Neural basis of familiar face recognition

Katja Weibert

Doctor of Philosophy

University of York

Psychology

July 2016

Abstract

Familiar identities can be easily recognised across different images. This contrasts with the difficulty in recognising unfamiliar identities across changes in image. Thus, cognitive models propose different neural processing of familiar and unfamiliar faces: familiar identities are suggested to be represented independent of face image (image-invariantly) while unfamiliar identities are suggested to be represented more image-dependently. This image-invariance proposed for familiar faces is suggested to underlie the familiar face recognition advantage. However, the neural correlate of this proposed difference in neural representation of familiar and unfamiliar identities remains unclear. This thesis presents multiple fMRI experiments conducted to investigate the neural correlate of familiar face recognition. The first empirical chapter shows a link between activity in the fusiform face area (FFA) and the familiar face recognition advantage. The second and third empirical chapters investigate the representation of faces within the FFA as well as other face-selective regions. The last empirical chapter demonstrates that the image-invariant representation proposed to underlie the familiar face recognition advantage seems to lie outside the FFA and other core face-selective regions. Taken together, these findings support the suggested role of the FFA in identity processing. Additionally, we show that regions outside the core face-selective regions might be necessary for the recognition of familiar identities.

List of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Timothy Andrews. Thank you for all the guidance and making me think like a researcher. Next, I would like to thank Dr. Karla Evans and Dr. Gary Green as the additional members of my thesis advisory panel. Your comments helped advance my research and made me think more critically about my findings.

Further, I would also like to thank the members of the York Neuroimaging Centre team, particularly Dr. Andre Gouws, who tried to answer all of my questions, as well as Ross Devlin who trained me to scan by myself. Finally, I am extremely thankful to the Psychology Department of York, whose funding allowed me to do my P.h.D. in such an amazing research department.

Und zuallerletzt möchte ich meiner Mama danken. Ohne deinen Glauben an meine Fähigkeiten und deinen ständigen Beistand hätte ich in meinem Leben nicht "nach den Sternen greifen können"!

Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

Chapter 1 – Introduction

We can recognise faces easily, particularly if the face is familiar to us. Familiar faces are recognised across different viewing conditions, facial expressions, and changes in appearance such as hairstyle. However, the neural processes underlying familiar face recognition remain unclear. Models of face perception propose that specific regions of the brain are selective for face processing. These models suggest that familiar and unfamiliar faces are represented differently in these face-selective regions. Using neuroimaging, this thesis investigates the potential difference in neural representations of familiar and unfamiliar faces, and explores the neural correlates of face recognition.

This chapter will: 1) contrast object with face processing; 2) describe models of face processing; and 3) contrast familiar with unfamiliar face processing.

1.1 Faces: a special case of object processing

Behavioural studies suggest a special mechanism for processing faces. Early evidence comes from studies investigating the effects of inversion (for review: Valentine, 1988). Yin et al. (1969) showed that inversion affects face recognition differently than other objects. In his study, participants viewed images of faces and other categories such as houses and were then tested on an old/new recognition task. Stimuli during encoding and test were presented either upright or inverted. For the upright orientation, performance for faces was best. However, for the inverted orientation, performance for faces was worst. Face recognition was disproportionally affected by inversion compared to other object categories. Participants made more than four times as many errors with inverted compared to upright faces. Hence, face recognition seems to differ from object recognition by being more orientation-sensitive.

In line with behavioural findings, neuropsychological case reports show that face processing can be affected independently from object processing. Injuries to occipitotemporal regions have been reported to impair face processing selectively leaving object processing intact: A condition known as prosopagnosia (Farah, 1996; Farah, Levinson, & Klein, 1995; McNeil & Warrington, 1993). For example, Farah et al. (1995) compared recognition ability of the prosopagnosic patient L.H. with normal subjects. They used different exemplars of two different categories: faces and eye glass frames, which similar to faces, show a high within category similarity. Normal subjects performed well for faces but much poorer with eye glass frames (mean accuracy: faces: 85%; eye glass frames: 67%). However, patient L.H. did not show an advantage for faces. He recognised faces and eye glass frames equally poorly (mean accuracy: faces: 64%; eye glass frames: 63%). Thus face but not object recognition in prosopagnosics seems to be severely impaired compared to normal subjects.

However, these results do not necessarily suggest that faces and objects rely on different processing. Face processing might simply be more difficult. Evidence against this alternative explanation provide cases of object agnosia, where participants' object processing is impaired but face processing remains intact (Moscovitch, Winocur, & Behrmann, 1997). Taken together, the observed double-dissociation, where face and object processing can be impaired independently, suggests differential neural processing of faces and other object categories.

Though providing evidence for differential neural processing of faces and other object categories, neuropsychological findings are limited in their generalisation, since they derive from patients with brain injuries. Neural processing in neuropsychological patients might be different compared to healthy participants. However,

electrophysiological studies testing healthy participants also show a difference between neural face and object processing.

EEG studies show a different processing for faces compared to other objects. When participants view faces and non-face objects, only faces elicit a negative evoked response around 170ms after stimulus onset (N170) (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Its face-selectivity is not influenced by task, since it remains significantly larger for faces than other objects regardless of whether faces were attended to or ignored (Carmel & Bentin, 2002). It is observed at electrodes located over the posterior temporal lobe, but due to the limited spatial resolution of electroencephalography its neural source cannot be localised more precisely.

The high spatial resolution of neuroimaging methods allows better localisation of the neural correlate of the differential processing of faces and other object categories. Studies, which compare brain response to faces and other object categories, reveal brain regions such as the fusiform face area (FFA), which are activated more strongly by faces than other objects (Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997). However, faces and objects differ not only on their category. Faces are more homogeneous than other object categories. Additionally, faces show living beings while objects are inanimate. Also, face images show a typical symmetry other objects mostly lack. Thus the observed neural difference between faces and other objects could be explained by these alternative explanations rather than the category itself.

An ingenious study by Kanwisher et al. (1997) systematically tested these alternative explanations. Faces were compared with different images of the same object category (houses) to control for within-category homogeneity and with another animate category (hands) to control for the difference between animate and inanimate. Using face

profiles instead of frontal views allowed to control for face symmetry. Each comparison yielded higher activity for faces than objects in areas such as the FFA ruling out all tested alternative explanations. This finding suggests that the neural processing difference between faces and objects reflects the face-selectivity of these regions.

Nevertheless, whether the stronger FFA response for faces reflects truly face-selectivity has been a matter of active debate. Alternatively, it has been suggested that this regions reflects within-category discrimination among highly familiar stimuli: FFA activity might reflect our encoding of individual faces rather than just the general category of faces automatically engaged due to our expertise with faces (Tarr & Gauthier, 2000). Some evidence for this claim comes from studies investigating expert populations. Experts with many years of experience show FFA activity for non-face objects of expertise such as cars or birds (Gauthier, Skudlarski, Gore, & Anderson, 2000), or even stimuli highly dissimilar to faces like radiographs (Harley et al., 2009) or chess stimuli (Bilalić, Langner, Ulrich, & Grodd, 2011). However, even for experts with many years of experience with a non-face object category, faces still activate the FFA more than non-face objects of expertise. Additionally, non-face objects of expertise have been shown to activate regions of the ventral occipitotemporal cortex beyond the FFA (Grill-Spector, Knouf, & Kanwisher, 2004; Rhodes, Byatt, Michie, & Puce, 2004). Thus the entire ventral occipitotemporal cortex might be involved in object recognition while neurons within the FFA might be engaged selectively by faces.

Combined all these findings show that the category of faces differs from other object categories. Faces result in different behaviour, their processing can be selectively impaired and they evoke different brain responses compared to other object categories.

1.2 Models of face processing

Two influential models have advanced our understanding of face processing. Bruce and Young's cognitive model of face processing (1986), which differentiates between processing of invariant and changeable aspects of faces and Haxby, Hoffman and Gobbini's (2000) neural model of face processing, which proposes potential underlying neural correlates.

1.2.1 Cognitive Model

The most influential model of face processing was proposed by Bruce and Young (1986). According to this model faces are processed via two separate routes. One processes invariant facial features such as identity and the other one changeable aspects such as facial expression. Each function is assumed to be processed by different cognitive systems (Fig. 1.1).
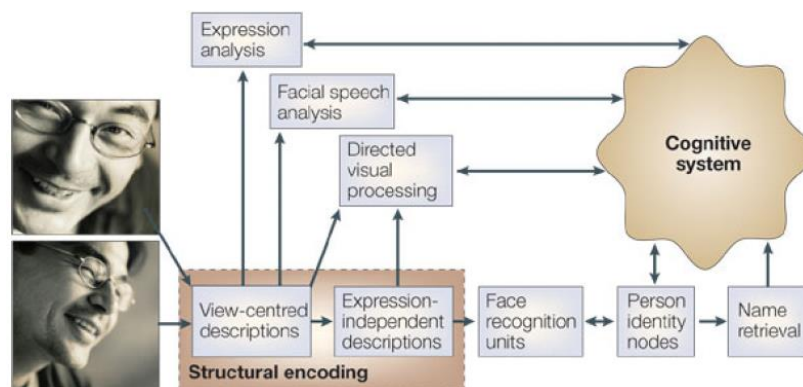
Figure 1.1   *The functional model of face processing (Bruce & Young, 1986). The model proposes two separate routes for processing of identity and changeable aspects. [From (Calder & Young, 2005)].*

Facial identity is proposed to be processed in several stages where its representation becomes more and more abstract. First, faces are encoded pictorially, a representation closely linked to the actual image. This representation is sufficient to recognise identical

images in an old/new recognition task (Bruce, 1982; Hay & Young, 1982). However, this representation is not sufficient to recognise a face across different image changes such as viewing angle, hairstyle or facial expression. A more abstract representation must be established: the structural code.

The structural code is proposed to capture face-invariant aspects which remain constant across different images, a necessity for successful face recognition. It is claimed to be held in face recognition units (FRUs) and to process identity independently of changeable aspects such as facial expressions. From this invariant representation, the FRUs project to corresponding person identity nodes (PINs), causing a sense of familiarity. Activation of the PINs allows the retrieval of names, and other semantic information associated with the person.

Taken together, the model predicts differential processing of identity and changeable aspects such as expressions. The identity representation becomes more and more abstract allowing recognition across different images.

1.2.2   Neural Model

Building upon the Bruce and Young's (1986) model, Haxby et al. (2000) proposed a model trying to explain the neural mechanisms underlying face processing. Numerous studies report a network of brain regions activated more by faces than other object categories (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999; Kanwisher et al., 1997; McCarthy et al., 1997). Haxby et al.'s (2000) model suggests that within this network different areas are involved in processing different cognitive aspects of face processing. The core of this network form the fusiform face area (FFA), the occipital face area (OFA) and the superior temporal sulcus (STS) which are involved in the visual analysis of face stimuli. Within

these core regions, different regions are proposed to process different aspects of faces (Fig. 1.2).
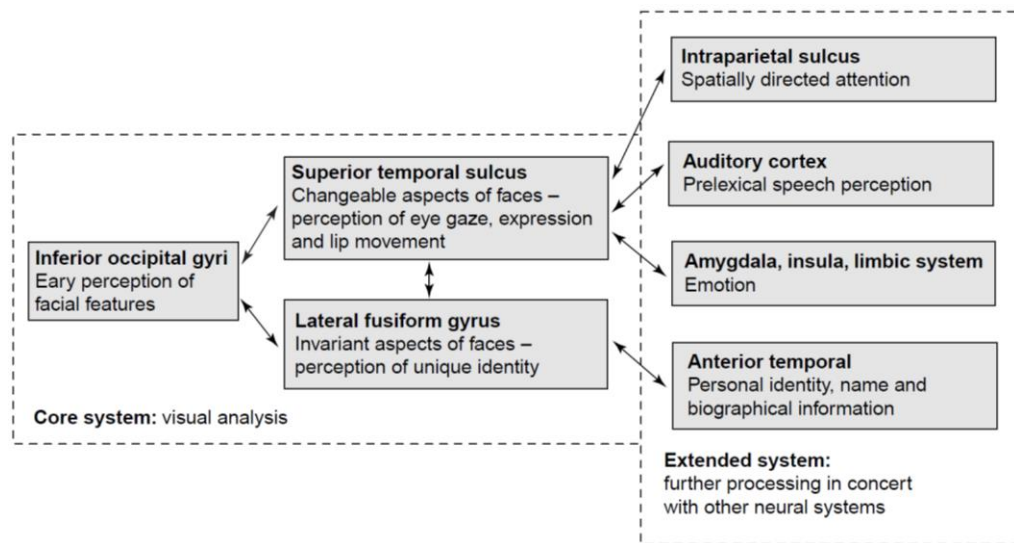
Figure 1.2 *The neural model of face processing (Haxby et al., 2000). The model proposes separate processing of identity and changeable aspects.*

The OFA is proposed to be involved in the early detection of faces (Haxby et al., 2000; Pitcher, Walsh, Yovel, & Duchaine, 2007; Rotshtein, Henson, Treves, Driver, & Dolan, 2004). After the detection, the information is passed to two pathways, a separation analogue to Bruce and Young's model. One pathway projects to the STS processing changeable aspects such as facial expressions (Allison, Puce, & McCarthy, 2000; Engell & Haxby, 2007, Harris, Young, & Andrews, 2012; Pelphrey, Viola, & McCarthy, 2004; Winston, Henson, Fine-Goulden, & Dolan, 2004). The other pathway projects to the FFA which is involved in identity processing.

Numerous studies have shown the FFA's role in processing facial identity (Grill-Spector et al., 2004; Hoffman & Haxby, 2000, Rotshtein et al., 2004). This processing is proposed to be independent of image, thus different images of the same identity should

activate the same representation (image-invariance). Our ability to recognise faces across different images is claimed to rely on this proposed image-invariant FFA representation.

