputs and modulatory inputs. Different models can then be compared in order to identify which models best predict the data. DCM enables inferences at different levels, on the one hand inference on model space, on the other hand inference on parameter space of any given model. In the following, we will describe (i) the extraction of time series in specified ROIs, (ii) the model space definition, and (iii) the assessment of model parameters. Last, we will outline (iv) critical assumptions of the neural models and discuss which violations would render the conclusions of the DCM analysis invalid.

(i) Time series extraction: While the GLM-based analyses in SPM allow the analysis of multi-session data, fMRI time series extraction for DCM operates on a per-session basis. We therefore initially concatenated the time series from the three sessions. First, we specified a new first level design matrix which included the functional images from all three sessions. The trial onsets of the second and third session were adapted to create a continuous experimental design. The design matrix included, as before, the face regressor, the no-face regressor and the six realignment parameters. A high-pass filter with a cut-off of 1/128 Hz was used. Second, we applied the `spm_fmri_concatenate` function to replace the usual mean column in the design

matrix with regressors modelling each session. It also corrected the high-pass filter and temporal non-sphericity calculations to account for the original session lengths. Third, the GLM was estimated as usual.

We extracted the times series from eight ROIs: early visual cortex (V1), OFA, IFG and OFC, each located in the left and right hemisphere. To account for inter-subject variability in the location of these regions, coordinates of the ROIs were determined for each subject individually (see supplementary material S3 and S5). First, group level peak activation coordinates were determined for each ROI. For the OFA and IFG, we used the “face detected > no-face detected” contrast. For the OFC, we used the “no-face detected > face detected” contrast. For V1, we used the “face detected + no-face detected” contrast. Second, individual ROI centre coordinates were identified as the local maximum in the single subjects’ activation maps thresholded at $p < 0.001$, uncorrected⁴, under the following anatomical constraints: V1 coordinates had to be located in Brodmann Area 17 (2D mask dilation: 2 voxels), while OFA coordinates were previously determined through the assessment of brain activity in the core system (analysis 1). IFG coordinates had to be in the inferior frontal gyrus (pars opercularis, triangularis and orbitalis; 2D dilation: 1 voxel; IBASPM116 atlas), and the OFC in the orbitofrontal cortex (2D dilation: 1 voxel; IBASPM116 atlas), as assessed by the WFU PickAtlas (v 3.0.5) implemented in SPM (Maldjian et al. 2003, 2004). Masks were constructed separately for each hemisphere. Third, BOLD signal time series were extracted from each subject-specific ROI as the first eigenvariate of all activated voxels within a 4 mm sphere centred on the individual coordinates ($p < 0.1$, uncorrected). Time series were mean-centred and movement-related variance was removed (by adjustment with regard to an effects-of-interest F-contrast).

⁴ More specifically, activity in 73 of 80 ROIs could be detected in the single subject activation patterns at $p < 0.001$ uncorrected. In some cases, the threshold had to be lowered. For one subject, activity in the left OFC was detected only at $p < 0.05$ uncorrected. For two subjects, activity in both the left and right OFC was detected only at $p < 0.05$ uncorrected. For two subjects, activity in the left IFG was detected only at $p < 0.05$ uncorrected.

(ii) *Definition of the model space:* We built separate DCMs for the left and right hemisphere. Each model included four regions: V1 (as input region), OFA, IFG and OFC. We explicitly chose the OFA and not the FFA as the region within the core system that receives information from the prefrontal cortex. On the one hand, the OFA is typically considered as the first stage in the hierarchical face processing network (Haxby et al. 2000; Duchaine and Yovel 2015). On the other hand, the OFA is associated in particular with the identification of face parts, such as eyes, nose, and mouth (Haxby et al. 2000; Pitcher et al. 2011). These early face perception processes are essential for the occurrence of illusory face perception. The FFA and pSTS are, in contrast, associated with higher-level processes (e.g. face identification and discrimination of facial expressions). Nine models were constructed for each hemisphere. All models were identical with regard to the endogenous connectivity (A-matrix) and the exogenous input (C-matrix), but differed with regard to the B- and D-matrix (for an overview, see Fig. 2). To model external manipulations, we created a new SPM design matrix with the two regressors “noise” (modelling both “face detected” and “no-face detected” trials) and “face detected” (modelling the “face detected” trials). The models were defined as follows: Endogenous connections were set from V1 to the OFA (representing bottom-up input to the core system) and from V1 to the OFC (representing an anatomical pathway conveying LSFs in early stages of visual processing to form a first quick classification of visual input (Bar 2003; Kveraga et al. 2007). Endogenous connections were further set bidirectionally between OFA and IFG, OFA and OFC, and IFG and OFC (representing interactions between the core system and prefrontal regions as well as interactions within the prefrontal cortex). Exogenous input was modelled by the “noise” regressor and was set to V1. B- and D-matrix were set to define different models representing distinct, mutually exclusive hypotheses on how prefrontal regions upregulate activity in the core system. These models were grouped into three families (Fig. 2). In family F1 (“OFC family”), OFA activity is upregulated directly by the OFC, either by modulatory input from the

“face detected” regressor on the OFC→OFA connection (model M1) or by gating the V1→OFA (model M2) or the IFG→OFA connection (model M3), respectively. In family F2 (“IFG-A family”), OFA activity is upregulated directly by the IFG, either by modulatory input of the “face detected” regressor on the IFG→OFA connection (model M4) or by gating the V1→OFA (model M5) or the OFC→OFA connection (model M6), respectively. The models in family F3 (“IFG-B family”) were identical to those in family 2, but with an additional modulatory upregulation of the OFC→IFG connection in case a face was detected.

(iii) Assessment of model parameters: Families were compared using random effects Bayesian model selection (BMS) at the family level (Penny et al., 2010; Stephan et al., 2009). Model parameters were subsequently estimated for each subject by random effects Bayesian model averaging (BMA) (Penny et al. 2010) across the models of the winning family within a pre-specified Occam's window ($p < 0.05$). The BMA parameter estimates were then entered into summary statistics at the group level. To test whether the BMA parameter estimates significantly differed from zero, we used a Wilcoxon Signed-Rank test ($p < 0.05$, uncorrected).

(iv) Critical assumptions of the neural models:

For the interpretation of the DCM results, it is important to clearly distinguish between prior model assumptions used to build neural models and the conclusions we can draw from the analysis. In the Bayesian framework of DCM, different assumptions can lead to different conclusions, even if the same data is used for inference. Therefore, it is important to clearly state these assumptions (for a discussion, see Stephan et al. (2010).

The first critical assumption is the existence of an anatomical connection from V1 to the OFC. We believe that the existence of such an anatomical connection is in general undisputed (see for example the Human Brainnetome Atlas (Fan et al. 2016) or the 1000BRAINS study (Caspers et al. 2014)). This connection might be the “dorsal magnocellular pathway”, as

proposed by Bar (2003), but it could likewise be any other route. Technically, the neural models specified in our DCM analysis only assume the existence of any connection between V1 and OFC (specified by the A-matrix), without making further assumptions on the specific fibre tract. The second critical assumption is the rapid transfer of stimulus information from V1 to OFC to facilitate face recognition. More specifically, we postulated that specific stimulus information described by LSFs are rapidly projected during the face detection process from the retina via the magnocellular pathway to V1 (Merigan and Maunsell 1993; Bullier and Nowak 1995). In our DCM analysis, this assumption is explicitly modelled by the C-matrix. Subsequently this information is rapidly transferred to the OFC via anatomical connections (as outlined above). The assumption of a rapid transfer of stimulus information is based on neuroimaging studies on object and face recognition, showing that prefrontal activity precedes occipito-temporal activity. Bar et al. (2006) found, for instance, that activity in the OFC developed 50 ms earlier compared to activity in fusiform regions implicated in object recognition. This suggests that the OFC receives at least some stimulus information earlier than classical object processing regions. In a similar vein, Smith et al. (2012) found frontal activity before lateral occipito-temporal activity in their EEG study with pure noise images. The assumption of a specific transfer of LSF information is also supported by Bar et al. (2006), reporting significant stronger functional connectivity between OFC and early visual areas when subjects were presented with images that contained LSFs compared to images that contained primarily HSFs. Likewise, Bognár and colleagues (2017) showed in a transcranial direct-current stimulation (tDCS) study that LSF information facilitates decision making through the OFC. The result of our DCM analysis will be still valid even if some of our assumptions are violated. We will come to the same conclusions (i) if the V1-OFC connection is not the speculative “dorsal magnocellular pathway” (as proposed by Bar (2003)), but another route, and (ii) if not LSF information but other aspects of the stimuli are transferred to the OFC for a first classification of the stimulus. We want to

point out, however, that a critical assumption of our model is the existence of an anatomical short-cut to the OFC that is used for a rapid information transfer of specific stimulus information facilitating face recognition.