1.3 Differential processing of familiar and unfamiliar faces

1.3.1. Behavioural studies

Although face processing differs from that of other object categories, behavioural studies show differential processing within the category of faces itself. When participants are asked to judge whether two different images depict the same face or different faces, they are more accurate for familiar compared to unfamiliar faces (Bruce et al., 1999, Davies-Thompson, Gouws, & Andrews, 2009, Hancock, Bruce, & Burton, 2000; Jenkins, White, Van Montfort, & Burton, 2011, Megreya & Burton, 2006). Particularly judging two different images as the same unfamiliar face causes participants to make the most mistakes. The familiar face recognition advantage is vast with up to 30% accuracy (Bruce et al., 1999) and is consistently reported across various image changes, such as viewpoint (Bruce, Valentine, & Baddeley, 1987), orientation and expression (Young, McWeeny, Hay, & Ellis, 1986) or poor image conditions (Bruce et al., 1999; Bruce, Henderson, Newman, & Burton, 2001).

These findings have been attributed to a more image-invariant representation for familiar than unfamiliar faces. According to Bruce and Young's cognitive model of face processing (1986) faces are first encoded pictorially. This representation is sufficient to recognise identical images in an old/new recognition task and explains while familiar and unfamiliar faces are recognised equally well, when the same images are used (Bruce, 1982; Hay & Young, 1982). However, it is not sufficient to recognise a face across different image changes such as viewing angle, hairstyle or facial expression. A more

abstract representation is needed which encodes identity image-invariantly (structural code). The difference between familiar and unfamiliar faces is suggested to occur at this stage. While the structural code for familiar faces results from multiple previous encounters, the formation of the structural code for an unfamiliar face will be limited by its initial exposure. Thus the familiar face advantage might result from a richer structural code compared to unfamiliar faces.

1.3.2 Neuropsychology

In line with the behavioural familiar face advantage neuropsychological studies show a dissociation between familiar and unfamiliar face processing (Benton, 1980; Malone, Morris, Kay, & Levin, 1982). For instance, Malone et al. (1982) report a neuropsychological patient who was unable to recognise familiar faces e.g. family members and yet able to match unfamiliar faces. They also observed the inverse pattern of impairment in another patient, who was unable to match unfamiliar faces yet was able to recognise familiar faces. This double dissociation suggests differential neural processing of familiar and unfamiliar faces. However, neuropsychological findings have to be interpreted with reservation as they rely on findings obtained from neuropsychological patients, whose neural processing might be different from that of healthy participants.

1.3.3 EEG

Additional evidence for differential neural processing of familiar and unfamiliar faces comes from electrophysiological studies testing healthy participants. Using electrophysiological methods such as electroencephalography (EEG) allows to investigate the time course of the brain response. Differences between familiar and unfamiliar face processing have been reported as early as 170ms after stimulus onset (Caharel et al.,

2002; Caharel, Ramon, & Rossion, 2014), but are more consistently found at 250ms after stimulus onset (N250r, "r" for repetition effect) (Doerr, Herzmann, & Sommer, 2011; Herzmann, Schweinberger, Sommer, & Jentzsch, 2004; Gosling & Eimer, 2011; Pfütze, Sommer, & Schweinberger, 2002; Schweinberger, Pfütze, & Sommer, 1995). This early difference has been interpreted as the first activation of long-term memory representations retrieved for familiar faces. Other studies have shown differences at later time points, namely after 400ms (Bentin & Deouell, 2000; Paller, Gonsalves, Grabowecky, Bozic, & Yamada, 2000), an effect attributed to the semantic information associated with familiar faces.

A recent study by Caharel, Ramon, and Rossion (2014) investigated differences in EEG signal between familiar and unfamiliar faces and asked which difference is linked to successfully recognition of a face as familiar. They used a familiarity judgement task requiring participants to judge whether a face was personally familiar or unfamiliar. Different EEG response for familiar compared to unfamiliar faces started to emerge at 170ms after stimulus onset. However, only the difference within a later time window (220 to 300ms) successfully predicted response time during correct judgements of a face as familiar. This finding suggests that appr. 200ms are sufficient for a neural response reflecting recognition of whether a face is familiar or not.

Though EEG allows the analysis of the time course of a neural response with a precision of milliseconds, its spatial resolution is low not allowing precise localisation of the neural source of the observed difference. The high spatial resolution of neuroimaging methods can be used to investigate the brain regions reflecting the difference between familiar and unfamiliar face processing.

1.3.4 Neuroimaging

The high resolution of neuroimaging techniques allows to investigate which brain regions are involved in processing familiar and unfamiliar faces. Neuroimaging studies addressing this question have used two main methods: direct contrasting and adaptation. fMRI-contrasts are used to investigate which regions respond more to familiar compared to unfamiliar faces and vice versa. fMR-adaptation is used to investigate whether familiar faces are represented more image-invariantly than unfamiliar faces, a mechanism potentially underlying the familiar face recognition advantage.

1.3.4.1 Direct comparison

Many neuroimaging studies directly compare brain activity elicited by familiar and unfamiliar faces to pinpoint the brain regions differentiating between the two. Identity of familiar faces is better recognised than of unfamiliar faces, thus different processing of facial identity could underlie the familiar face advantage. Hence, one focus of interest is the FFA, since it is involved in processing invariant aspects of a face such as identity (Haxby et al., 2000). However, the results are mixed. Some studies report greater FFA activity for familiar compared to unfamiliar faces (Elfgren et al., 2006; Nielson et al., 2010; Sergent, Ohta, & MacDonald, 1992), while others fail to find such a difference (Gorno-Tempini & Price, 2001; Leveroni et al., 2000; Pierce, Haist, Sedaghat, & Courchesne, 2004) [for a review (Natu & O'Toole, 2011)].

Although FFA involvement is not consistently reported, another region emerges in most studies: the hippocampus. Studies comparing activity of famous or personally familiar faces with unfamiliar faces report stronger hippocampal activity for familiar vs. unfamiliar faces (Elfgren et al., 2006; Leveroni et al., 2000; Nielson et al., 2010; Pierce et al., 2004; Sergent et al., 1992). Hence, the difference between familiar and unfamiliar face

processing might lie outside the core face-selective regions, where according to Gobbini and Haxby's (2007) person knowledge associated with familiar faces is processed. The hippocampus has been shown to abstractly encode facial identity since it is activated by different images of the same identity, its name, and even faces associated with it (Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Therefore, the hippocampus might form the neural substrate of person knowledge activated by familiar faces. Higher hippocampal response to familiar compared to unfamiliar faces might thus reflect the retrieval of semantic information associated with familiar faces, such as the name or profession. As retrieval of semantic information is not possible for unfamiliar faces, this might form the source of the difference between familiar and unfamiliar face processing.

1.3.4.2 Underlying neural representation

Numerous studies did not just directly compare familiar and unfamiliar face processing but investigated potential differences in the underlying neural representations. Of particular interest is whether the neural representation for familiar faces is more image-invariant than for unfamiliar faces, since this could potentially account for the observed familiar face recognition advantage.

An image-invariant representation might be the mechanism underlying the behavioural familiar face recognition advantage. It is defined as a representation which remains the same across different images. Familiar faces are claimed to be represented more image-invariantly than unfamiliar faces, explaining our superior recognition of familiar faces across different images (Bruce & Young, 1986).

Image-invariant representations within the face processing network have been reported for unfamiliar faces. For instance, Andrews and Ewbank (2004) investigated

whether face-selective regions such as the OFA represent faces independently of image size. To test this, they presented the same face image in three different sizes. The OFA adapted across these changes suggesting that this area represents faces independent of image size. An even greater image-invariance has been observed in the FFA.

The FFA represents unfamiliar faces invariant to changes in size, viewpoint, lighting, hairstyle, and facial expressions (Andrews & Ewbank, 2004; Eger, Schweinberger, Dolan, & Henson, 2005; Ewbank & Andrews, 2008; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005a, 2005b; Winston et al., 2004). For instance, Winston et al. (2004) tested FFA response to changes in identity and facial expression. On each trial two face images were shown successively. The identity and facial expression of the second stimulus could independently change or remain the same. FFA response adapted when the identity was repeated. However, no FFA adaptation was observed when identity changed and facial expressions were repeated. These findings supports FFA's role in processing facial identity invariantly of other changes.

Though these findings suggest that FFA processes identity, these findings have a major limitation: Same images rather than different images were used to represent the same identity. The observed adaptation might thus be due to the same image rather than the same depicted identity. Indeed, when different images of the same identity are used, the FFA adaptation seems to vanish (Davies-Thompson et al., 2009; Davies-Thompson, Newling, & Andrews, 2013; Eger et al., 2005; Pourtois et al., 2005a, 2005b; Xu, Yue, Lescroart, Biederman, & Kim, 2009).

Nevertheless, the few studies reporting image-invariant representations using different images of the same identity, do not concordantly report a more image-invariant identity

representation for familiar compared to unfamiliar faces in the FFA (Davies-Thompson et al., 2013; Eger, Schweinberger, Dolan, & Henson, 2005; Pourtois et al., 2005b).

For instance, Pourtois et al. (2005b) tested image-invariance across different images of the same identity in an event-related design for familiar and unfamiliar faces. On each trial two different images of the same identity were presented differing in viewpoint. No image-invariant identity representations were observed in the FFA, neither for familiar nor for unfamiliar faces. In contrast, Eger and colleagues (2005) employed the same design but allowed greater changes between the two different images of each identity (viewing angle, lighting, hairstyle, etc.). They found an image-invariant representation in the FFA for familiar but not unfamiliar faces.

In conclusion, the underlying mechanism of the familiar face advantage remains unclear. A more image-invariant representation for familiar faces compared to unfamiliar faces is proposed. Although greater image-invariance in the FFA for familiar faces has been observed in some studies the results are mixed.

Chapter 2 – fMRI methods review

To investigate the neural correlates of familiar face processing, I used functional magnetic resonance imaging (fMRI). fMRI measures brain responses by tracking changes in blood flow that are associated with increased neural activity. As neurons increase their firing rate, their energy intake increases. Consequently, the transfer of oxygen to those cells via the bloodstream increases as well. This leads to a local change in the supply of oxygenated blood. Ogawa, Lee, Kay, & Tank (1990) showed that this Blood Oxygenation Level-Dependent (BOLD) change caused a change in the local magnetic signal that could be detected with MRI. Thus, changes in the BOLD signal measured by MRI can be used to infer the underlying brain response.

fMRI is an indirect measure of brain response limited by the temporal resolution of the haemodynamic response. Nevertheless, changes in BOLD response have been found to correlate with changes in local field potentials and action potentials which occur within milliseconds (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). Additionally, its high spatial resolution allows the resolution of differences in brain response within millimetres and has thus made fMRI a fundamental technique to investigate brain response. A number of statistical techniques have been proposed for analysing fMRI data. These technique include univariate and multivariate analyses.

2.1 Univariate Analysis

Many fMRI analyses have used a univariate general linear model (GLM) approach to analyse the BOLD signal (Friston et al., 1995). Brain responses, within each voxel or within a region of interest (ROI), are modelled by regressors, e.g. one for each condition. To

model the response to one condition, a box-car model can be used. This model predicts a

non-zero response when this condition is present and a zero response when it is absent.

Since the changes in blood flow do not occur abruptly, the regressor should be convolved

with a hemodynamic response function (HRF) to produce a better model of the expected

time series of brain response. This model can then be used to explain the measured fMRI

BOLD signal within each voxel or ROI (Fig. 2.1).
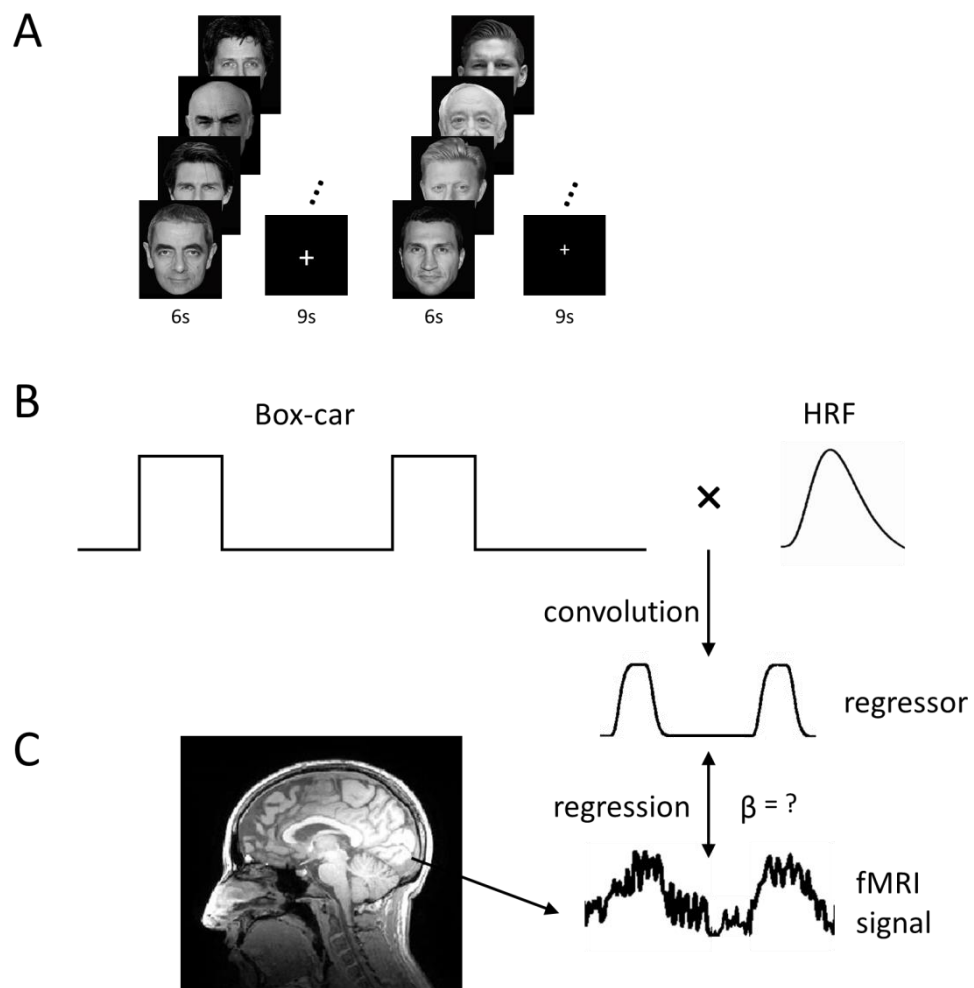
Figure 2.1 *Example of a univariate GLM analysis of fMRI data. A) fMRI experiment where images are*

presented in blocks followed by fixation periods. B) Box-car model of brain response based solely on block

on- and off-sets and hemodynamic response function (HRF), which can be used to convolve the box-car-

model to create a better model. C) Measured fMRI response will be modelled by the convolved box-car-

model.

Modelling the brain response, will generate a statistical map of parameter estimates (regression coefficients) of the whole brain that reflect the model fit at each voxel; Voxels activated by the condition will be predicted well and receive a large coefficient value.

A comparison or contrast between parameter estimates can be used to infer properties of the brain. For example, the parameter estimates can be contrasted against a baseline or the response to other experimental conditions. This analysis is performed independently for each voxel (univariate) and will generate a statistical map depicting voxels which show a significant difference in brain response (Huettel, Song, & McCarthy, 2004). To determine whether these patterns are consistent across different participants, these statistical maps can then be submitted to a higher-level analysis.

2.2 fMR-adaptation

One limitation when using fMRI contrasts is that only one signal is obtained per voxel, thus its resolution is restricted by the voxel size. A voxel is a three-dimensional volume element and forms the smallest unit fMRI can measure (Huettel et al., 2004). However, even voxels of $1mm^3$ contain several hundred thousands of neurons. Therefore, difference in response within a single voxel may either reflect response of all neurons within that population, or alternatively greater change in only a subset of neurons (Fig. 2.2).
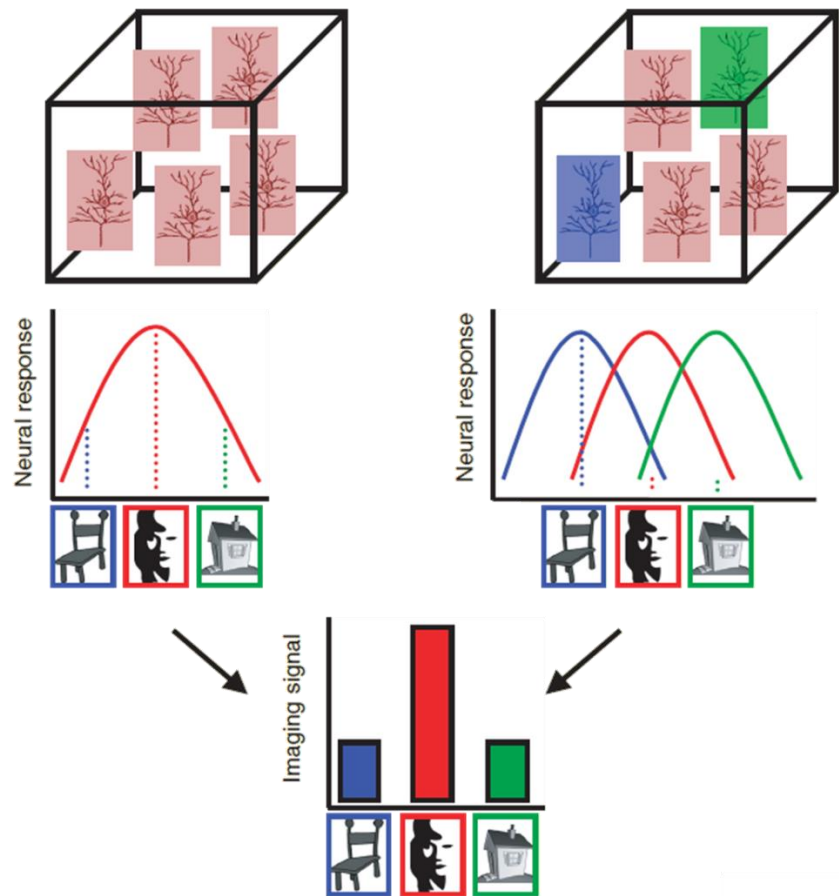
Figure 2.2 *Possible neural ensembles underlying brain response within a voxel (Andrews, 2005). Response from a single voxel may either reflect response of all neurons within that population, or alternatively greater change in only a subset of voxels.*

One way to differentiate between these potential underlying sources of the signal is fMR-adaptation (Grill-Spector & Malach, 2001).

fMR-adaptation allows to target neural populations responding to particular functional properties (Grill-Spector & Malach, 2001). To do so, it relies upon the repetition effect on brain response: when a stimulus is presented repeatedly, fMRI signal decreases, whilst changing the stimulus results in a recovery of brain signal (Henson, Shallice, & Dolan, 2000). fMR-adaptation operationalises these two effects. First, a stimulus is repeated so that neurons responsive to it will adapt (= decrease their firing

rate). Second, a property of the stimulus is varied. Only neurons which are sensitive to the varied property will recover from the adaptation and increase their response. Thus, fMR-adaptation can be used to distinguish between neural populations with different functional properties within a voxel.

2.3 Region of Interest (ROI) analysis

Though fMRI analysis allows simultaneous analysis of the entire brain, it can be beneficial to restrict the analysis to regions of interest (ROIs). A ROI approach focusses on particular brain areas, e.g. the face-selective network. An advantage of the ROI analysis is that it minimises the number of voxels analysed, and therefore decreases the high family-wise error rates associated with the multiple comparisons across many voxels (Poldrack, 2007).

ROIs can be defined in numerous ways. They can be defined anatomically or functionally at a group level and then projected to individual brains, or separately in each individual. A problem with ROI analyses is that size of regions can vary within and across individuals. This creates problems in the interpretation of statistical analyses. To address this issue, I have developed an approach that generates ROIs of the same size for different regions and individuals. This involves measuring the brain response of a fixed number of most selective voxels in each ROI. This way differences between ROIs could not be due to differences in ROI size. To do so, I defined ROIs independently for each participant. Next, the peak voxel within each ROI was localized and finally a flood-fill algorithm systematically decreasing the threshold was used to include a fixed number of most selective voxels. This allowed me to generate ROIs with the same number of voxels across individuals. I varied the size of the ROIs systematically from 25 - 100 voxels in 25

voxel steps to investigate potential effects of ROI size. Finally, this method allowed me to compare activity across different ROIs as well as different participants controlling effects of ROI size.

2.4 Multivariate Pattern Analysis (MVPA)

A limitation of univariate methods is that they analyse each voxel independently. Multi-voxel pattern analysis (MVPA) allows the comparison of the pattern of neural activity across multiple voxels simultaneously, independent of the overall mean response (Haxby et al., 2001). This can be useful when two conditions may produce the same overall mean response within a ROI, but may have different underlying neural patterns. This difference will be obscured by averaging within a ROI resulting in no significant difference between the conditions within this ROI. Therefore, analysing the pattern of neural activity across voxels within a ROI can find differences between conditions that could otherwise be masked by a standard ROI approach.

2.4.1 Correlation-based method

One of the simplest forms of MVPA of fMRI data is the correlation method first proposed in the original study by Haxby et al. (2001). This approach assesses how similar brain response patterns of two conditions are, by simply correlating them.

Firstly, the data is split, e.g. into odd and even runs of the stimulus presentation of each condition. Next, parameter estimates are generated for each condition, separately for each run, using a univariate GLM analysis at each voxel within a region. Then, pairwise correlations are calculated between the patterns of parameter estimates across the splits of the data. This can be done comparing within-condition (e.g. faces-even with faces-odd) and between-condition (e.g. faces-even with houses-odd) (Fig. 2.3).
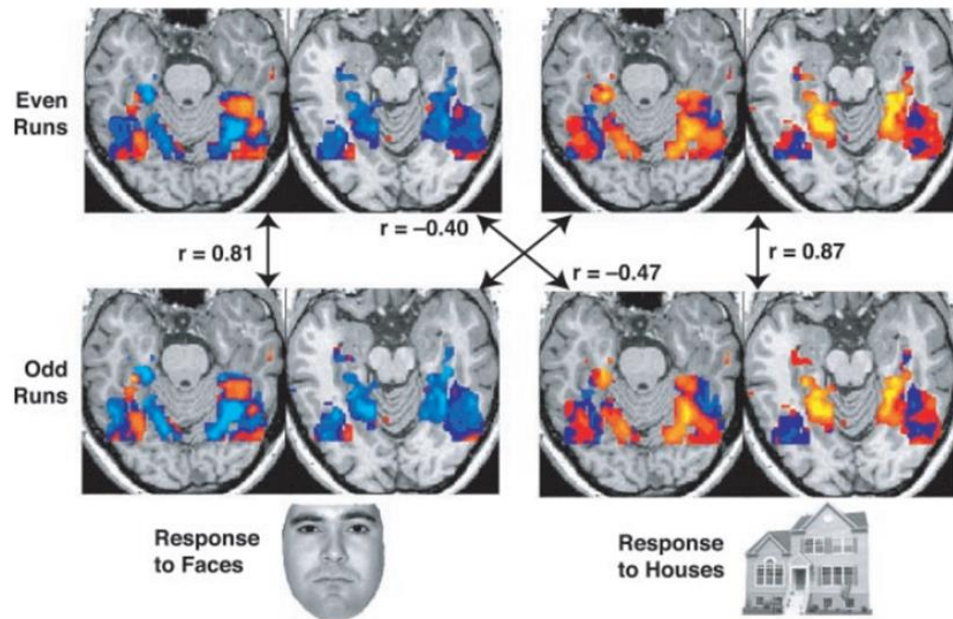
Figure 2.3 *Illustration of a correlation-based MVPA paradigm from Haxby et al. (2001). Patterns were estimated for two stimulus conditions (faces and houses), each across two independent splits of the data (even and odd stimulus runs). Patterns are restricted to a region of interest (ROI), and correlated pairwise within- and between-conditions across the data splits.*

If stimulus conditions evoke distinctive brain response patterns, within-condition correlations should be higher than between-condition correlations. Haxby et al. (2001) demonstrated higher within- than between-category correlations for multiple stimulus conditions, such as faces, houses and chairs, indicating that distinct response patterns to different object categories could be discriminated. Furthermore, this result remained even when the analysis was restricted to only those voxels which did not show a selective univariate response to any condition. This demonstrates that while standard univariate analysis did not yield distinct brain responses to different categories, multivariate analysis was sensitive enough to reveal category-specific brain response patterns.

2.4.2 Representational Similarity Analysis (RSA)

Correlation-based MVPA methods can be used to determine whether conditions elicit distinct response patterns. However, this analysis does not reveal the organizing principles. Representational similarity analysis (RSA) provides a method for understanding what is driving distinctive brain response patterns of conditions; (Kriegeskorte, Mur, & Bandettini, 2008). RSA allows us to test how well a model can explain the observed fMRI data similarity. Multiple models can be tested simultaneously against each other. Thus, RSA can be used to test which model can explain the observed brain response similarities best. Therefore, it contributes to understanding the functional dimension driving the observed brain response patterns.

To perform a RSA at least two similarity matrices are needed, one of the brain response and one of one model. Similarity matrices can be constructed from any data available provided they all correspond to the same set of conditions. Figure 2.4 shows one example where one matrix is constructed from neural data using correlation-based MVPA and another from a model of the stimulus. Next, the similarity matrices are compared to one another, e.g. by correlating the corresponding elements between matrices. If the model is not a good predictor we will observe little similarity between observed brain response pattern similarities and similarities predicted by the model. However, a good model will predict the observed brain response similarities well and result in a high correlation. A high correlation suggests that similar functional dimensions underlie both measures. In this way, we can explicitly test hypotheses about the dimensions underlying the neural response.
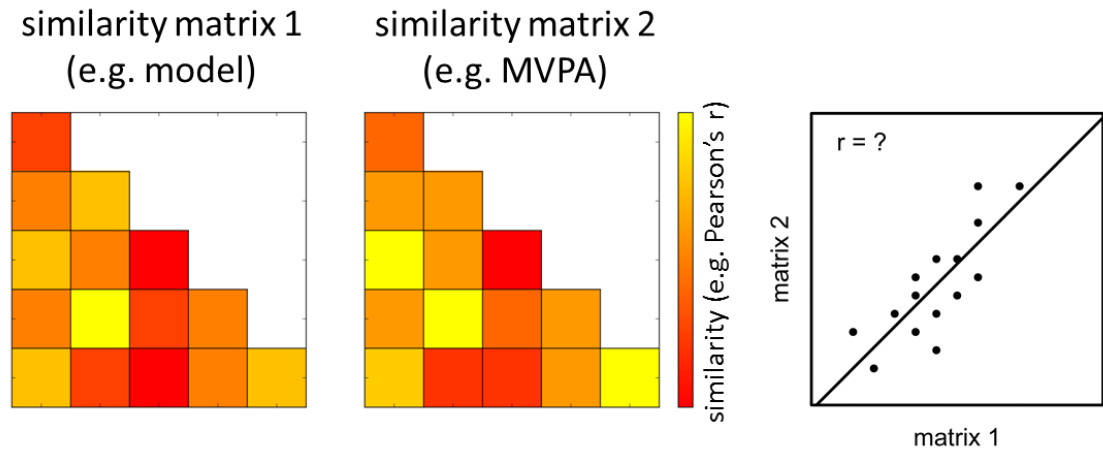
Figure 2.4 *Illustration of the Representational Similarity Analysis (RSA) paradigm (Kriegeskorte et al. ,2008). At least two similarity matrices are constructed, for example from neural data via MVPA, and from a model of the stimuli. To assess the representational similarity these two matrices can be compared, e.g. via correlation. High similarity between the matrices indicates that each measure is a good predictor of the relative similarity between conditions in the other measure.*

Chapter 3 – Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces

This chapter is adapted from: Weibert, K., & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, *75*, 588-596.[1]

3.1 Abstract

People are extremely proficient at discriminating the identity of familiar faces, but are significantly worse with unfamiliar faces. Despite this clear behavioural difference in perception, the neural correlates of the advantage for familiar faces remain unclear. Here, we use an individual differences approach to explore the link between neural responses in face-selective regions and the behavioural advantage for the perception of familiar faces. First, we compared performance on an identity matching task with either familiar or unfamiliar faces. We found that participants were significantly better at matching the identity of familiar compared to unfamiliar faces. Next, we used fMRI to measure the response to familiar and unfamiliar faces. Consistent with the behavioural data, there was a significant difference in the neural response to familiar and unfamiliar faces in face-selective regions. Finally, we asked whether interindividual variation in behavioural performance could be predicted by corresponding variation in fMRI response. We found a significant correlation in the right fusiform face area (rFFA)

---

[1] The author, Katja Weibert, designed the experiment, analysed the results, and wrote the article under the supervision of Prof. Timothy Andrews.

between the difference in response to familiar and unfamiliar faces and corresponding differences on the face-matching task. That is, participants who showed a larger response to familiar compared to unfamiliar faces in the rFFA also matched familiar faces much more accurately than unfamiliar faces. No other face-selective region showed a correlation between neural and matching accuracy. These results provide a link between activity in the rFFA and the perception of familiar faces.