3. RESULTS

The results section is structured as follows: First, we present the behavioural data (3.1). Second, we describe the illusory face detection network (3.2), in particular with regard to the involvement of the core system (3.2.1, analysis 1), its hemispheric lateralisation (3.2.2, analysis 2), the activation of the prefrontal cortex (3.2.3, analysis 3) and the functional connectivity between core system and prefrontal cortex (3.2.4, analysis 4). Last, we will present the results from the DCM analysis assessing how the prefrontal cortex upregulates brain activity in the core system during illusory face detection (3.3, analysis 5).

3.1 Behavioural data

Face response rate: Although there was never a face present, subjects detected a face in 37.66% (SD = 9.63%) of trials. The face response rate did not significantly change over sessions (Fig. 3, one-way repeated measures ANOVA with Greenhouse-Geisser correction, $F(1.21, 10.89) = 0.08$, $p = 0.83$). The mean face response rate and SD was within the same range as reported in earlier studies. For example, Smith et al. (2012) found a face response rate of 44.0% (SD = 5.9%) and Liu and colleagues (2014) reported 34.23% (SD = 15.95%).

Reaction time: A two-way repeated measures ANOVA was conducted to compare the effects of illusory face detection (two levels: face/no-face rating) on reaction times over three experimental sessions. When subjects detected a face they responded slightly earlier (mean = 1913 ms, SD = 404 ms) compared to no-face detection (mean = 2061 ms, SD = 425 ms), $F(1,9) = 3.422$, $p = 0.097$, not significant. With each session, participants answered

significantly faster (session 1 mean = 2111 ms, session 2 mean = 2021 ms, session 3 mean = 1829 ms), $F(2,18) = 6.804$, $p = 0.006$.

3.2 The illusory face detection network: Activity, lateralisation and functional connectivity

The illusory face detection task activated a distributed network spanning from frontal over parietal to occipito-temporal regions (contrast “face detected > no-face detected”, Fig. 4, Table S3). These regions encompassed the core system of face perception, in particular bilateral OFA and FFA, and the prefrontal cortex with the main focus on the IFG. The OFC was however, as outlined below, unexpectedly deactivated, i.e. stronger active for the contrast “no-face detected > face detected”. In the following, we will present the results in more detail.

3.2.1 Analysis 1: Does illusory face detection activate the core network of face perception?

First, we analysed whether the illusory face detection task activates, as assessed by the contrast “face detected > no-face detected”, the core system of face perception. Our results clearly showed, even though participants never observed a real face, activity in the bilateral core network of face perception (i.e. OFA and FFA) (Fig. 5; Table 1, $p < 0.05$ corrected for multiple comparisons). Importantly, the activity in the occipito-temporal cortex elicited by the illusory face detection task overlapped with activation elicited by the face localizer paradigm, i.e. by real faces (see supplementary material S2). The illusory perception of faces thus activated the same regions in the core system as the perception of real faces.

==== insert Table 1 here ====

3.2.2 Analysis 2: Hemispheric lateralisation of the core network of face perception

Second, we calculated a LI to compare hemispheric lateralisation in the core system between the illusory face detection and the face localizer task. For the illusory face detection task, the LI was 0.32 ± 0.29 for the OFA and -0.35 ± 0.26 for the FFA. For the localizer task, the LI was 0.00 ± 0.51 for the OFA and -0.27 ± 0.39 for the FFA (Fig. 6). OFA activation was thus significantly stronger left-lateralised for illusory face detection than for real face processing (two-sided paired t-test, $t(9) = 2.23$, $p = 0.02$). In contrast, no significant difference was found in the right-hemispheric lateralisation of the FFA between the illusory face detection and the localizer task (two-sided paired t-test, $t(9) = -0.48$, $p = 0.64$).

3.2.3 Analysis 3: Activation of the prefrontal cortex during illusory face detection

Third, we analysed which brain regions, apart from the core system, were activated in the illusory face detection task. For the contrast “face detected > no-face detected” we found, additionally to the core system, a large network also encompassing parietal (bilateral BA 7 in inferior parietal lobule), insular and prefrontal regions (supplementary material Table S3). Based on findings by Li et al. (2009, 2010) we specifically hypothesised that illusory face detection would elicit activity in prefrontal cortex areas, especially in the bilateral OFC and IFG. Our results, however, showed that brain activity was found only in the IFG (BA 44 and BA 46, with stronger activity in the right than the left hemisphere) (Table 2, $p < 0.05$, FWE corrected for multiple comparisons). In contrast, no significant activation was found in the OFC. The activation analysis therefore did not provide supporting evidence for OFC involvement in the illusory face detection network. We will later show, however, that the OFC nevertheless plays an important role in illusory face detection, being functionally coupled to the core system and, most unexpectedly, significantly “deactivated” in the contrast “face detected” > “no-face detected” (see 3.2.4).

==== insert Table 2 here ====

3.2.4 Analysis 4: Functional connectivity between core system and prefrontal cortex

Fourth, we analysed the functional connectivity pattern between PFC and core system using four PPI analyses, one for each core system's region (i.e. left and right OFA, left and right FFA) as seed region. The PPI analysis revealed interactions with a vast network including parietal, occipito-temporal and prefrontal brain regions. Whole brain functional connections are shown in Fig. 7. Detailed lists with peak voxel coordinates ($p < 0.05$, FWE corrected for multiple comparisons) for PFC regions are shown in Table 3; extensive lists with all other regions can be found in the supplementary material (Table S5-S8).

The main finding is that both bilateral OFC and bilateral IFG were stronger connected to the core system during the “face detected” compared to the “no-face detected” trials. More specifically, the left OFA was significantly connected to bilateral IFG and bilateral OFC and the right OFA was linked with bilateral IFG ($p < 0.05$, FWE corrected for multiple comparisons). No significant results were found for the left and right FFA. However, if thresholds were lowered to $p < 0.001$ uncorrected (whole brain), the left FFA was connected to bilateral IFG and for the right FFA at least connections to the left IFG were detected (see supplementary material S4). Importantly, the cluster in the IFG revealed by the PPI analysis was identical to the cluster revealed by the activation analysis (3.2.3), showing that indeed the same region within the IFG was both activated and stronger connected to the core system during face detection trials (Fig. 8 bottom). In contrast, the OFC was characterised by a different functional profile. Although it was, just as the IFG, stronger connected to the core system during “face detected” trials, it showed significant “deactivation” in the activation analysis (i.e. more activity for the “no-face detected” > “face detected” contrast) (Fig. 8 top). This showed that both OFC and IFG were involved in the illusory face detection network, albeit functionally in

different roles.

A further finding emerged from the comparison of the OFA and the FFA connectivity patterns. Left and right OFA exhibited considerably more connections to various brain regions compared to both FFAs (see supplementary Tables S5-S8). A formal statistical comparison (flexible factorial design, with factors subject and seed region) of all four PPIs showed that the connectivity pattern of the left OFA seed was significantly more pronounced compared to all the other seed regions.

==== insert Table 3 here ====

3.3 Upregulation of brain activity in the core system by prefrontal regions

While the first four analyses assessed general properties of the illusory face detection network, analysis 5 tested specific models of how brain activity in the core system is upregulated by prefrontal regions during illusory face perception. In the following, we will describe the results of the Bayesian model selection (3.3.1) and present the averaged parameter estimates of the winning family (3.3.2).

3.3.1 Bayesian model selection

We used a random effects BMS procedure at the family level to compare a set of a-priori models representing distinct hypotheses of how the prefrontal cortex upregulates brain activity in the core system (Fig. 2). BMS was carried out for both hemispheres separately. Posterior evidence was strongest for family F3 (“IFG-B”) for both left- and right-hemispheric models, revealing an exceedance probability of 94.6% (left hemisphere) and 94.8% (right hemisphere), respectively (Table 4). Family F3 grouped models with both a modulatory effect of the face regressor on the OFC→IFG connections and an influence of the IFG on OFA activity. The latter

could be achieved by a modulatory effect of the face regressor on the IFG→OFA connection (model M7) or by either a gating effect of IFG activity on the V1→OFA connection (model M8) or on the OFC→OFA connection (model M9). After model selection, we used the Matlab routine *spm_dcm_fmri_check* as a diagnostic tool to ensure model inversion has converged. It compares the predicted and observed responses over all regions and provides a coefficient of determination (“percent variance explained”). As a rule of thumb, the explained variance should not drop well below 10%, as this indicates a lack of convergence. For each subject, we calculated the explained variance averaged over the models of the winning family F3 (supplementary material S5). For left-hemispheric models, mean variance explained was $25.92 \pm 12.43\%$, for right-hemispheric models $27.90 \pm 10.02\%$.

==== insert Table 4 here ====

3.3.2 Averaged DCM connectivity parameter estimates

Individual DCM connectivity parameters were estimated using BMA across the three models of the winning family within the pre-specified Occam’s window. The resulting averaged parameter estimates were then entered into summary statistics at the group level (two-sided Wilcoxon Signed-Rank test, $p < 0.05$). A full overview of the parameter estimates for both left- and right-hemispheric models is given in Fig. 9 and Table 5.

The main finding was that OFA activity was upregulated mostly by an excitatory face-specific modulation of the IFG→OFA connection. This connection was characterized by an excitatory endogenous parameter estimate of +0.247 Hz for the left-hemispheric and of +0.188 Hz for the right-hemispheric model. The connection strength was increased by a modulatory input of +1.233 Hz and +0.745 Hz for the left- and right-hemispheric model, respectively. This represents a 6.0-fold (left hemisphere) and 5.0-fold (right hemisphere) increase in connectivity

strength when subjects experience illusory face perception. Furthermore, the modulatory IFG→OFA parameter ranges between 0.0 and +2.5 for the left-hemispheric and 0.0 and +2.1 for the right-hemispheric model, showing that this connection is excitatory not only at the group level, but also in each individual subject. In contrast, a gating of connections to the OFA by IFG activity (i.e. OFC→OFA, V1→OFA) was negligible, being characterized by D-matrix parameters of 0.0 at the group level.

A further and at first view unexpected finding was the inhibitory modulatory input of the “face detected” regressor on the OFC→IFG connection (-0.568 Hz for the left-hemispheric model, -0.546 Hz for the right-hemispheric model). The connection was otherwise characterized by excitatory endogenous parameters (+0.291 Hz for the left-hemispheric and +0.060 Hz for the right-hemispheric model). This finding, however, has to be interpreted with care. Unlike the modulatory IFG→OFA parameter, the modulatory OFC→IFG parameter shows much higher inter-individual variability. It ranges from -1.4 to +3.5 for the left-hemispheric and from -2.9 to +2.4 for the right-hemispheric model (Table 5), showing that the modulation of the OFC→IFG connection is excitatory in some subjects, but inhibitory in others.

Interestingly, our findings were equal for both left- and right-hemispheric models, indicating that the intrahemispheric functional integration was similar in both the left and right hemisphere. The only significant difference between both models was the strength of the endogenous OFC→IFG parameter, with a stronger coupling of the left-hemispheric connection (paired samples two-sided Wilcoxon Signed-Rank Test, $p = 0.049$).

==== insert Table 5 here ====

4. DISCUSSION

In the present study, we investigated the neural correlates of illusory face detection. Our main goal was to explore how the interaction of prefrontal cortex and core system leads to the illusory perception of faces. Our main findings can be summarized, in accordance with the five analyses outlined in the introduction, as follows: First, illusory face perception activated the core network of face perception (i.e. left and right OFA, left and right FFA). Second, OFA activity was, unlike for real faces, left-lateralised. Third, illusory face perception activated the bilateral IFG, but not – as hypothesized – the OFC. Unexpectedly, the OFC was even significantly stronger deactivated during illusory face detection. Fourth, the core system, in particular the OFA, was stronger coupled (i.e. showed increased functional connectivity) with both bilateral IFG and bilateral OFC during illusory face perception. Lastly, activity in the core system during illusory face detection was mostly driven by a modulatory face-specific influence of the IFG on the OFA.

In the following, we discuss properties of the illusory face detection network (4.1) and describe how the prefrontal cortex upregulates brain activity in the core system during the illusory perception of faces (4.2). Finally, we summarise our results and give an outlook for future studies (4.3).

4.1 The illusory face detection network: Activity, lateralisation and functional connectivity

4.1.1 Analysis 1: Does illusory face detection activate the core network of face perception?

Our data showed that illusory face perception, just like real faces, significantly activated the core system of face processing (i.e. bilateral OFA and FFA). Importantly, brain activation during the illusory face detection task strongly overlapped with brain activity during the face localizer task, suggesting that the same network was recruited for processing of real and illusory

faces. Overall, our data suggests, consistent with findings by Zhang et al. (2008), Li et al. (2009, 2010) and Liu et al. (2014), that the core system of face processing is not only involved in bottom-up driven processing of real faces, but also in top-down modulated processing of illusory faces.

4.1.2 Analysis 2: Hemispheric lateralisation of the core network of face perception

Although the neural network underlying face processing is distributed across both hemispheres, the lateralisation of the core system is typically right-dominant, as shown by functional imaging, lesion and behavioural studies (Kanwisher et al., 1997; Kanwisher & Yovel, 2006; Rossion et al., 2012). We currently lack a precise characterization of the lateral differences in face processing (Meng et al. 2012). Rossion et al. (2000) however proposed that the right hemisphere might be specialized for the processing of faces as an overall representation (“holistic processing”), while the left hemisphere is specialized for earlier stages, in particular the processing of single features. They showed that brain activity during a face matching task was increased in the left FFA when subjects had to match specific face parts (e.g. eyes), while activity in the right FFA was higher when faces were matched as a whole. Their hypothesis is supported by behavioural and lesion studies. Faces are typically processed faster and more accurately when they are presented to the left compared to the right visual field (Leehey et al. 1978; Levine et al. 1988; Hillger and Koenig 1991; Rhodes et al. 1993). This right hemisphere advantage is reduced when faces are inverted, which disrupts the configural processing (Leehey et al. 1978; Hillger and Koenig 1991). Furthermore, patients with right-hemispheric brain injuries are often no longer better at matching upright compared to inverted faces (Yin 1970). Their hypothesis got further support by a recent fMRI study investigating objects in which the face semblance was continually manipulated (Meng et al. 2012). The authors showed that brain activity in the right FFA was increased after a stimulus was categorized as a face (i.e. holistic

processing was established), while activity in the left FFA corresponded to the image-level face similarity. Furthermore, responses in the left FFA had an earlier onset and dissolution relative to those on the right.

For illusory face detection, hemispheric lateralisation is far less studied and even yielded contradictory results. While some studies highlighted a pivotal role of the right FFA (Li et al. 2009; Liu et al. 2014; Wardle et al. 2020) or right OFA (Li et al. 2010), others detected only activation in the left FFA (Zhang et al. 2008). In the present study, subjects were presented with pure noise images. They were instructed that faces would be present on half of the trials, filling the images' space and facing towards them like in a passport photograph. To solve the task, they had to actively search for the presence of specific features, typically first-order face characteristics (eyes, nose, mouth), and had to decide whether or not a specific feature might be discernible in the stimulus. This type of stimulus processing might be called a “feature based processing style” since the subject is actively searching for the presence of specific features (e.g. eyes). This kind of processing is possible even if no actual first-order face features are present in the stimulus. The processing style is less “holistic” because the stimuli cannot be processed as an overall face presentation. A “holistic” processing would require the detection of a face configuration and the subsequent analysis of further stimulus elements in relation to the first-order face characteristics. Based on the hypotheses on the processing style of the left and right hemisphere outlined above, we therefore hypothesized that illusory face detection would be associated with a reduced right-hemispheric lateralisation compared to real face processing. Our results partly confirmed this hypothesis. We showed that illusory face detection led to left-lateralised activity in the OFA (but not the FFA). The degree of lateralisation was significantly different between the illusory face detection and the real face processing task.

4.1.3 Analysis 3 and 4: Involvement of the prefrontal cortex during illusory face detection

In analyses 3 and 4, we explored which parts of the PFC were associated with illusory face detection. We tackled this question using two analyses. On the one hand, we assessed activation differences between events classified as face and no-face (analysis 3, “activation analysis”). On the other hand, we analysed whether the functional connectivity between the core system’s brain regions (i.e. left and right OFA, left and right FFA) and the PFC differed between the face and no-face events (analysis 4, “PPI analysis”)⁵.