3.2 Introduction

The ability to recognize familiar faces across a variety of changes in illumination, expression, viewing angle, and appearance contrasts with the inherent difficulty found in the perception and matching of unfamiliar faces across similar image manipulations (Bruce, Valentine, & Baddeley, 1987; Davies-Thompson et al., 2009; Hancock et al., 2000; Jenkins et al., 2011; Megreya & Burton, 2006; Johnston & Edmonds, 2009). This difference in perception has been incorporated into cognitive models of face processing, which propose that familiar and unfamiliar faces are represented differently in the human visual system (Bruce & Young, 1986; Burton, Bruce, & Hancock, 1999). These models propose that faces are initially encoded in a pictorial or image-dependent representation. This image dependent representation can be used for the perception and matching of unfamiliar faces. In contrast, the identification of a familiar face requires an image-invariant representation that can be used to recognize familiar identity across different images.

Despite the clear behavioural difference in the perception of familiar and unfamiliar faces, the neural correlates of this difference are much less clear (Natu & O'Toole, 2011). Functional imaging studies have consistently found regions in the

occipital and temporal lobes that respond selectively to faces (Kanwisher et al., 1997; McCarthy et al., 1997). Models of face processing suggest that one region within this core network — the fusiform face area (FFA) — is important for the representation of invariant facial characteristics that are necessary for familiar face recognition (Grill-Spector et al., 2004; Haxby et al., 2000; Ishai, 2008). Consistent with these predictions, neuropsychological studies show that lesions in the approximate location of the FFA can lead to selective impairments in face recognition (Barton, Press, Keenan, & O'Connor, 2002; Damasio, Damasio, & Van Hoesen, 1982). In contrast to evidence from brain lesions, neuroimaging studies that have directly compared the response in the FFA to familiar and unfamiliar faces have shown mixed results. Some studies report stronger FFA activity for familiar compared to unfamiliar faces (Elfgren et al., 2006; Gobbini, Leibenluft, Santiago, & Haxby, 2004; Pierce et al., 2004; Sergent et al., 1992), while others find no difference in response (Gorno-Tempini & Price, 2001; Leveroni et al., 2000).

Support for the role of the FFA in familiar face recognition comes from studies that directly correlate neural responses with behavioural measures of face processing. For example, Grill-Spector and colleagues (2004) showed that the activation in the FFA was higher in trials when participants successfully detected and identified a familiar face than when they did not. Other studies have used an individual differences approach in which neural responses are correlated with behavioural performance in face recognition (Yovel & Kanwisher, 2005; Furl, Garrido, Dolan, Driver, & Duchaine, 2011; Huang, Song, Zhen, Yang, & Liu, 2014). For example, Furl and colleagues showed that performance across tasks that measure face matching and recognition correlated with face-selectivity in the FFA. However, this study included both normal and prosopagnosic participants, so it is not clear whether this relationship was primarily based on the difference between

37

these two participant groups. A more recent study, using a large population of normal participants found that differences in a memory task for faces compared to non-face objects was correlated with the selectivity of the FFA (Huang et al., 2014). However, this study only used unfamiliar faces and only tested recognition performance with identical images. So, it is unclear whether this relationship reflects image-based rather than identity-based face processing. A stronger test for a link between neural activity and the recognition of facial identity is to determine whether this correlation is still evident when different images from the same identity are shown (i.e. image-invariant perception).

The aim of this study was (1) to determine whether there is a difference in the response to familiar and unfamiliar faces in face-selective regions and (2) to determine if individual differences in the response to familiar and unfamiliar faces can predict the behavioural advantage for familiar faces. To address these questions, we first measured the difference in neural response to familiar and unfamiliar faces in face-selective regions using fMRI. Next, participants performed a matching task that involved discriminating whether two different face images belonged to the same identity or a different identity. Participants performed this task with both familiar and unfamiliar faces. Our hypothesis was that individual differences in the matching task for familiar and unfamiliar faces would be reflected in corresponding differences in the neural response in face-selective regions, such as the FFA. To determine the face-specific nature of any relationship between neural and behavioural measures of face processing, we repeated the experiment with familiar and unfamiliar scenes.

3.3 Methods

3.3.1 Participants

40 right-handed participants with normal or corrected to normal vision participated in the experiment. Data from two participants had to be excluded: one participant showed uncorrectable motion artefacts, while data from another participant was withdrawn for routine screening. Data sets from 38 participants were therefore analysed (19 females; mean age: 24.3 ± 1.3 years). All participants gave their written informed consent. The study was approved by the York Neuroimaging Centre Ethics Committee.

3.3.2 Behavioural Experiment

There were two behavioural experiments: a face matching experiment and a scenematching experiment. A matching task was employed since it is commonly used to assess differences in recognition ability between familiar and unfamiliar stimuli (Bruce et al., 2001; Davies-Thompson et al., 2009; Epstein, Higgins, Jablonski, & Feiler, 2007; Young et al., 1986). There were 4 trial types in the face matching task (Fig. 3.1A):

  (i) familiar face same (two different images with the same identity)

  (ii) familiar face different (two different images with different identities)

  (iii) unfamiliar face same (two different images with the same identity)

  (iv) unfamiliar face different (two different images with different identities).

Face stimuli were taken from a previous study (Davies-Thompson et al., 2013) and included male and female identities. Familiar faces consisted of pictures depicting famous UK/US celebrities. Unfamiliar faces depicted celebrities from other countries and were unknown to the participants. Participants had to judge whether the two images depicted the same identity or different identities with a button press. On each trial, a fixation cross

on a black screen was presented for 1.5s.This was followed by a face image for 800ms, a

black screen for 200ms and then a second face image for 800ms. The gender was always

the same in each trial. Eight images of 36 identities (288 images) were used to create 144

trials. Half the trials used familiar faces and half the trials used unfamiliar faces.

Moreover, half the trials included faces from the same identity and half the trials used

faces from different identities. Accuracy and reaction time were measured for each trial.

Response accuracy was converted to d'[z-score (hits) - z-score(false alarms)]. The scene

matching task had an identical design (Fig. 3.1B).
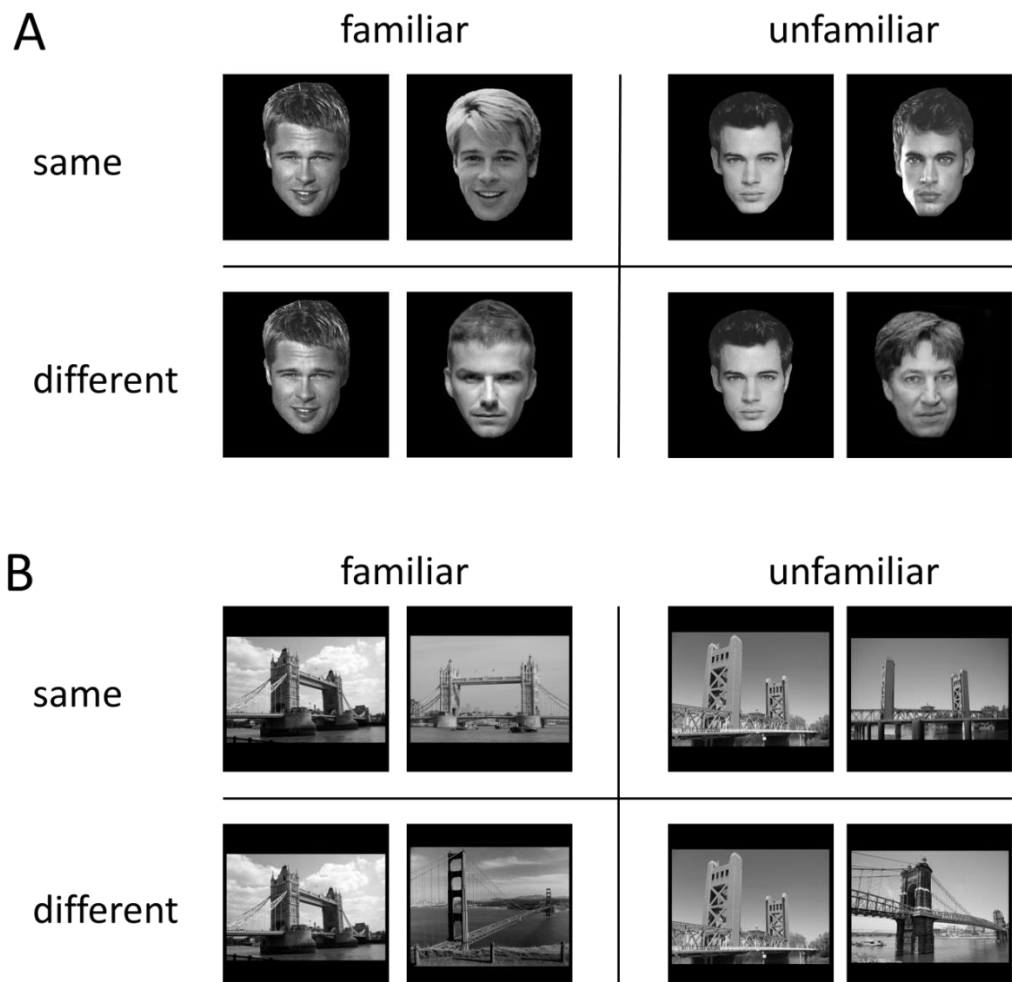
Figure 3.1 *Example trials from the matching task. Participants had to judge (A) whether two successive face images had the same identity or (B) whether two successive scenes depicted the same landmark. Images were either familiar or unfamiliar to the participants.*

Scene stimuli were taken from public sources on the world-wide web and consisted of buildings from the following architectural styles: bridges, domed structures, classical architecture, neoclassical architecture, skyscrapers (art deco), and skyscrapers (high-tech). Pairs of images in each trial were taken from the same subcategory of scene to minimize visual differences between the images.

After the behavioural tasks, participants were given a recognition test. Recognition ability was measured with a questionnaire depicting one image of each of the familiar stimuli used (18 familiar faces, 18 familiar scenes). Participants were asked to name or provide some information of the images. Participants correctly recognized 91.5% (SD ± 13.7%) of the familiar faces and 56.1% (SD ± 17.1%) of the familiar scenes.

3.3.3 Experimental Design

There were 6 stimulus conditions in the fMRI experiment (Fig. 3.2):

(i) familiar faces

(ii) unfamiliar faces

(iii) scrambled faces

(iv) familiar scenes

(v) unfamiliar scenes

(vi) scrambled scenes

Face and scene stimuli showed the same identities and landmarks that were used in the behavioural experiment. However, different images were used. Scrambled images were created by scrambling the phase of the face and scene images. All images were presented in greyscale. Stimuli from each condition were presented in a blocked design.
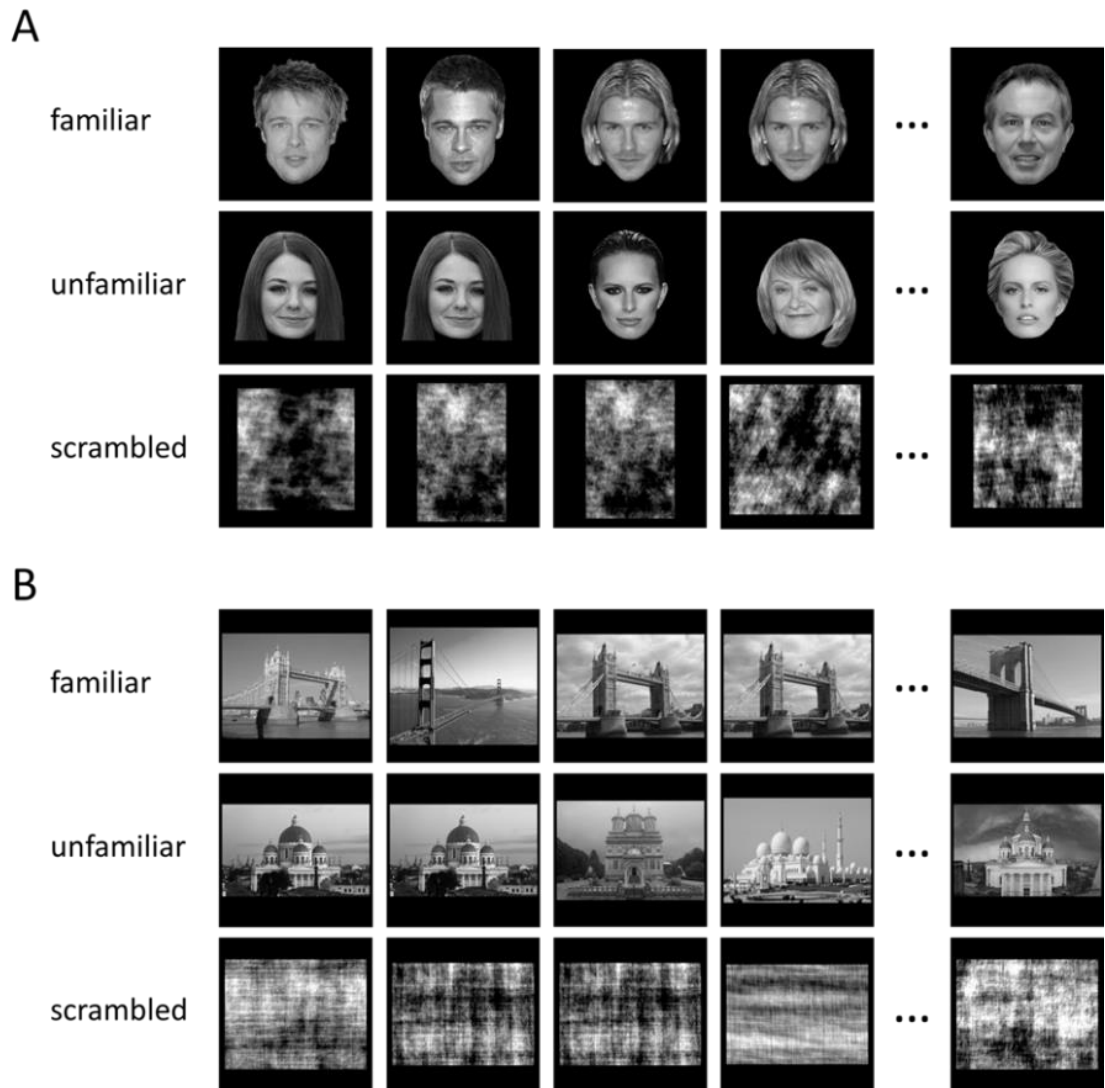
Figure 3.2 *Example blocks of different stimulus conditions in the fMRI experiment: A) face stimuli and B) scene stimuli.*

Faces in each block always had the same gender. Scenes in each block were taken from the same subcategory of scene. Each stimulus block lasted for 9s and contained 9 images. Within each stimulus block, each image was presented for 800ms followed by a 200ms blank screen. Each of the six stimulus conditions was repeated 6 times. This gave a total of 36 blocks. The blocks were presented in a counterbalanced order. Blocks were separated by a 9s fixation screen (a white fixation cross on a black background). To maintain attention, participants performed a one-back task in which participants

responded with a button press every time an identical image was directly repeated. This allowed us to use a common behavioural task across all conditions and minimised the influence of task on neural response (Kanwisher et al., 1997; Davies-Thompson et al., 2009).