Our results showed that both bilateral OFC and bilateral IFG were associated with illusory face detection, albeit in different functional roles. The ***OFC*** increased its coupling to the core network during “successful” face perception, but was, in contrast to our a-priori expectations, even stronger activated during the “no-face detected” compared to the “face detected” trials. These findings are in accordance with a more general, non-face specific role of the OFC in illusory face perception. As previously proposed, its main function might lie in the attempt to detect specific patterns in a stimulus and thus enabling top-down object processing (Bar et al., 2006, 2009; Frey et al., 2004; Frey & Petrides, 2000; Kveraga et al., 2007). The OFC would then be constantly active during the whole task⁶, trying to make sense of the visual input by matching LSFs of the stimulus with internal face templates. When a match with face templates

⁵ The activation analysis compares events in which illusory face detection did or did not occur. Although this analysis is sufficient to determine brain regions that are part of the illusory face detection network, it has two limitations. First, it does not show in which function a brain region is involved in the illusory face detection network. Here, the PPI analysis can supplement the activation analysis by giving additional information whether an activated prefrontal brain region has higher functional connectivity during illusory face detection to one or more specific areas of the core system. Second, the activation analysis might not highlight brain regions that are specifically activated by the attempt to detect faces in noise images since these regions can be active independent of the actual detection of a face. These regions, however, are believed to increase their coupling to one or more regions of the core system in case of a successful face detection. These brain regions, clearly being part of the illusory face detection network, can therefore be assessed by a PPI analysis, but not by a classical activation analysis.

⁶ Strictly speaking, from our data we can just infer that the OFC is stronger activated during the no-face detected compared to the face detected condition. To also show that the OFC is “constantly active during the whole task”, we would have to introduce a third condition in which a stimulus is presented without that the subject tries to match the stimulus with internal templates. This design, however, is difficult to implement, at least when the same type of stimulus is presented, since it is not trivial to stop the subject to match it with internal templates. A suitable, but less elegant solution might be to present a different stimulus (e.g. a checkerboard) as low-level control condition (as used in Zhang et al. (2008) and Zimmermann et al. (2019)).

is achieved, the OFC begins to interact with the core system (indirectly via the IFG, see 4.2). These “face detected” signals would explain the observed increased functional connectivity between OFC and core network. After successful face detection, activity in the OFC is subsequently decreased, potentially due to negative feedback signals from the core system or the IFG (“mission accomplished”). This might explain the lower activity in the “face detected” compared to the “no-face detected” condition since in the latter condition the OFC is active for a longer time, continuously trying to detect face patterns in the stimulus.

Our results yield a more differentiated view on the functional role of the OFC in illusory face detection than previous studies. These studies simply postulated that the OFC plays a major role in illusory face detection by its interaction with the core system, without trying to specify its function. A closer analysis of the imaging literature even showed that this assumption is furthermore based on inconsistent findings. On the one hand, the anatomical localisation of the brain regions that were classified as “OFC” highly varied between studies on illusory face detection (see supplementary material S6 for an overview). It is therefore questionable whether these findings can be ascribed to the same region at all. On the other hand, OFC involvement was reported for different analyses. The same research group, for instance, reported OFC involvement in one study for an activation analysis (Li et al. 2010), in another study for a PPI analysis (Li et al. 2009). In yet other studies, no OFC involvement in illusory face detection was reported at all (Zhang et al. 2008; Liu et al. 2014).

In contrast to the OFC, a specific cluster in the ***IFG*** was both stronger activated and stronger connected to core system’s brain regions in the “face detected” compared to the “no-face detected” condition. Different from the OFC, however, the IFG has not been previously ascribed a major role in illusory face detection. This is somewhat remarkable since previous studies on illusory face perception consistently showed an involvement of this region (Zhang

et al. 2008; Li et al. 2009, 2010; Liu et al. 2014), albeit without further discussing it. The findings of IFG involvement in illusory face detection of both this study and previous studies are depicted in Fig. 10 (also see supplementary material S7).

Unlike the non-face specific role of the OFC (see above), one might speculate that the IFG is involved in illusory face detection in a face-specific function. The IFG has been increasingly discussed in recent years as a crucial area of the neural network underlying the perception of real faces. The (right) IFG was for instance recruited when important cues were detected (Hampshire et al., 2010), responded selectively to items that were of most relevance to the currently intended task (Hampshire et al., 2009) and was stronger activated when subjects imagined faces of famous people and answered questions about envisioned facial features like mouth or eyes (Ishai et al., 2002). An extensive overview on the involvement of the IFG in face perception is given in the “revised face perception network” by Duchaine and Yovel (2015).

4.2 Upregulation of brain activity in the core system by prefrontal regions

In analysis 5, we assessed how – and which – regions in the prefrontal cortex upregulate brain activity in the core system during illusory face detection. We constructed models for both hemispheres encompassing V1 as input region, OFC and IFG as potential top-down mediators, and the OFA as the region in the core system that primarily receives input from the prefrontal cortex. We had explicitly chosen the OFA instead of the FFA as recipient of top-down signals from the PFC, in accordance with previous studies on illusory face detection (Li et al. 2010). The OFA is believed to analyse invariant facial features like eyes or mouth and the decision face or no-face is made in this region (Pitcher et al. 2011), while the FFA, on the other hand, is associated with the processing of higher-level information (such as identity, see for example Kanwisher and Yovel 2006). This a priori decision for the OFA was supported by our PPI

results, showing significant stronger functional connectivity between both the OFC and IFG with the OFA compared to the FFA.

While the PPI analysis was largely model-free and simply assessed which regions were coupled with the core system during illusory face detection, the DCM analysis compared different neuroanatomical models describing specific hypotheses on which region in the PFC potentially drives activity in the core network in a top-down manner. Bayesian model selection clearly favoured those models in which OFA activity during illusory face perception was driven by the IFG, not the OFC. Bayesian model averaging showed that OFA activity was upregulated primarily by an excitatory face-specific modulation of the IFG→OFA connection. This modulation led to a 5- to 6-fold increase in connection strength. This modulation was furthermore highly consistent between subjects, being excitatory not only at the group level, but also in each individual subject. Since intrinsic connectivity from OFC to IFG and IFG to OFA was positive as well, a possible role of the IFG could be active integration of information during visual search to determine whether a seen structure resembles a face or not. This information could consequently be transmitted to the OFA in case a face is perceived, which would be explained by the respective modulatory influence.

The model parameters were highly similar between left- and right-hemispheric models. The only exception was the intrinsic coupling from OFC to IFG. This connection showed a significant difference in connectivity strength between the right and left hemisphere, with the latter expressing a stronger coupling which also significantly deviated from zero. This might be due to differences in functional integration or a left-hemispheric bias of the underlying mechanism yet remains speculative in the present model space.

At the group level, we found unexpectedly a negative modulatory effect of perceiving a face on the coupling from OFC to IFG. This finding could be interpreted in the light of a downregulation mechanism. In this case, the OFC can be thought of as a supervisory unit to

allow for decreasing IFG activity as soon as an internal face representation gets putatively matched to a viewed random noise image and visual search can be terminated. One has to be aware, however, that this parameter showed, in comparison to the aforementioned modulatory IFG→OFA parameter, much higher inter-individual variability being excitatory in some subjects, but inhibitory in others.

Our models did not include specific modulatory connections that were able to explain the deactivation of the OFC. We had already speculated that the lower activity in the “face detected” condition might be caused by a negative feedback signal from the core system or the IFG (“mission accomplished”, see 4.1.3). Our models, however, were constructed before data analysis. At this time, we had expected, in accordance with previous reports (Li et al. 2009, 2010), that the OFC would be stronger activated during illusory face detection. Therefore, the modelling of such a connection was not necessary. The post-hoc implementation of for instance a modulatory input on the OFA→OFC connection would have been “double dipping”, i.e. the application of the same data set to establish and test hypothesis. In further studies, using independent data sets, however, we will be able to also test this hypothesis.

4.3 Summary, outlook and limitations

At the neuroanatomical level, illusory face detection has been previously associated with the interaction of prefrontal brain areas with the core system of face perception. More specifically, low spatial frequencies in the perceived noise image are thought to be projected directly from early visual areas over the dorsal magnocellular pathway to the PFC. When a match with internal face templates is achieved, the PFC is assumed to upregulate brain areas in the core system. Our study confirmed, in line with previous studies, that both real and illusory face detection activates the core system of face processing (Zhang et al. 2008; Li et al. 2009, 2010; Liu et al. 2014), more specifically the bilateral OFA and FFA. We additionally showed that

OFA activity during illusory face detection is stronger left-lateralised than during real face processing. We speculated that this atypical lateralisation pattern is associated with a “feature based processing style”. This hypothesis might be tested in future studies for instance by applying face detection paradigms in which the face processing style (feature based vs. holistic) is manipulated.

Our study further showed for the first time that not only one, but at least two bilateral regions in the PFC, namely OFC and IFG, are associated with illusory face detection, albeit in different functional roles. Both regions were functionally coupled with the core system (in particular the OFA) during illusory face detection. The IFG activity however was face-sensitive (i.e. stronger activated during illusory face detection), in line with current neuroanatomical models of face perception, while the OFC showed a reverse pattern, presumably due to its role as a general, non-face specific matching point between external input and internal templates. Future studies can now further investigate these assumptions. The face-specific role of the IFG can be tested more thoroughly for instance by assessing other forms of illusions (e.g. illusory letter detection; Liu et al. 2014). The task-specific role of the OFC can be assessed by additionally introducing task-free baseline events (Zhang et al. 2008; Liu et al. 2014; Zimmermann et al. 2019).

Last, our study provided further evidence on how the PFC upregulates brain activity in the core system during illusory face detection. Our results showed that OFA activity is upregulated primarily by an excitatory face-specific modulation of the IFG→OFA connection rather than the OFC→OFA connection as previously assumed (Li et al. 2010). Future studies can now test neural models in which feedback parameters from the core system to the OFC are incorporated, potentially explaining the higher OFC activity in the “no-face detected” compared to the “face detected” events. Future studies should also implement bilateral models, encompassing both

the left and right hemisphere, similar to previous studies of our group on real face processing (Frässle et al, 2016a, 2016b, 2016c).

Finally, we would like to discuss potential limitations of the present study. One limitation is that to date it is unclear whether the findings from the present study are specific for the illusory face detection paradigm or generalize to other face detection tasks. In future studies, top-down processing during face detection might be also investigated using classical illusion paradigms, typical pareidolia objects (e.g. Akdeniz et al. 2018; Wardle et al. 2020), noisy faces (Summerfield et al. 2006) or ambiguous face stimuli (Meng et al. 2012). One might propose, however, that in all tasks the detection of faces is accomplished by the same mechanisms, i.e. typically by the identification of first-order face information (eyes, nose, mouth) within a stimulus. For face detection, subjects most likely search for specific features in the stimulus that correspond to the LSF of the canonical face template.

A second limitation is related to the slow temporal resolution of fMRI. Face detection is typically an extremely rapid process. Although in the present study the recognition process is substantially prolonged since top-down and bottom-up input have to be integrated multiple times until a decision can be reached whether a face is present in a stimulus, it cannot be excluded that also other cognitive processes are included in the top-down modulation we assess with the DCM analysis. This modulatory process might not only be associated with the facilitation of the face detection process by matching LSF information with internal face templates but could also include for instance active attentional explorations.

A third limitation is related to the forced 50/50 choice of the paradigm. The instruction that faces are hidden in half of the stimuli might skew the response so that subjects (at least those who are less prone to detect a face in noise images) are guessing even if they do not really see a face. The results might therefore be more robust if images are contrasted in which subjects confidently do and do not detect faces, respectively. To tackle this question, future studies could

introduce more response options (e.g. “yes, definitely a face”, “yes, might be a face”, “no, probably not a face”, “no, definitely not a face”). Another possibility would be to change the instructions. Instead of using a 50/50 choice, one might use a 30/70 choice. This will most likely increase the ratio of trials within the face-detected condition in which the subjects are confident to have seen a face (albeit perhaps at the cost of including overall less trials in the face-detected condition).

A further limitation is, at least at first view, the relatively low number of subjects. This might lead some readers to categorize the study out of reflex as underpowered. The study design, however, was motivated, by a previous pilot study, as outlined in detail before (§2.1). We also implemented additional multivariate pattern analyses (see Fig. 11 and 12 for group results). Single subject results and MVPA methods can be found in supplementary material S8. These analyses independently showed, also at the single subject level, that activity in the core system of face perception (OFA and FFA) as well as in the IFG and OFC is highly predictive for the decision whether or not a stimulus is categorized as a face. The mean classification performance was significantly different from a classifier performing at chance level (Fig. 12).

FUNDING

This work was supported by the German Research Council (Deutsche Forschungsgemeinschaft, DFG, Grant no. JA 1890/11-1 to A.J.).

AUTHOR CONTRIBUTIONS

I.T.: conception, data acquisition, data analysis, writing of the manuscript; D.M.H.: conception, data analysis, writing of the manuscript; K.M.Z.: conception, data analysis, writing of the manuscript; M.L.S.: conception, writing of the manuscript; R.K.; conception, data analysis; A.J.: conception, writing of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

The t-maps that support the findings of this study are available in the NeuroVault collection <https://identifiers.org/neurovault.collection:8614> and raw data is available from the corresponding author on reasonable request.

CODE AVAILABILITY

The custom computer code used in this study is available from the corresponding author upon reasonable request.

REFERENCES

- Akdeniz G, Toker S, Atli I. 2018. Neural mechanisms underlying visual pareidolia processing: An fMRI study. *Pakistan J Med Sci.* 34:1560–1566.
- Ashburner J, Friston KJ. 2005. Unified segmentation. *Neuroimage.* 26:839–851.
- Bar M. 2003. A cortical mechanism for triggering top-down facilitation in visual object recognition. *J Cogn Neurosci.* 15:600–609.
- Bar M. 2004. Visual objects in context. *Nat Rev Neurosci.* 5:617–629.
- Bar M. 2009. The proactive brain: memory for predictions. *Philos Trans R Soc Lond B Biol Sci.* 364:1235–1243.
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmid AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E. 2006. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A.* 103:449–454.
- Berman MG, Park J, Gonzalez R, Polk TA, Gehrke A, Knaffla S, Jonides J. 2010. Evaluating functional localizers: The case of the FFA. *Neuroimage.* 50:56–71.
- Bognár A, Csete G, Németh M, Csibri P, Kincses TZ, Sárosi G. 2017. Transcranial Stimulation of the Orbitofrontal Cortex Affects Decisions about Magnocellular Optimized Stimuli. *Front Neurosci.* 11:234.
- Bullier J, Nowak LG. 1995. Parallel versus serial processing: new vistas on the distributed organization of the visual system. *Curr Opin Neurobiol.* 5:497–503.
- Buxton RB, Wong EC, Frank LR. 1998. Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magn Reson Med.* 39:855–864.
- Caspers S, Moebus S, Lux S, Pundt N, Schütz H, Mühleisen TW, Gras V, Eickhoff SB, Romanzetti S, Stöcker T, Stirnberg R, Kirlangic ME, Minnerop M, Pieperhoff P, Mödler U, Das S, Evans AC, Jöckel KH, Erbel R, Cichon S, Nöthen MM, Stürma D, Bauer A, Jon Shah N, Zilles K, Amunts K. 2014. Studying variability in human brain aging in a

- population-based German cohort-rationale and design of 1000BRAINS. *Front Aging Neurosci.* 6:149.
- Corballis MC. 1991. Left brain, right brain. *Science* (80-). 251:575–576.
- Dolan RJ, Fink GR, Rolls E, Booth M, Holmes A, Frackowiak RSJ, Friston KJ. 1997. How the brain learns to see objects and faces in an impoverished context. *Nature.* 389:596–599.
- Duchaine B, Yovel G. 2015. A Revised Neural Framework for Face Processing. *Annu Rev Vis Sci.* 1:393–416.
- Ebner NC. 2008. Age of face matters: age-group differences in ratings of young and old faces. *Behav Res Methods.* 40:130–136.
- Fairhall SL, Ishai A. 2007. Effective connectivity within the distributed cortical network for face perception. *Cereb Cortex.* 17:2400–2406.
- Fan L, Li H, Zhuo J, Zhang Y, Wang J, Chen L, Yang Z, Chu C, Xie S, Laird AR, Fox PT, Eickhoff SB, Yu C, Jiang T. 2016. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cereb Cortex.* 26:3508–3526.
- Fox CJ, Iaria G, Barton JJS. 2009. Defining the face processing network: Optimization of the functional localizer in fMRI. *Hum Brain Mapp.* 30:1637–1651.
- Fox CJ, Moon SY, Iaria G, Barton JJS. 2009. The correlates of subjective perception of identity and expression in the face network: An fMRI adaptation study. *Neuroimage.* 44:569–580.
- Frässle S, Krach S, Paulus FM, Jansen A. 2016. Handedness is related to neural mechanisms underlying hemispheric lateralization of face processing. *Sci Rep.* 6:1–17.
- Frässle S, Paulus FM, Krach S, Jansen A. 2016. Test-retest reliability of effective connectivity in the face perception network. *Hum Brain Mapp.* 37:730–744.
- Frässle S, Paulus FM, Krach S, Schweinberger SR, Stephan KE, Jansen A. 2016. Mechanisms