3.3.4 Imaging Parameters

Data was collected using a GE 3 Tesla HD Excite MRI scanner at the York Neuroimaging Centre at the University of York. For each participant, a T1-weighted structural MRI (1x1x1mm voxel) and a gradient-echo EPI were acquired. The EPI sequence was recorded with a radio-frequency coil tuned to 127.4 MHz and was used to acquire 38 axial slices (TR = 3s, TE = 33ms, flip angle = 90°, FOV = 260mm, matrix size = 128 x 128, slice thickness = 3mm, voxel size: 2.25 x 2.25 x 3mm). Data was analysed with FEAT version 4.1 (http://www.fmrib.ox.ac.uk/fsl). The first 9 seconds (3 volumes) from each scan were discarded, and MCFLIRT motion correction, spatial smoothing (Gaussian, FWHM 5mm), and temporal high-pass filtering (cutoff 0.0093Hz) were applied.

3.3.5 fMRI Analysis

The BOLD response for each condition was modelled with a box-car function convolved with a standard haemodynamic response function. Face-selective regions were defined by the average statistical maps created by familiar faces > scrambled faces and unfamiliar faces > scrambled faces. Scene-selective regions were defined by the average statistical maps created by familiar scenes > scrambled scenes and unfamiliar scenes > scrambled scenes. Face-selective ROIs were defined independently for each participant. First, the threshold of the face-selective statistical map was adjusted until it divided into individual clusters that corresponded with the core face-selective regions (FFA, OFA, STS). The

cluster function in FSL was used to identify the peak voxel within each cluster. Next, a flood-fill algorithm was used to increase the size of the region incrementally based on the z-scores of individual voxels. This allowed us an objective method to generate ROIs with the same number of voxels across individuals. Our initial analysis was based on a ROI size of 50 most-selective voxels, but we also varied the size of the ROIs from 25 - 100 voxels. Only ROIs where all voxels lied above the minimum threshold of Z = 2.3 were included in the analysis.

To measure the neural response to faces and scenes, the response of each voxel was averaged within a ROI. The units were converted from units of image intensity to percentage signal change (% MR signal change) and normalized to the start of each block (Fig. 3.3).
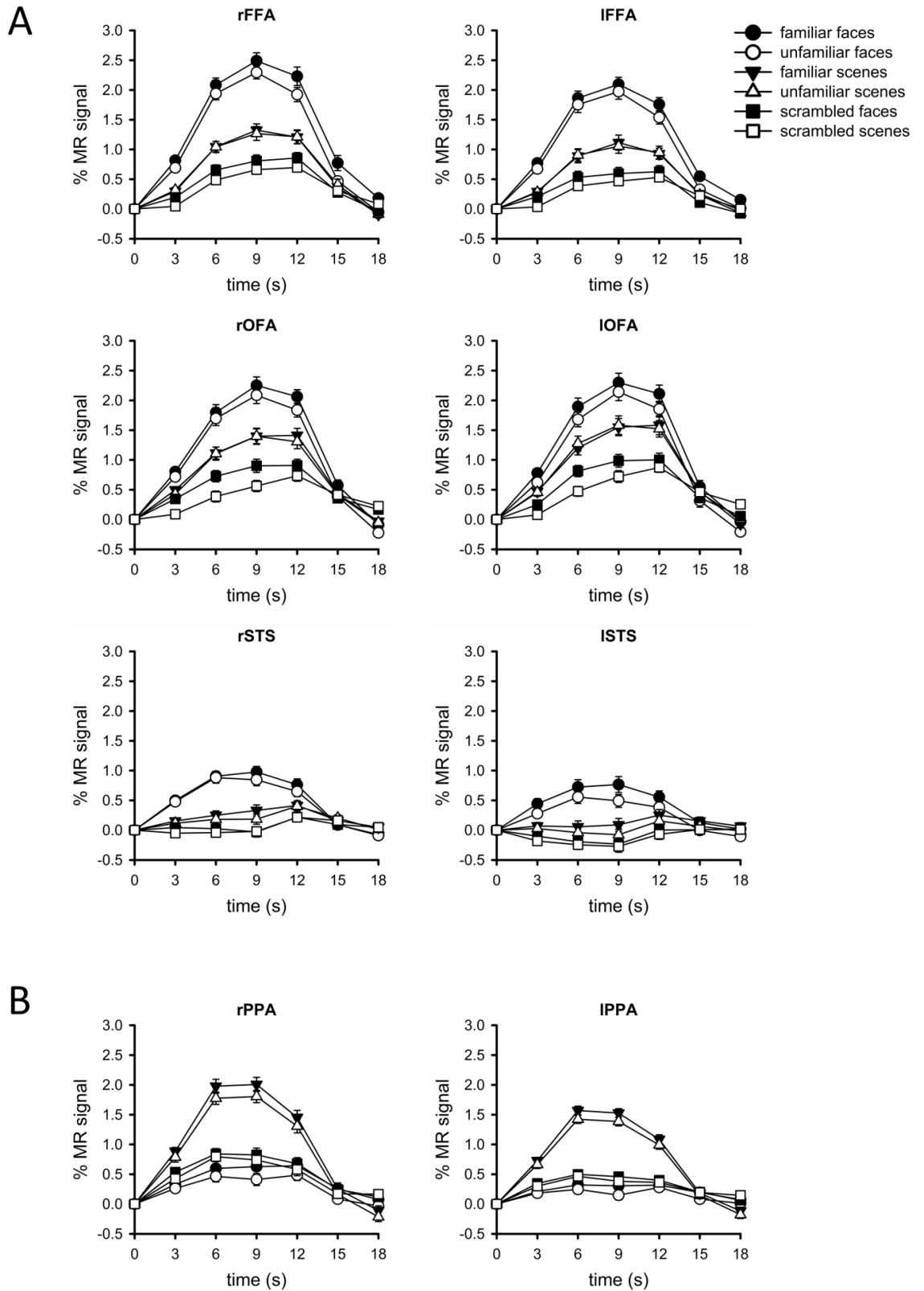
Figure 3.3 *Average time course of response in A) face-selective regions and B) scene-selective regions across participants. Error bars represent 1 SE.*

The peak response across the time series was taken at 9s after the onset of the block for each stimulus condition in each participant. To examine a potential relationship with behaviour, we calculated the difference between peak response to familiar and unfamiliar faces. This normalization is critical to remove overall differences in the amplitude of response across individuals. The relative difference between the neural response to familiar and unfamiliar faces was then compared with corresponding difference between familiar and unfamiliar faces in the behavioural task.

Finally, a whole brain analysis was run to test whether effects observed during the ROI analysis could be found in regions outside the defined ROIs. First, the behavioural difference in d' between familiar–unfamiliar face matching was measured for each participant. These values were demeaned across participants and then used as a regressor in the group analysis of the contrast of familiar > unfamiliar faces. This allowed us to determine if there were any regions in which individual differences in fMRI response between familiar and unfamiliar faces could predict the corresponding difference in matching performance. An analogue analysis was performed for scenes. To account for multiple comparisons, statistical thresholding was carried out using clusters determined by $z > 2.3$ and a corrected cluster significance of $p = 0.05$.

3.4 Results

3.4.1 Face matching task

A matching task was used to determine differences in the perception of familiar and unfamiliar faces. Accuracy (d') in the identification of familiar faces was significantly greater than for unfamiliar faces (familiar: M = 3.56, SD = 0.89; unfamiliar: M = 1.99, SD = 0.51; $t(37) = 11.61$, $p < .001$). Although most participants showed an advantage for

familiar faces, there was individual variation in the magnitude of the effect. Figure 3.4A

shows the range of performance across participants (range: -0.15 – 3.24). Response times

also revealed a familiar face advantage: familiar faces were matched significantly faster

than unfamiliar faces (familiar: M = 0.77s, SD = 0.16s; unfamiliar: M = 0.84s, SD = 0.18s;

t(37) = -6.60, p < .001). There was also individual variation in the difference in response

time to familiar and unfamiliar faces (range: -.26s — .04s).

3.4.2 fMRI response in face-selective regions

The localiser scan revealed a number of face-selective regions: the fusiform face area

(FFA), occipital face area (OFA) and superior temporal sulcus (STS). The peak coordinates

of these regions and the number of ROIs that could be localized in each participant are

shown in Table 3.1. The location of these ROIs correspond closely to those found in

previous studies (Davies-Thompson & Andrews, 2012; Grill-Spector et al., 2004;

Kanwisher et al., 1997).

Table 3.1
Peak coordinates of face-selective and scene-selective regions yielded by group-contrast.

	Hemisphere	MNI coordinates			Number of participants
		x	y	z	
FFA	right	40	−52	−26	38
	left	−42	−48	−24	36
OFA	right	48	−76	−8	37
	left	−40	−82	−18	30
STS	right	46	−66	8	37
	left	−56	−60	6	22
PPA	right	26	−46	−18	37
	left	−28	−46	−22	37

Next, a repeated-measures ANOVA was performed separately for each ROI with the factors Category (faces and scenes) and Familiarity (familiar and unfamiliar). Table 3.2 shows the effect of Category and Familiarity for each ROI. There was a significant main effect of Category for each ROI. This was due to higher responses to faces compared to scenes in face-selective regions and higher responses to scenes compared to faces in scene-selective regions. There was also a significant main effect of Familiarity in each ROI. This was due to higher responses to familiar compared to unfamiliar stimuli. There was no interaction between Category and Familiarity in all ROIs except the rOFA. This interaction was due to a higher response to familiar compared to unfamiliar faces, but no difference in the response to familiar and unfamiliar scenes.

3.4.3 Correlation between fMRI and behavioural responses to faces

To determine whether there was a link between responses in face-selective regions and the perception of familiar faces, we compared individual differences in fMRI response (Fig. 3.4B) with corresponding differences on the face matching task. Table 3.3 shows the correlation between neural response (familiar-unfamiliar) and behavioural accuracy (familiar-unfamiliar) across individuals in all ROIs. The rFFA showed a significant correlation ($r(36) = 0.38$, $p < 0.05$, confidence interval: $0.07 - 0.69$) with accuracy on the face matching task (Fig. 3.4C). The correlation was not dependent on the size of the ROI (Fig. 3.4D). In contrast to the rFFA, no other face ROIs showed a correlation with behavioural accuracy on the face-matching experiment. The rOFA showed a marginal trend towards a correlation, which did not reach significance ($r(35) = 0.29$, $p = 0.08$, confidence interval: $-0.04 - 0.62$). In contrast to accuracy, response times on the face matching task and neural response of any ROI did not correlate significantly (Table 3.4).

48

A   behavioural matching
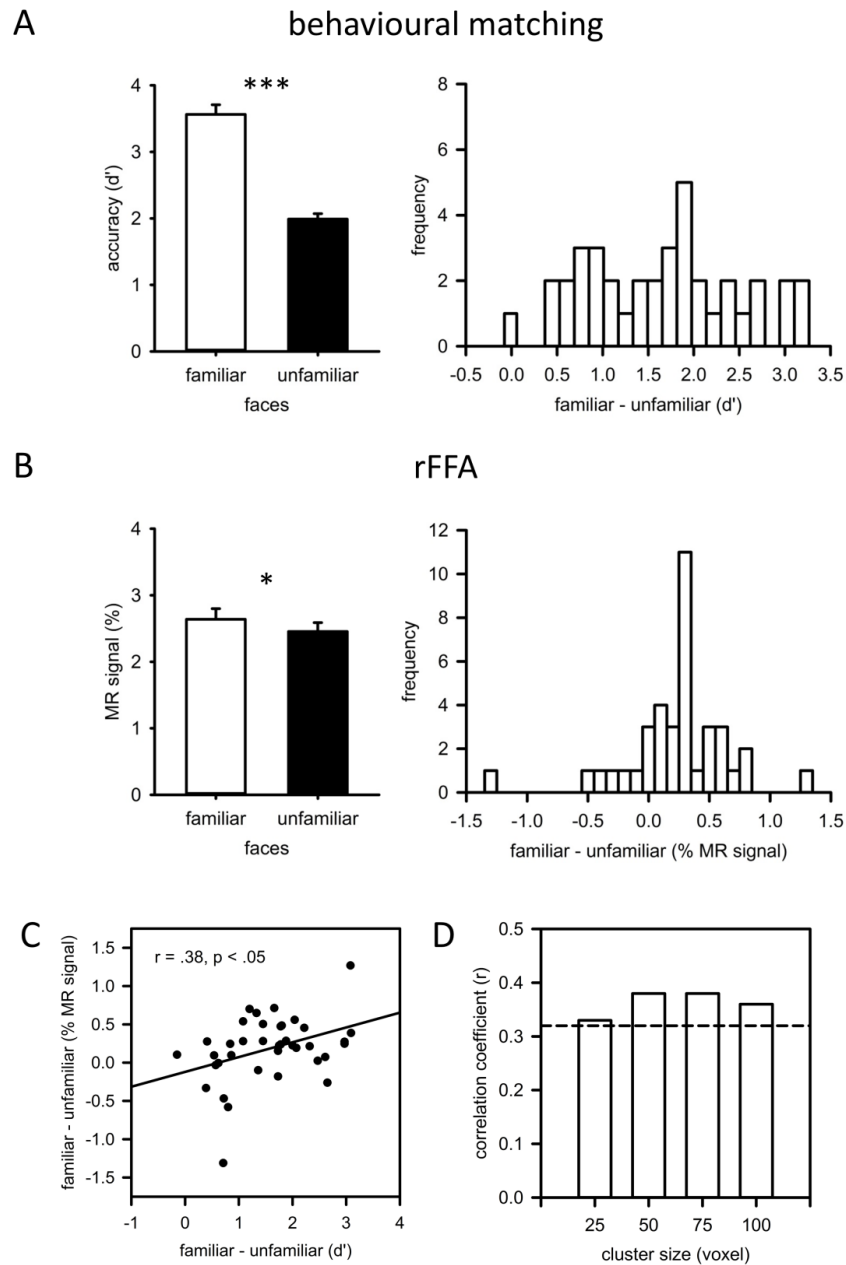
B   rFFA

Figure 3.4   *Behaviour and neural responses to familiar and unfamiliar faces and their correlation.*

(A) behavioural matching performance and (B) rFFA response. Left graphs show group average with SE

bars, right graphs show individual variation in the difference between familiar and unfamiliar faces.