- of hemispheric lateralization: Asymmetric interhemispheric recruitment in the face perception network. *Neuroimage*. 124:977–988.
- Frey S, Kostopoulos P, Petrides M. 2004. Orbitofrontal contribution to auditory encoding. *Neuroimage*. 22:1384–1389.
- Frey S, Petrides M. 2000. Orbitofrontal cortex: A key prefrontal region for encoding information. *Proc Natl Acad Sci U S A*. 97:8723–8727.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. 1997. Psychophysiological and Modulatory Interactions in Neuroimaging. *Neuroimage*. 6:218–229.
- Friston KJ, Holmes AP, Worsley KJ, Poline J-P, Frith CD, Frackowiak RSJ. 1995. Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp*. 2:189–210.
- Goh JOS, Leshikar ED, Sutton BP, Tan JC, Sim SKY, Hebrank AC, Park DC. 2010. Culture differences in neural processing of faces and houses in the ventral visual cortex. *Soc Cogn Affect Neurosci*. 5:227–235.
- Hampshire A, Chamberlain SR, Monti MM, Duncan J, Owen AM. 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage*. 50:1313–1319.
- Hampshire A, Thompson R, Duncan J, Owen AM. 2009. Selective tuning of the right inferior frontal gyrus during target detection. *Cogn Affect Behav Neurosci*. 9:103–112.
- Harris A, Aguirre GK. 2010. Neural Tuning for Face Wholes and Parts in Human Fusiform Gyrus Revealed by fMRI Adaptation. *J Neurophysiol*. 104:336–345.
- Hasselmo ME, Rolls ET, Baylis GC. 1989. The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behav Brain Res*. 32:203–218.
- Haxby J V. 2012. Multivariate pattern analysis of fMRI: The early beginnings. *Neuroimage*.
- Haxby J V., Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face

- perception. *Trends Cogn Sci.* 4:223–233.
- Haxby J V, Gobbini MI. 2011. *Distributed Neural Systems for Face Perception*. 1st ed. Oxford: Oxford University Press.
- Hillger LA, Koenig O. 1991. Separable mechanisms in face processing: Evidence from hemispheric specialization. *J Cogn Neurosci.* 3:42–58.
- Ishai A. 2008. Let's face it: it's a cortical network. *Neuroimage.* 40:415–419.
- Ishai A, Haxby J V., Ungerleider LG. 2002. Visual Imagery of Famous Faces: Effects of Memory and Attention Revealed by fMRI. *Neuroimage.* 17:1729–1741.
- Ishai A, Ungerleider LG, Haxby J V. 2000. Distributed neural systems for the generation of visual images. *Neuron.* 28:979–990.
- Jansen A, Menke R, Sommer J, Förster AF, Bruchmann S, Hempleman J, Weber B, Knecht S. 2006. The assessment of hemispheric lateralization in functional MRI-Robustness and reproducibility. *Neuroimage.* 33:204–217.
- Johnson MH. 2005. Subcortical face processing. *Nat Rev Neurosci.*
- Kanwisher N, McDermott J, Chun MM. 1997. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J Neurosci.* 17:4302–4311.
- Kanwisher N, Tong F, Nakayama K. 1998. The effect of face inversion on the human fusiform face area. *Cognition.* 68.
- Kanwisher N, Yovel G. 2006. The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci.* 361:2109–2128.
- Kveraga K, Ghuman AS, Bar M. 2007. Top-down predictions in the cognitive brain. *Brain Cogn.* 65:145–168.
- Lacadie CM, Fulbright RK, Rajeevan N, Constable RT, Papademetris X. 2008. More accurate Talairach coordinates for neuroimaging using non-linear registration. *Neuroimage.* 42:717–725.

- Leehey S, Carey S, Diamond R, Cahn A. 1978. Upright and Inverted Faces: The Right Hemisphere Knows the Difference. *Cortex*. 14:411–419.
- Levine SC, Banich MT, Koch-Weser MP. 1988. Face recognition: A general or specific right hemisphere capacity? *Brain Cogn*. 8:303–325.
- Li J, Liu J, Liang J, Zhang H, Zhao J, Huber DE, Rieth CA, Lee K, Tian J, Shi G. 2009. A distributed neural system for top-down face processing. *Neurosci Lett*. 451:6–10.
- Li J, Liu J, Liang J, Zhang H, Zhao J, Rieth CA, Huber DE, Li W, Shi G, Ai L, Tian J, Lee K. 2010. Effective connectivities of cortical regions for top-down face processing: A Dynamic Causal Modeling study. *Brain Res*. 1340:40–51.
- 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.
- Maldjian JA, Laurienti PJ, Burdette JH. 2004. Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *Neuroimage*. 21:450–455.
- Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage*. 19:1233–1239.
- McLaren DG, Ries ML, Xu G, Johnson SC. 2012. A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage*. 61:1277–1286.
- Mechelli A. 2004. Where Bottom-up Meets Top-down: Neuronal Interactions during Perception and Imagery. *Cereb Cortex*. 14:1256–1265.
- Meng M, Cherian T, Singal G, Sinha P. 2012. Lateralization of face processing in the human brain. *Proc R Soc B Biol Sci*. 279:2052–2061.
- Merigan WH, Maunsell JHR. 1993. How parallel are the primate visual pathways? *Annu Rev Neurosci*.

- Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 9:97–113.
- Papademetris X, Jackowski MP, Rajeevan N, DiStasio M, Okuda H, Constable RT, Staib LH. 2006. BioImage Suite: An integrated medical image analysis suite: An update. *Insight J*. 2006:209.
- Penny WD, Stephan KE, Daunizeau J, Rosa MJ, Friston KJ, Schofield TM, Leff AP. 2010. Comparing families of dynamic causal models. *PLoS Comput Biol*. 6:1000709.
- Pitcher D, Walsh V, Duchaine B. 2011. The role of the occipital face area in the cortical face perception network. *Exp Brain Res*. 209:481–493.
- Rauss K, Pourtois G. 2013. What is Bottom-Up and What is Top-Down in Predictive Coding? *Front Psychol*. 4:276.
- Rempel-Clower NL, Barbas H. 2000. The laminar pattern of connections between prefrontal and anterior temporal cortices in the Rhesus monkey is related to cortical structure and function. *Cereb Cortex*. 10:851–865.
- Rhodes G, Brake S, Atkinson AP. 1993. What's lost in inverted faces? *Cognition*. 47:25–57.
- Rossion B, Dricot L, Devolder A, Bodart JM, Crommelinck M, De Gelder B, Zoontjes R. 2000. Hemispheric asymmetries for whole-based and part-based face processing in the human fusiform gyrus. *J Cogn Neurosci*. 12:793–802.
- Rossion B, Hanseeuw B, Dricot L. 2012. Defining face perception areas in the human brain: a large-scale factorial fMRI face localizer analysis. *Brain Cogn*. 79:138–157.
- Smith ML, Gosselin F, Schyns PG. 2012. Measuring Internal Representations from Behavioral and Brain Data. *Curr Biol*. 22:191–196.
- Stephan KE, Kasper L, Harrison LM, Daunizeau J, den Ouden HEM, Breakspear M, Friston KJ. 2008. Nonlinear dynamic causal models for fMRI. *Neuroimage*. 42:649–662.
- Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009. Bayesian model selection