(C) Correlation between the rFFA response and matching performance. (D) The correlation remained

significant when ROI size was varied from 25-100 voxels. This shows a link between responses in the rFFA

and behavioural matching advantage for familiar faces.

To test whether a relationship between fMRI activity and the behavioural matching

advantage could be found in brain regions outside the core face-selective regions, we

performed a whole brain analysis. Consistent with the ROI analysis, a region in the right fusiform gyrus showed a significant correlation (Fig. 3.5 and Table 3.5). Only one other region in the ventral striatum showed a significant correlation between neural and behavioural responses.

3.4.4 Behavioural and neural scene measures

To test whether the patterns of response were specific to face perception, an identical analysis was performed using behavioural and neural responses to scenes. In contrast to faces, we did not observe a significant advantage on the perceptual matching task for familiar scenes. In fact, there was a small advantage for matching unfamiliar scenes in accuracy (famous: M = 3.06, SD = 0.59; unfamiliar: M = 3.45, SD = 0.66; t(37) = 3.93, p < .001) and response time (famous: M = 0.79s, SD = 0.18s; unfamiliar: M = 0.77s, SD = 0.18s; t(37) = 3.60, p < .001). This finding was unexpected, since a matching advantage for familiar scenes has previously been shown (Epstein et al., 2007). Nevertheless, there was individual variation in the difference between familiar and unfamiliar scene matching (d' range: -1.94 — 0.96).
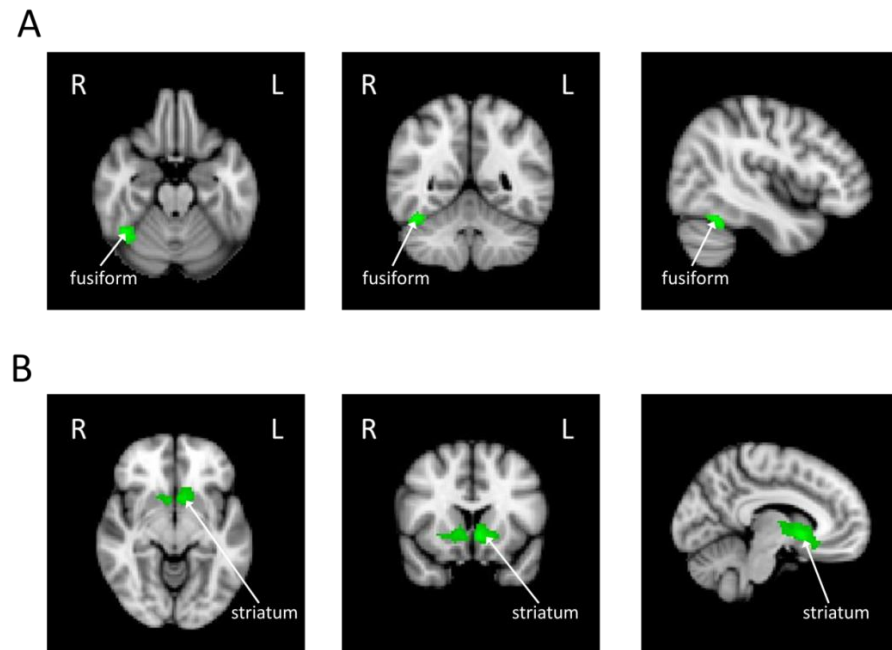
Figure 3.5 *Whole brain regression analysis between the contrast of familiar > unfamiliar faces and behavioural familiar face advantage. This analysis revealed significant clusters in the (A) right fusiform gyrus (fusiform) and (B) the ventral Striatum (striatum).*

Next, we tested whether individual variation in the perceptual matching of scenes could be predicted by activity in the rFFA. The rFFA did not show a correlation between neural responses and either accuracy (Table 3.6) or response time (Table 3.7) to scenes. None of the other face-selective ROIs showed a correlation between neural responses and either accuracy or response time for scenes, except for the rSTS which showed a negative correlation with response time for scenes (r(36) = -.33, p < .05, confidence interval: -0.64 — -0.01).

Then we analysed the response in scene-selective regions using a repeated-measures ANOVA with the factors Category (faces and scenes) and Familiarity (familiar and unfamiliar). The rPPA and lPPA responded significantly more to scenes than faces and to familiar compared to unfamiliar stimuli. There was no significant interaction between Category and Familiarity (Table 3.2).

Table 3.2
Neural response within face-selective regions to all stimulus conditions. Descriptives, mean and standard error, and corresponding statistical results.

ROI		Familiar faces	Unfamiliar faces	Familiar scenes	Unfamiliar scenes	Category	Familiarity	category × familiarity
FFA	right	2.64 (0.16)	2.46 (0.13)	1.28 (0.11)	1.24 (0.12)	$F_{(1,37)}=242.47, p<.001$	$F_{(1,37)}=7.35, p<.05$	$F_{(1,37)}=2.34, p=.135$
	left	2.35 (0.15)	2.22 (0.16)	1.20 (0.13)	1.12 (0.13)	$F_{(1,35)}=121.32, p<.001$	$F_{(1,35)}=7.15, p<.05$	$F_{(1,35)}=0.27, p=.604$
OFA	right	2.41 (0.15)	2.22 (0.15)	1.37 (0.14)	1.37 (0.15)	$F_{(1,36)}=151.25, p<.001$	$F_{(1,36)}=4.90, p<.05$	$F_{(1,36)}=5.02, p<.05$
	left	2.46 (0.17)	2.24 (0.16)	1.57 (0.16)	1.57 (0.17)	$F_{(1,29)}=64.69, p<.001$	$F_{(1,29)}=4.31, p<.05$	$F_{(1,29)}=3.59, p=.068$
STS	right	1.05 (0.10)	0.91 (0.11)	0.34 (0.10)	0.18 (0.09)	$F_{(1,36)}=112.08, p<.001$	$F_{(1,36)}=15.41, p<.001$	$F_{(1,36)}=0.09, p=.766$
	left	0.91 (0.16)	0.55 (0.14)	0.11 (0.13)	−0.09 (0.11)	$F_{(1,21)}=58.31, p<.001$	$F_{(1,21)}=21.07, p<.001$	$F_{(1,21)}=2.90, p=.103$
PPA	right	0.52 (0.75)	0.35 (0.68)	2.03 (0.58)	1.89 (0.70)	$F_{(1,36)}=247.74, p<.001$	$F_{(1,36)}=8.46, p<.01$	$F_{(1,36)}=0.05, p=.817$
	left	0.30 (0.56)	0.09 (0.53)	1.62 (0.54)	1.49 (0.53)	$F_{(1,36)}=186.85, p<.001$	$F_{(1,36)}=9.06, p<.01$	$F_{(1,36)}=0.80, p=.377$

Table 3.3
Correlation between the fMRI response in different regions and accuracy (d') on the behavioural matching of familiar and unfamiliar faces with 95% confidence intervals.

	Right hemisphere	Left hemisphere
FFA	$r=.38, p<.05$, CI=0.07–0.69	$r=.03, p=.88$, CI=−0.32–0.38
OFA	$r=.29, p=.08$, CI=−0.04–0.62	$r=.00, p=.99$, CI=−0.39–0.39
STS	$r=-.08, p=.63$, CI=−0.43–0.26	$r=-.12, p=.60$, CI=−0.34–0.58
PPA	$r=-.03, p=.85$, CI=−0.38–0.31	$r=.16, p=.35$, CI=−0.18–0.50

Table 3.4
Correlation between the fMRI response in different regions and response time on the behavioural matching of familiar and unfamiliar faces with 95% confidence intervals.

	Right hemisphere	Left hemisphere
FFA	$r=.05, p=.79$, CI=−0.29–0.38	$r=.03, p=.85$, CI=−0.31–0.38
OFA	$r=-.02, p=.92$, CI=−0.36–0.33	$r=.04, p=.86$, CI=−0.35–0.42
STS	$r=.00, p=.99$, CI=−0.34–0.35	$r=-.19, p=.41$, CI=−0.64–0.27
PPA	$r=.19, p=.27$, CI=−0.15–0.53	$r=.06, p=.74$, CI=−0.29–0.40

Table 3.5
Whole brain regression analysis between behavioural face matching and the contrast of familiar > unfamiliar faces.

	Z_{max}	MNI coordinates			Size (voxel)
		x	y	z	
Fusiform gyrus	3.27	54	−54	−20	593
Ventral Striatum	4.56	−6	12	−4	1849

Table 3.6
Correlation between the fMRI response in different regions and accuracy (d') on the behavioural matching of familiar and unfamiliar scenes with 95% confidence intervals.

	Right hemisphere	Left hemisphere
FFA	$r=-.31, p=.06$, CI=−0.63–0.02	$r=-.24, p=.16$, CI=−0.58–0.10
OFA	$r=-.16, p=.35$, CI=−0.50–0.18	$r=-.15, p=.43$, CI=−0.53–0.23
STS	$r=.04, p=.80$, CI=−0.30–0.39	$r=.03, p=.91$, CI=−0.44–0.49
PPA	$r=-.20, p=.25$, CI=−0.53–0.14	$r=-.26, p=.12$, CI=−0.59–0.07

Finally, there was no correlation between neural and behavioural responses to scenes in the scene-selective PPA. A whole-brain analysis was used to determine if there were any regions outside the ROIs that were defined that showed a correlation between neural and behavioural responses to scenes. However, we did not find any regions that showed a significant relationship.

Table 3.7
Correlation between the fMRI response in different regions and response time on the behavioural matching of familiar and unfamiliar scenes with 95% confidence intervals.

	Right hemisphere	Left hemisphere
FFA	$r = -.12$, $p = .46$, CI $= -0.46$–0.21	$r = .01$, $p = .94$, CI $= -0.34$–0.36
OFA	$r = .05$, $p = .79$, CI $= -0.29$–0.38	$r = -.13$, $p = .50$, CI $= -0.51$–0.26
STS	$r = -.33$, $p < .05$, CI $= -0.64$–-0.01	$r = -.18$, $p = .43$, CI $= -0.64$–0.28
PPA	$r = -.01$, $p = .94$, CI $= -0.35$–0.33	$r = -.08$, $p = .65$, CI $= -0.41$–0.26

3.5 Discussion

The aim of this study was to use an individual differences approach to determine the link between activity in face-selective regions and the perception of familiar faces. In a behavioural experiment, we found that participants were significantly more accurate when matching familiar faces compared to unfamiliar faces. Similarly, we found that the response in the rFFA was significantly higher to familiar faces compared to unfamiliar faces. Next, we determined whether there was a link between these neural and behavioural measures. Our results show a correlation between neural processing in the rFFA and the behavioural advantage for the perception of familiar faces.

Our results are consistent with previous studies that have shown familiar faces are easily recognized across changes in viewpoint and illumination, whereas unfamiliar faces

are more difficult to recognize across similar image manipulations (Bruce et al., 1987; Davies-Thompson et al., 2009; Hancock et al., 2000; Jenkins et al., 2011; Megreya & Burton, 2006). However, we also found that the magnitude of the difference on perceptual matching varied across participants (d' range = -0.15 – 3.09). To determine the neural correlates of this individual variation in perceptual matching, we measured the response to familiar and unfamiliar faces in the FFA. We found that familiar faces elicited a significantly higher response in the FFA compared to unfamiliar faces (see also Elfgren et al., 2006; Gobbini et al., 2004; Pierce et al., 2004; Sergent et al., 1992). Moreover, the individual variation in neural response in the rFFA was positively correlated with individual differences in perceptual matching. That is, participants who showed a larger difference in the neural response to familiar and unfamiliar faces showed larger differences in perceptual matching of familiar and unfamiliar faces.

A number of previous studies have used an individual differences approach to explore the link between activity in the FFA and behavioural performance in face processing (Bi, Chen, Zhou, He, & Fang, 2014; Furl et al., 2011; Huang et al., 2014; Yovel & Kanwisher, 2005). For example, Yovel and Kanwisher (2005) investigated whether response in any face-selective region could account for the behavioural face inversion effect. They found that individual differences in neural response in the FFA between upright and inverted faces correlated with difference in accuracy when matching upright versus inverted faces. A related finding was reported by Huang and colleagues (2014) who found that individual differences in a face-memory task correlated with face-selectivity in the FFA. However, these studies used unfamiliar faces and only tested performance with identical images. Furl and colleagues (2011) addressed this issue by measuring performance on tests of face perception and recognition that could not be

solved by image matching. In their study, they showed that performance across these tasks correlated with face-selectivity in the FFA. However, this study included both normal and prosopagnosic participants, so it is not clear whether this relationship was primarily based on the difference between these two participant groups. The findings from our study demonstrate a clear link between the recognition of familiar faces and neural processing in the rFFA in normal participants.

The responses in other face-selective regions, such as the OFA and STS, were also higher for familiar compared to unfamiliar faces. However, these regions did not correlate with accuracy on the matching task. There was also no correlation between neural and behavioural measures in the lFFA. This finding contrasts with previous studies that have reported a relationship between perception and neural response in both the right and left FFA (Bi et al., 2014; Furl et al., 2011; Huang et al., 2014; Yovel & Kanwisher, 2005). It is possible that the difference in this study reflects the use of familiar faces and an experimental design that could not be performed by image matching. There was also no correlation between activity in the STS and accuracy on the face matching task. This is consistent with the role of this region in the processing of changeable aspects of faces, such as expression and viewpoint (Allison et al., 2000; Baseler, Harris, Young, & Andrews, 2014; Engell & Haxby, 2007; Harris, Young, & Andrews, 2012).

A number of studies have used fMR-adaptation paradigms to probe the neural basis of familiar face perception in the FFA. Evidence that the FFA is important for face recognition has been shown in fMRI studies in which a reduced response (adaptation) is found for repeated images of the same face image (Andrews & Ewbank, 2004; Ewbank & Andrews, 2008; Grill-Spector et al., 1999; Loffler, Yourganov, Wilkinson, & Wilson, 2005;

Rotshtein et al., 2004). However, a stronger test for whether this demonstrates selectivity for the identity of the face is if the adaptation is still evident when different images from the same identity are shown (i.e. image-invariant adaptation). Although a number of studies have shown image-invariant adaptation to familiar faces in the FFA (Davies-Thompson et al., 2013; Eger et al., 2005; Ewbank & Andrews, 2008; Henson et al., 2000; Loffler et al., 2005; Rotshtein et al., 2004), other studies have failed to find any image-invariant adaptation (Davies-Thompson et al., 2009; Pourtois et al., 2005b).

Studies using multivariate analyses have investigated if the pattern of response in the FFA can predict facial identity. Although early studies failed to find identity-selective patterns in the FFA (Kriegeskorte, Formisano, Sorger, & Goebel, 2007; Natu, Jiang, Narvekar, Keshvari, Blanz, & O'Toole,2010), more recent reports suggest that it is possible to decode identity in the pattern of response from the FFA (Nestor, Plaut, & Behrmann, 2011; Anzellotti, Fairhall, & Caramazza, 2013; Goesaert, & Op de Beeck, 2013, Axelrod & Yovel, 2015).

To investigate whether regions outside the face processing network could predict the behavioural familiar face matching advantage, we performed a whole brain analysis using difference in matching performance between familiar and unfamiliar faces as a regressor. Consistent with the ROI analysis, we found a significant effect in a region of the right fusiform gyrus overlapping with the average rFFA region of the ROI analysis. This analysis also revealed a region in the ventral striatum. The ventral striatum is part of the reward network (Haber & Knutson, 2010; Olds & Milner, 1954). It is not clear why this region was correlated with behavioural performance. However, it has also been shown to be activated by attractive faces which were not associated with any reward outcome

(Aharon et al., 2001). Therefore, it is possible that this relationship might reflect other aspects of the images that covary with familiarity.

A number of studies have suggested that the FFA may be selective for within-category discrimination of familiar objects rather than being specifically face-selective (Gauthier et al., 2000; Harley et al., 2009; McGugin, Gatenby, Gore, & Gauthier, 2012). For example, McGugin, Gauthier and colleagues (2012) reported that individual differences in the behavioural ability of car experts could be predicted by the difference in the selectivity of the FFA for cars. To address this issue, we compared behavioural performance on scenes with the response to scenes in the rFFA. Scenes were chosen since they are visually distinct from faces and, similarly to faces, there is a network of brain regions highly selective for scenes, like the PPA. In contrast to faces, we found no correlation between the difference in rFFA response to famous and unfamiliar scenes and corresponding differences in accuracy on the perceptual matching task. We also did not see any correlation between these neural and behavioural measures in the PPA. One explanation could be that participants were not familiar enough with the famous scenes used in this study. Only about half of the famous scenes were reported as being familiar in the recognition test. Our results contrast with previous studies which have shown a matching advantage for familiar scenes across different views (Epstein et al. 2007). However, in this previous study, they used pictures of personally familiar scenes. Thus our famous scene stimuli might have not been sufficiently familiar, which is in line with the low recognition score for famous scenes.

In conclusion, our results show that individual responses in the rFFA response can predict individual differences in perception of familiar and unfamiliar faces. This

relationship seems to be specific to the rFFA as it was not observed in any other face-selective region. Taken together our results demonstrate a link between activity in the rFFA and the perception and recognition of familiar faces.

Chapter 4 – Spatial patterns of response in face regions to identity, viewpoint, and expression can be predicted by variance in image properties

This chapter is adapted from: Weibert, Flack, Young & Andrews (2016). Spatial patterns of response in face regions to identity, viewpoint, and expression can be predicted by variance in image properties. *(in preparation)*[2]

4.1 Abstract

Models of face processing propose that the fusiform face area (FFA) plays an important role in processing facial identity. For example, patterns of response in the FFA can successfully decode facial identity across changes in viewpoint or expression. However, other studies have reported that the patterns of response in the FFA can decode viewpoint or expression across different identities. Thus, the relative role of identity, viewpoint, and expression on the pattern of response in the FFA remains unclear. Here, we used fMRI and a correlation-based MVPA to explore how identity, viewpoint, and expression are represented in face-selective regions. First, we contrasted the relative role of identity and viewpoint on patterns of response in face regions. Faces from three famous identities were shown from three different viewpoints. A representational similarity analysis of the fMRI data showed that viewpoint explained the pattern of response within face-selective regions significantly better than identity. Next, we compared effects of two changeable aspects, viewpoint and expression. Again, viewpoint predicted brain response within face-selective regions significantly better than expression. Finally, we asked whether lower-level image properties could explain these

---

[2] The author, Katja Weibert, designed the experiment, analysed the results, and wrote the article under the supervision of Prof. Timothy Andrews.

findings. Lower-level image properties predicted the neural patterns in all face-selective regions. Our results suggest that coarse topographical organization of the face regions is based on the image properties conveyed by facial cues.

4.2 Introduction

The identity of familiar faces is easily recognised across different viewing conditions, such as changes in viewpoint and expression. Models of face processing propose separate processing of information about identity and changeable aspects of faces, such as viewpoint and expression (Bruce & Young, 1986; Haxby et al., 2000). While the early face processing occipital face area (OFA) is proposed to be sensitive to the visual properties of the face, the fusiform face area (FFA) is thought to show an image-invariant representation. This image-invariant representation might underlie our familiar face recognition ability across changes in viewpoint and expression (Bruce & Young, 1986; Burton, Bruce, & Hancock, 1999).

Studies using multi-voxel pattern analysis (MVPA) provide evidence for an invariant representation of identity in the spatial pattern of response in the FFA (Anzellotti et al., 2013; Guntupalli, Wheeler, & Gobbini, 2016; Natu et al., 2010; Nestor et al., 2011). For instance, Anzellotti and colleagues (2013) measured brain response to different identities shown from different views. Then they trained a classifier on brain responses to all-but-one view of two identities and finally tested whether it could decode identity from the left-out view. Brain responses in the FFA allowed successful decoding of identity across different views. In line with these findings, other MVPA studies successfully decoded identity from FFA response across other changes such as facial

expressions (Nestor et al., 2011; Axelrod & Yovel, 2015) suggesting that the FFA represents the identity of a face change-independently.

However, these findings contrast with other studies showing that there are distinct representations of changeable aspects of faces in the FFA (Axelrod & Yovel, 2012; Kietzmann, Swisher, König, & Tong, 2012; Said, Moore, Engell, Todorov, & Haxby, 2010; Wegrzyn et al., 2015; Zhang et al., 2016). For instance, Axelrod and Yovel (2012) measured brain response to different viewpoints shown of different identities. Then they trained a classifier to decode the view from the brain response. They could successfully decode viewpoint across different identities from FFA response. Thus while some studies report change-independent identity representations, others report identity-independent representations of changeable aspects within the FFA.

The ability to decode identity and viewpoint in face regions might be explained by overlapping representations of viewpoint and identity. Freiwald and Tsao (2010) recorded from face-selective neurons in the macaque temporal lobe and found that, while posterior temporal regions represented the viewpoint of faces, a more view-invariant identity representation emerges towards more anterior regions. Similar findings have been reported by Guntupalli and colleagues (2016) who found that the identity representation became progressively view-invariant towards anterior regions of the temporal lobe. In this study, the response in the FFA contained viewpoint and identity information suggesting overlapping representations in this region might reflect an intermediate stage in the processing of facial identity.

An alternative explanation for the organization of face regions could be that they represent lower-level image properties. Images of different viewpoints vary in terms of lower-level image properties. Thus a region differentiating between different viewpoints

might in fact differentiate between different lower-level image properties associated with changes in viewpoint. Indeed, previous studies which successfully decoded identity or viewpoint information from FFA, also report successful decoding from the early face processing region OFA (Anzellotti et al., 2013; Natu et al., 2010; Wegrzyn et al., 2015) suggesting that image information might contribute to the reported successful decoding. Support for this alternative explanation provide studies reporting response in regions within the ventral stream to reflect lower-level image properties (Andrews, Watson, Rice, & Hartley, 2015; Rice, Watson, Hartley, & Andrews, 2014; Watson, Young, & Andrews, 2016).

The aim of this study was thus to investigate the relative contribution of identity, viewpoint, and expression on brain response in face-selective regions. Additionally, we tested whether the observed findings could be alternatively explained by lower-level image properties associated with these facial cues. To address this question, we performed two fMRI experiments. The first one compared identity and viewpoint effects using three identities shown from three different viewpoints (9 conditions). The second experiment used an analogue design showing three expressions from three different viewpoints. Finally, we compared lower-level image properties across conditions. We predict that if a region represents identity, different views of the same identity would elicit similar brain responses. On the other hand, if a region represents viewpoint independent of identity, we predict that same views of different identities elicit similar brain responses. However, if brain response in face-selective regions reflects lower-level image properties, we predict that conditions with similar lower-level image properties would elicit similar brain responses.

4.3 Methods

4.3.1 Participants

For experiment 1, data were collected from 19 participants (7 males, mean age = 22.8 ± 1.0 years). For experiment 2, data were collected from 24 participants (9 males, mean age 23.5 ± 2.5 years). All participants were right-handed with normal or corrected to normal vision. Written informed consent was obtained from all participants. The study was approved by the York Neuroimaging Centre Ethics Committee.

4.3.2 Experiment 1

Experiment 1 compared the effects of the invariant aspect identity and the changeable aspect viewpoint on brain response within face-selective regions. To investigate identity and viewpoint processing in face-selective regions, three different identities (Brad Pitt (BP), David Beckham (DB), and Justin Timberlake (JT)) were shown from three different views resulting in 9 stimulus conditions (Fig. 4.1):

(i)   Brad Pitt, left

(ii)   Brad Pitt, frontal

(iii)   Brad Pitt, right

(iv)   David Beckham, left

(v)   David Beckham, frontal

(vi)   David Beckham, right

(vii)   Justin Timberlake, left

(viii) Justin Timberlake, frontal

(ix)   Justin Timberlake, right
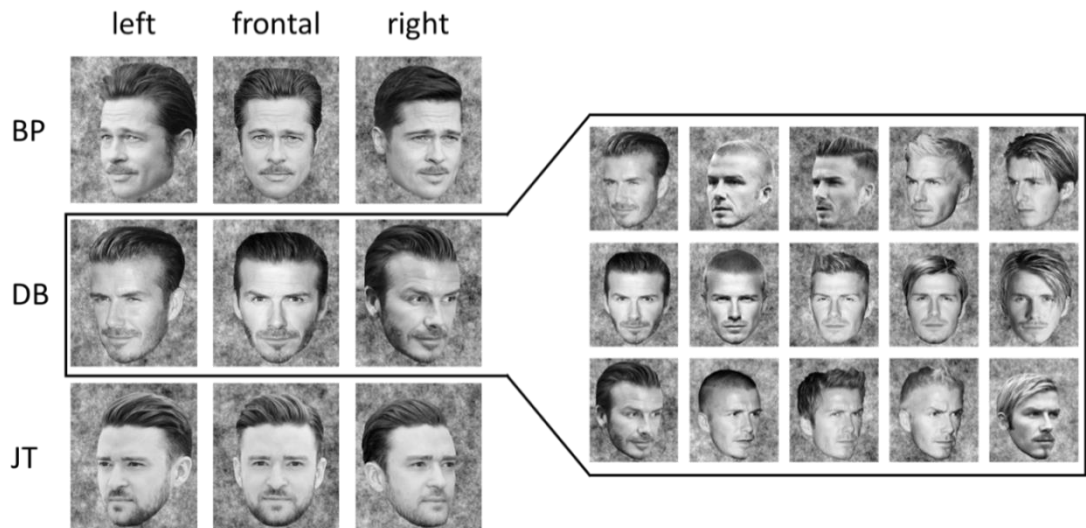
Figure 4.1 *Experiment 1: Experimental design and example blocks of one identity. Each identity was shown from three viewpoints (left, frontal and right). Within each block images shared identity and viewpoint. To measure response independent of image five different images were used in each block varying in hairstyle, lighting etc..*

Stimuli were taken from a variety of internet sources using Google Images and were of three famous male identities: Brad Pitt, David Beckham, and Justin Timberlake. The use of famous identities had two advantages. First, it ensured recognition across different different viewpoints, which has been reported to be impaired for unfamiliar identities (Bruce et al., 1999; Davies-Thompson et al., 2009; Hancock et al., 2000). Personally familiar stimuli are also easily recognised across viewpoint changes. However, they have the disadvantage, that the stimulus sets must be individually tailored to each participant. Thus stimuli sets differ between participants in terms of lower-level image properties making them unsuitable for our experiment. Second advantage of using famous identities was that it allowed us to select different images for each condition. This ensured, that response to each condition does not reflect response to a certain image but rather across different images. Additionally, we could roughly match the appearance of our identities in

terms of hairstyles etc. as well as control lower-level image properties of this newly created stimulus set (see Table 4.1 for analysis of image properties).