- for group studies. *Neuroimage*. 46:1004–1017.
- Stephan KE, Penny WD, Moran RJ, den Ouden HEM, Daunizeau J, Friston KJ. 2010. Ten simple rules for dynamic causal modeling. *Neuroimage*.
- Summerfield C, Egnér T, Greene M, Koechlin E, Mangels J, Hirsch J. 2006. Predictive Codes for Forthcoming Perception in the Frontal Cortex. *Science* (80-). 314:1311–1314.
- Tsantani mariatsantani M, Kriegeskorte N, Storrs K, Lloyd Williams A, McGettigan C, Garrido luciagarrido L. 2020. FFA and OFA encode distinct types of face identity information. *bioRxiv*. 2020.05.12.090878.
- Wardle SG, Taubert J, Teichmann L, Baker CI. 2020. Rapid and dynamic processing of face pareidolia in the human brain. *Nat Commun*. 11:1–14.
- Weiner KS, Grill-Spector K. 2013. Neural representations of faces and limbs neighbor in human high-level visual cortex: Evidence for a new organization principle. *Psychol Res*.
- Wilke M, Lidzba K. 2007. LI-tool: A new toolbox to assess lateralization in functional MR-data. *J Neurosci Methods*. 163:128–136.
- Wilke M, Schmithorst VJ. 2006. A combined bootstrap/histogram analysis approach for computing a lateralization index from neuroimaging data. *Neuroimage*. 33:522–530.
- Worsley KJ, Friston KJ. 1995. Analysis of fMRI time-series revisited — Again. *Neuroimage*. 2:173–181.
- Yin RK. 1970. Face recognition by brain-injured patients: A dissociable ability? *Neuropsychologia*. 8:395–402.
- Zhang H, Liu J, Huber DE, Rieth CA, Tian J, Lee K. 2008. Detecting faces in pure noise images: a functional MRI study on top-down perception. *Neuroreport*. 19:229–233.
- Zimmermann KM, Stratil AS, Thome I, Sommer J, Jansen A. 2019. Illusory face detection in pure noise images: The role of interindividual variability in fMRI activation patterns. *PLoS One*. 14.

TABLES

Table 1 Brain activation in the core system of face perception elicited by the illusory face perception task

Region	BA	Cluster size	T	p _{FWE-corr}	MNI coordinates		
					x	y	Z
L Inferior Occipital Gyrus (L OFA)	12	339	8.92	0.001	-34	-78	-12
R Inferior Occipital Gyrus (R OFA)	19	3	6.73	0.033	48	-80	-2
L Fusiform Gyrus (L FFA)	37	included in L OFA cluster	8.26	0.003	-38	-58	-18
R Fusiform Gyrus (R FFA)	37	11	7.53	0.010	42	-44	-12

Peak activations in the core system elicited by face responses relative to no-face responses; $p < 0.05$, FWE corrected for multiple comparisons; ROI analysis with a occipital-temporal lobe mask designed with the WFU PickAtlas (Maldjian et al. 2003, 2004). MNI coordinates are given for the peak voxel in each cluster. Additionally, for these voxels, p values FWE corrected for multiple comparisons at the voxel level are displayed.

Table 2 Brain activation in the prefrontal cortex elicited by the illusory face perception task

Region	BA	Cluster size	T	p _{FWE-corr}	MNI coordinates		
					x	y	z
L Inferior Frontal Gyrus (p. Opercularis)	44	1	6.68	0.046	-58	10	10
R Inferior Frontal Gyrus (p. Opercularis)	44	108	7.44	0.015	40	6	28
R Inferior Frontal Gyrus (p. Triangularis)	46	110	9.51	0.001	44	38	10
R Inferior Frontal Gyrus (p. Orbitalis)	47	5	7.19	0.021	38	34	-6

Peak activations in the prefrontal cortex elicited by face responses relative to no-face responses; $p < 0.05$, FWE corrected for multiple comparisons; ROI analysis with a frontal lobe mask designed with the WFU PickAtlas (Maldjian et al. 2003, 2004). MNI coordinates are given for the peak voxel in each cluster. Additionally, for these voxels, p values FWE corrected for multiple comparisons at the voxel level are displayed.

Table 3 Psycho-Physiological Interaction (PPI) for left and right OFA in the prefrontal cortex

Region	BA	Cluster size	T	p _{FWE-corr}	MNI coordinates		
					x	y	z
<i>Left OFA PPI</i>							
R Inferior Frontal Gyrus	8	6	6.72	0.011	36	6	32
L Inferior Frontal Gyrus	8	1	5.80	0.047	-34	10	32
L Inferior Frontal Gyrus	46	2	5.83	0.044	-46	-36	16
L Inferior Frontal Gyrus	45	1	5.83	0.045	-36	-26	14
L Orbitofrontal Cortex	10	1	5.64	0.050	-8	52	-12
R Orbitofrontal Cortex	10	22	9.38	0.000	14	48	-10
<i>Right OFA PPI</i>							
L Inferior Frontal Gyrus	9	4	6.75	0.011	-34	26	18
R Inferior Frontal Gyrus	9	11	6.66	0.013	46	22	28

Connected brain areas in the PFC for “face detected > no-face detected” trials (PPI analysis, $p < 0.05$, FWE corrected for multiple comparisons). ROI analysis with OFC and IFG masks designed with the WFU PickAtlas (Maldjian et al. 2003, 2004), IBASPM116, with a 2D dilation of +1.

Table 4 Results of the Bayesian model selection (BMS)

Family			Model		
Family name	Expected probability	Exceedance probability	Model name	Expected probability	Exceedance probability
<i>Left hemisphere</i>					
F1	0.213	0.042	M1	0.162	0.043
			M2	0.026	0.000
			M3	0.025	0.001
F2	0.152	0.012	M4	0.101	0.012
			M5	0.025	0.001
			M6	0.026	0.000
F3	0.635	0.946	M7	0.549	0.933
			M8	0.045	0.006
			M9	0.042	0.004
<i>Right hemisphere</i>					
F1	0.190	0.032	M1	0.138	0.051
			M2	0.025	0.001
			M3	0.026	0.002
F2	0.163	0.020	M4	0.111	0.030
			M5	0.027	0.002
			M6	0.026	0.002
F3	0.647	0.948	M7	0.461	0.836
			M8	0.131	0.064
			M9	0.055	0.013

Random effects BMS results for the comparison of nine models, carried out for both left and right hemisphere. Models were grouped into three families. Expected and exceedance probabilities are shown separately for each family (family level) and for each model (model level). Posterior evidence was strongest for family F3 for both left- and right-hemispheric models with an exceedance probability of 0.946 and 0.948, respectively.

Table 5 Results of Bayesian model averaging (BMA) for left- and right-hemispheric models

Connection	Median \pm IQR	Range	P
<i>Left hemisphere</i>			
Endogenous parameters (A)			
V1 \rightarrow OFA	0.090 \pm 0.270	-0.110 – 0.655	0.110
V1 \rightarrow OFC	-0.112 \pm 0.243	-0.263 – 0.125	0.131
OFA \rightarrow IFG	-0.014 \pm 0.241	-0.643 – 0.743	0.695
OFA \rightarrow OFC	-0.124 \pm 0.274	-0.408 – 0.718	0.193
IFG \rightarrow OFA	0.247 \pm 0.218	-0.260 – 0.752	0.020
IFG \rightarrow OFC	-0.070 \pm 0.111	-0.400 – 0.079	0.037
OFC \rightarrow OFA	0.256 \pm 0.852	-0.726 – 0.972	0.322
OFC \rightarrow IFG	0.291 \pm 0.474	-0.126 – 0.677	0.020
Modulatory parameters (B)			
“face”: IFG \rightarrow OFA	1.223 \pm 2.173	0.000 – 2.484	0.008
“face”: OFC \rightarrow IFG	-0.568 \pm 3.225	-1.422 – 3.533	0.557
Exogenous input (C)			
“noise”: V1	0.634 \pm 0.529	0.354 – 1.200	0.002
Gating parameters (D)			
IFG: V1 \rightarrow OFA	0.000 \pm 0.079	0.000 – 0.607	0.250
IFG: OFC \rightarrow OFA	0.000 \pm 0.000	-0.200 – 0.031	0.750
<i>Right hemisphere</i>			
Endogenous parameters (A)			
V1 \rightarrow OFA	0.172 \pm 0.402	-0.276 – 0.465	0.193
V1 \rightarrow OFC	-0.079 \pm 0.197	-0.256 – 0.188	0.232
OFA \rightarrow IFG	0.168 \pm 0.663	-0.910 – 0.765	0.322
OFA \rightarrow OFC	-0.189 \pm 0.285	-0.461 – 0.023	0.004
IFG \rightarrow OFA	0.188 \pm 0.287	-0.880 – 0.495	0.131
IFG \rightarrow OFC	-0.048 \pm 0.118	-0.221 – 0.380	0.375
OFC \rightarrow OFA	0.006 \pm 1.394	-0.616 – 1.191	0.432
OFC \rightarrow IFG	0.060 \pm 0.372	-0.153 – 0.409	0.193
Modulatory parameters (B)			

“face”: IFG→OFA	0.745 ± 0.750	0.000 – 2.011	0.004
“face”: OFC→IFG	-0.546 ± 1.879	-2.905 – 2.241	0.232
Exogenous input (C)			
“noise”: V1	0.891 ± 0.493	0.243 – 1.216	0.002
Gating parameters (D)			
IFG: V1→OFA	0.000 ± 0.003	0.000 – 0.722	0.250
IFG: OFC→OFA	0.000 ± 0.000	-0.040 – 0.003	1.000

Group-level results for the endogenous (A), modulatory (B), exogenous (C) and gating (D) parameters after BMA. For each parameter estimate, we present the median, interquartile range (IQR), range and p value. Parameters significantly deviating from zero are printed bold (two-sided Wilcoxon Signed-Rank test, $p < 0.05$).

CAPTIONS TO FIGURES

Figure 1. Stimulus material and classification image. Left: Stimulus material: Example of a pure noise stimulus with the aspect ratio of a face. No stimulus was shown twice. All stimuli were generated individually for each subject. Right: Classification image from a pilot study. The classification image was calculated post-hoc by the difference of the sum of all “face detected” and the sum of all “no-face detected” stimuli. It shows that illusory face detection does not occur randomly, for instance caused by stochastic variations in neural processes, but because some aspects of the noise stimulus match sufficiently with the individual internal representations of faces. If subjects had decided randomly whether or not a noise stimulus is a face, the resulting classification image would have simply shown white noise, not facial structures like for instance eyes and mouth as in this example. More detailed information in supplementary material S1.

Figure 2. Model space. Nine models (M1-M9) were specified for each hemisphere. They were grouped into three families. White arrows depict endogenous connectivity (A-matrix). Red arrows show which connections are additionally modulated by the “face detected” regressor (B-matrix). Yellow arrows represent the driving input (C-matrix) to V1 via the “noise” regressor (“face detected” + “no-face detected” trials). Purple indicates the gating of specific connections by the OFC and IFG, respectively.

Figure 3. Face response rate per session. Although there was never a face present, subjects detected a face in 37.66% (SD = 9.63%) of trials. The face response rate did not change significantly over sessions. Results are presented both at the group level and at the single subject level, showing the variations across the individual participants. Abbreviations: ns – not significant.

Figure 4. Brain activity associated with illusory face detection (group analysis, contrast “face detected > no-face detected”). Illusory face detection activated a large network spanning from frontal over parietal to occipito-temporal regions. Although participants never observed a real face, both the bilateral core network of face perception (i.e. OFA and FFA) and lateral prefrontal (i.e. Insula and IFG), but not orbitofrontal cortices were activated. Of note, for illustrative purposes the brain activation pattern is displayed at $p < 0.001$, uncorrected (cluster size 50). The main activation clusters however were significant at $p < 0.05$, FWE corrected (see Table 1 and 2). Unthresholded t-maps are available on NeuroVault (see data availability).

Figure 5. Distribution of beta weights for “face detected” and “no-face detected” trials in the core network. Beta weights were calculated for the 18-neighbourhood around the peak voxel found for OFA and FFA (see Table 1). Beta weights between face detected and no-face detected trials were compared with two-sided paired t-tests (significant differences are marked with ** for $p < 0.01$, *** for $p < 0.001$, **** for $p < 0.0001$). Please note that subjects whose activity strength (or, more precisely, whose beta weights) strongly differ from the mean group results (i.e. subjects 4 and 9 for the left OFA, subjects 5, 8 and 9 for the right OFA and subject 8 for the right FFA) consistently show the same deviations for both the “face detected” and the “no-

face detected” condition. Subject 9, for instance, has higher beta weights for both conditions compared to the group mean. The difference between both conditions, however, is also for subject 9 comparable to the group difference between conditions. Individual differences thus exist, but are simply caused by an overall higher or lower activity level of some subjects, not by unequal differences between conditions.

Figure 6. *Hemispheric lateralisation of the core network of face perception. Lateralisation indices of BOLD activity in the core system (OFA and FFA) for the illusory face detection task (fuchsia) and the real face localizer task (blue). Negative LI values denote right-hemispheric dominance, positive LI values left-hemispheric dominance. As expected, the face localizer task led to right-lateralised activity in the FFA and bilateral activity in the OFA. While the illusory face detection task elicited a similar brain activation in the FFA, the OFA showed a significantly different (i.e. predominantly left-hemispheric) lateralisation pattern.*

Figure 7. *Psycho-Physiological Interactions (PPI) of the left and right OFA (group analysis, contrast “face detected > no-face detected”). During illusory face detection, the left and right OFA were significantly stronger connected to various brain regions, including bilateral OFC and IFG. Of note, for illustrative purposes, PPI patterns are displayed at $p < 0.001$, uncorrected. OFC and IFG clusters however were significant at $p < 0.05$, FWE corrected for multiple comparisons (see Table 3). Unthresholded t -maps are available on NeuroVault (see data availability statement).*

Figure 8. Distribution of beta weights for “face detected” and “no-face detected” trials in the activation and the PPI analyses. While beta weights for IFG and OFC display the same pattern in the PPI analysis, they show an opposite distribution in the activation analysis. Beta weights were calculated for the 18-neighbourhood around the peak voxel found in the left OFA’s PPI analysis (peak voxel coordinates for each region are denoted in the Figure). Beta weights between face detected and no-face detected trials were compared with two-sided paired t-tests (significant differences are marked with * for $p < 0.05$ and ** for $p < 0.01$). OFC and IFG coordinates were exemplarily chosen based on PPI results for the left OFA.

Figure 9. Averaged DCM connectivity parameter estimates for left- and right-hemispheric models. Each parameter is plotted as the median of the subject parameters after BMA. Endogenous parameters (A-matrix) are white, modulatory parameters (B-matrix) red, extrinsic stimulation parameters (C-matrix) yellow and gating parameters (D-matrix) purple. Parameters significantly deviating from zero are printed bold (two-sided Wilcoxon Signed-Rank test, $p < 0.05$). Parameters that differ significantly between hemispheres are marked with blue circles (paired samples two-sided Wilcoxon Signed-Rank Test, $p < 0.05$).

Figure 10. Involvement of right IFG in illusory face detection across different studies. The right IFG is involved in illusory face detection, as consistently reported across different studies and analyses. All clusters in the right IFG are highlighted with a yellow rectangle. The most reliable overlap is detectable in BAs 44 and 46. Activation further expands into BAs 8 and 9. MNI coordinates are given in supplementary Table S14. Core network regions are marked with a blue circle. A: Present study, activation analysis; $p < 0.0001$ unc. and $k \geq 50$. B: Li et al. (2010), activation analysis; $p < 0.0001$ unc. and $k \geq 50$. C: Liu et al. (2014), activation analysis; $p <$

0.001 unc. and $k \geq 12$. D: Present study, PPI analysis with left OFA as seed; $p < 0.001$ unc. and $k \geq 30$. E: Present study, PPI analysis, with right OFA as seed; $p < 0.001$ unc. and $k \geq 5$.

Figure 11. Overlay of multivariate pattern analysis with univariate effect of the illusory face detection task. All results are displayed both for $p < 0.001$ uncorrected and $p < 0.05$ FWE corrected for multiple comparisons. Violet: Group results from a cross-validated leave-one-run-out MVPA analysis using a Support Vector Machine and a searchlight of 12 mm. A one-sample *t*-test was used to identify cortical areas where the classification performance (AUC) for the “face detected” vs. “no-face detected” condition differed significantly from chance. Red: Group results from the univariate task activation, one-way ANOVA, contrast ‘face detected > no-face detected’. Blue: group results from the univariate task activation, one-way ANOVA, contrast ‘no-face detected > face detected’. Orange: The overlay clearly shows that brain activation in core areas of the illusory face detection network (e.g. OFA, FFA, IFG and OFC) is also highly predictive for behavioural ‘face’ and ‘no-face’ responses.

Figure 12. Classification performance (AUC) distribution across participants. For each ROI the mean classification performance (AUC minus chance) was calculated within a mask based on literature coordinates (OFA and FFA, see analysis 2 (2.6.2)) or anatomical constraints (IFG and OFC, see analysis 5 (2.6.4)). All distributions are significantly different from a classifier performing at chance level (dashed horizontal line); one-sided Wilcoxon Signed-Rank Test, * = $p < 0.05$, ** = $p < 0.001$).