Prior to scanning we ensured that participants were familiar with our stimuli. Participants were asked to name the identities used. All participants could name all identities. After scanning, we also checked that participants could recognise the identities from the different images used in our experiment. Participants performed an identity matching task on the images used during the scan. On each trial they were shown two different images sequentially and asked whether they depicted the same identity or different identities. Performance was at ceiling (mean accuracy: 93.69%, ± 3.21%) suggesting that participants were very familiar with the identities shown.

Five different face images were used in each condition. Greyscale versions of these images were presented in a blocked design. During the scan a one-back task was used to maintain attention, where participants responded with a button press every time an identical image was directly repeated (one target per block). Each stimulus block contained 6 images (5 unique images, 1 repeated). Performance was at ceiling (mean accuracy: 98.0% ± 0.01). Within each stimulus block, each image was presented for 800ms followed by a 200ms blank screen. So, each stimulus block lasted 6s. Blocks were separated by a 9s fixation screen (a white fixation cross on a black background). Experiment 1 consisted of two experimental runs, each with 4 repetitions of each of the nine stimulus (total of 36 blocks per run) presented in a counterbalanced order.

4.3.3. Experiment 2

Experiment 2 compared the effects of the two changeable aspects, expression and viewpoint, on brain response within face-selective regions. Face images were taken from the Radboud Faces Database (Langner et al., 2010) and depicted different people posing three different expressions (happy, disgust, and fear) taken from three different viewpoints. This resulted in 9 stimulus conditions (Fig. 4.2):

      I.   happy, left

      II.   happy, frontal

      III.   happy, right

      IV.   disgust, left

      V.   disgust, frontal

      VI.   disgust, right

      VII.   fear, left

   VIII.   fear, frontal

      IX.   fear, right

Figure 4.2 *Experiment 2: Experimental design and example blocks of one expression. Each expression was shown from three viewpoints (left, frontal and right). Within each block images shared expression and viewpoint. To measure response independent of image images depicting five different identities were used.*

66

A red-spot task was used to maintain attention during the scan, where participants responded with a button press every time a red spot appeared (one target per block). Performance was at ceiling (mean accuracy: 99.0% ± 1.9). Each stimulus block contained 5 images. Within each stimulus block, each image was presented for 1000ms followed by a 200ms blank screen. Each condition was repeated 6 times (total of 54 blocks) presented in a counterbalanced order.

4.3.4 Imaging parameters

Data from both fMRI experiments were collected using a GE 3 Tesla HD Excite MRI scanner at the York Neuroimaging Centre at the University of York. A T1-weighted structural MRI (1 x 1 x 1mm voxel) and a gradient-echo EPI were acquired for each participant. Functional data was collected using a gradient-echo EPI sequence with a radio-frequency coil tuned to 127.4 MHz was used to acquire 38 axial slices (TR = 3s, TE = 33ms, flip angle = 90°, FOV = 260mm, matrix size = 128 x 128, slice thickness = 3mm, voxel size: 2.25 x 2.25 x 3mm).

4.3.5 fMRI analysis

First, we examined the magnitude of brain response to each stimulus condition in voxels within the ROI. Statistical analysis of the fMRI data was carried out using FEAT version 5.0 in the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). The first 3 volumes (9 s) of each scan were removed to minimize the effects of magnetic saturation, and slice-timing correction was applied. Motion correction was followed by spatial smoothing (Gaussian, full width at half maximum 6 mm) and temporal high-pass filtering (cut-off, 0.01 Hz). Separately for each run, parameter estimates for each condition in the general linear model (GLM) were generated by regressing the hemodynamic response of each voxel against a box-car

regressor convolved with a single-gamma hemodynamic response function. In Experiment 1, we additionally performed a fixed-effects analysis to compute the average parameter estimate for each condition across the two runs.

Analyses were restricted to face-selective regions FFA, OFA and STS using masks from a previous study (Sormaz, Watson, Smith, Young, & Andrews, 2016). These masks were derived from data from a series of localiser scans using blocks of faces varying in identity and viewpoint as well as scrambled faces. Face-selective regions of interest (ROIs) were defined at the group level using the contrast faces > scrambled faces. First, the peak voxels for the OFA, FFA and STS in each hemisphere were localised. Then the 500 voxels with the highest z-scores within each region were selected and finally masks were combined across hemispheres generating 3 masks of equal size (1000 voxels) for the OFA, FFA and posterior STS (see Haxby et al., 2000).

Next, we measured patterns of response to each stimulus condition within these face-selective regions. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Poldrack, Halchenko, & Hanson, 2009; Shinkareva et al., 2008). First, parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. These data were then submitted to correlation-based pattern analyses (Haxby, Connolly, & Guntupalli, 2014; Haxby et al., 2001) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the mean response per voxel across all experimental conditions (see Haxby et al., 2001). For each iteration of the LOPO cross-validation, the normalised patterns of response to each stimulus condition were

correlated between the group and the left-out participant. The final correlation matrix provides a measure of the similarity in the pattern of response across different combinations of conditions. These neural correlations were transformed using Fisher's Z-transform and then converted into z scores.

4.3.5.1 Regression analyses

We next used a representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) utilising multiple regression to assess the relative contributions of identity, viewpoint, and expression to the neural response patterns. For each aspect (identity, viewpoint, expression) a binary regressor was generated representing a model correlations matrix whereby ones were placed on those elements where the relevant factor was shared and zeroes on all other elements. The regressors were then entered into the multiple regression analysis of the LOPO data. This analysis yielded a beta value and associated standard error for each regressor which would be expected to differ significantly from zero if that regressor were able to explain a significant amount of the variance in the MVPA correlations. A t-contrast was used to assess the significance of the differences between the regression coefficients for each model. This was performed separately for Experiment 1 comparing identity with viewpoint and then Experiment 2 comparing expression and viewpoint.

4.3.5.2 Analysis of lower-level image properties

To investigate the effects of lower-level image properties on patterns of neural response in face-selective regions, the image statistics of each object were computed using the GIST descriptor (http://people.csail.mit.edu/torralba/code/spatialenvelope/) (Oliva & Torralba, 2001). For each image, a vector of 2048 values was obtained by passing the image through a series of 32 Gabor filters (eight orientations at four spatial frequencies),

and windowing the filtered images along a 8x8 grid or 64 spatial locations (Fig. 4.3). Each vector represents the image in terms of the output of each gabor filter at each position across the image.
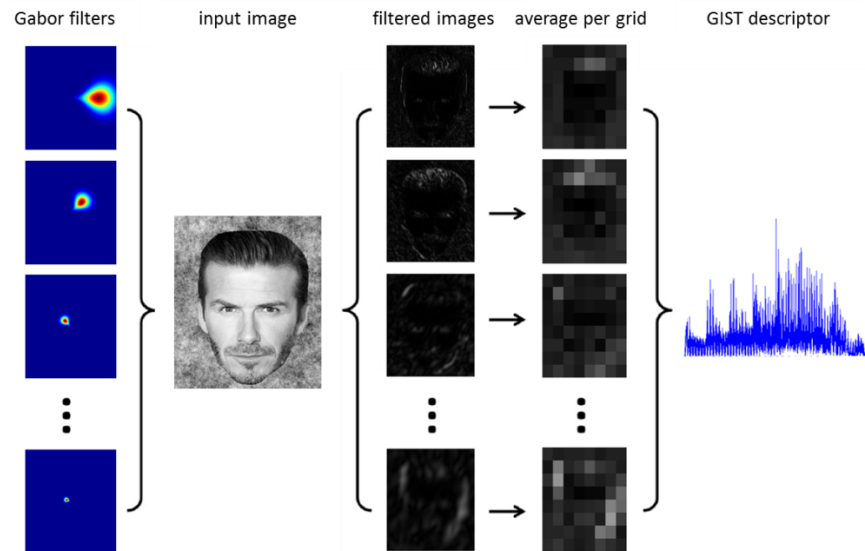
Figure 4.3 Schematic illustration of the calculation of a GIST descriptor for an example image. A series of Gabor filters across eight orientations and four spatial frequencies are applied to the image. Each of the resulting 32 filtered images is then windowed along a 8x8 grid to give a final GIST descriptor of 2048 values (right).

A Leave-One-Image-Out (LOIO) cross-validation procedure was used to assess the similarity of images in different conditions. GIST descriptors were averaged across all but one of the images within each condition. These average descriptors were then compared with each unique image creating within- and between-category correlations for each combination of condition. The similarity matrix of the correlation values for the GIST descriptor across all pairwise combinations of conditions was then used as a regressor in a multiple regression analysis to predict the fMRI pattern of response within face-selective regions. This analysis yielded whether beta value and associated standard error differed significantly from zero and thus being able to explain a significant amount of the variance in the MVPA correlations.

4.4 Results

4.4.1 Experiment 1: Identity versus viewpoint processing

First, we measured the mean % signal change of each ROI to each identity at each viewpoint. Fig. 4.4 shows similar levels of response to all conditions within each region. Repeated measures ANOVA showed that there was no effect of identity in any face-selective region (OFA: $F(2,36) = 0.37$, $p = .694$; FFA: $F(2,36) = 0.01$, $p = .995$; STS: $F(2,36) = 0.01$, $p = .994$). There was also no effect of viewpoint in the FFA ($F(2,36) = 1.69$, $p = .198$). However, there was an effect of viewpoint in the OFA ($F(2,36) = 7.62$, $p < .01$) and the STS ($F(2,36) = 8.98$, $p < .01$). All regions showed significant interactions. The interactions were driven by higher responses to the left compared to the right view of BP (OFA: $F(4,72) = 4.83$, $p < .01$; FFA: $F(4,72) = 5.40$, $p < .01$; STS: $F(4,72) = 4.22$, $p < .01$).
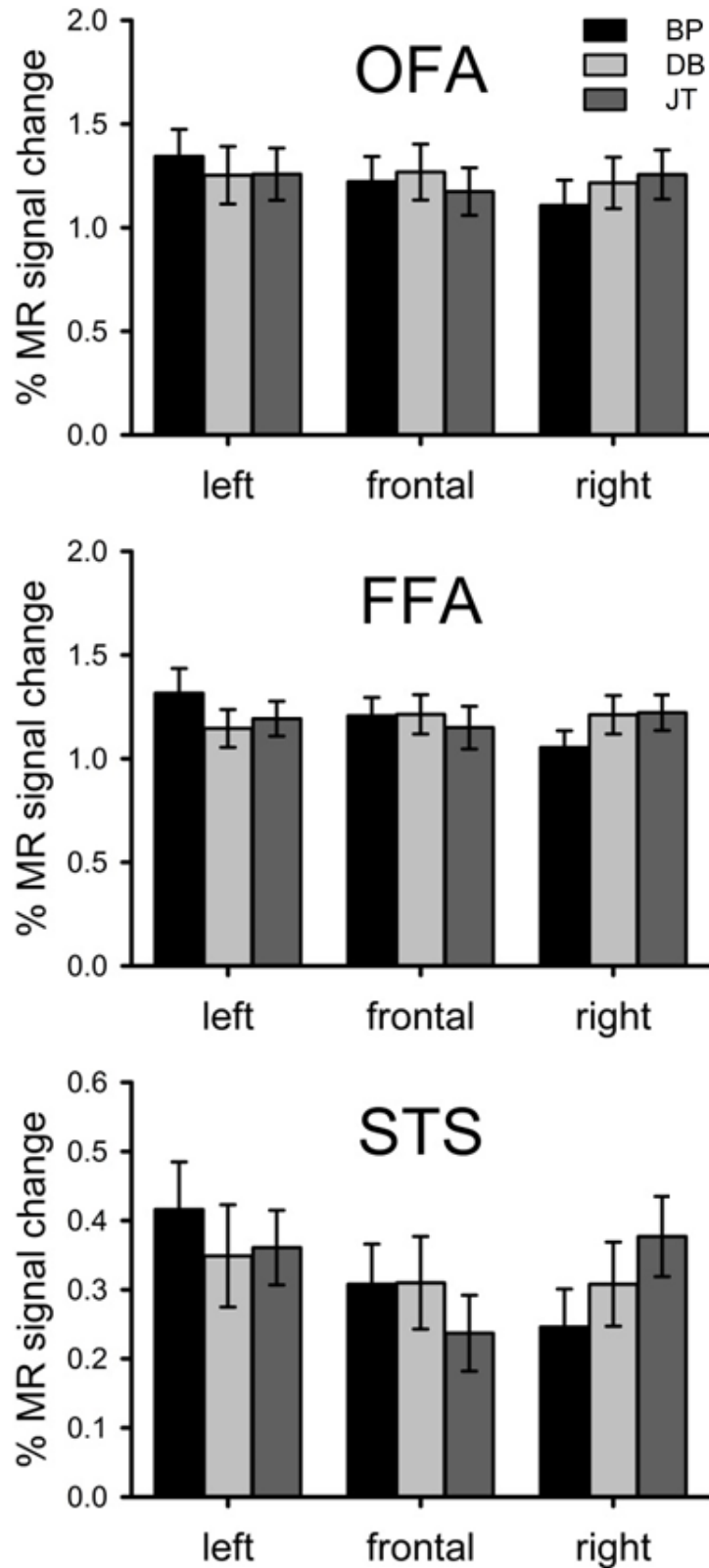
Figure 4.4 *Brain response within face-selective regions to each condition: Left, frontal, and right view of*

Brad Pitt (BP), David Beckham (DB), and Justin Timberlake (JT).

Next, we examined the pattern of brain response. For each ROI, we measured patterns of brain response for each of our conditions and then conducted a correlation based MVPA (Haxby et al., 2001) to measure their similarity (Fig. 4.5b). To test the effect of identity and viewpoint on the brain responses, we used a representational similarity analysis (Kriegeskorte et al., 2008). Model correlation matrices were generated where the patterns of brain response are entirely predicted by identity or by viewpoint (Fig. 4.5a). These were then used as regressors in a multiple regression analysis of the fMRI data (Fig. 4.5c).

Identity did not predict patterns of response in any face-selective region (OFA: β = 0.02, *p = .209*; FFA: β = 0.00, *p = .875*; and STS: β = 0.02, *p = .169*). However, viewpoint predicted patterns of response in all face-selective regions (OFA: β = 0.34, *p < .001*; FFA: β = 0.18, *p < .001*; and STS: β = 0.21, *p < .001*). Indeed, viewpoint was a significantly better predictor of brain response in all face-selective regions (OFA: t = 13.97, *p < .001*; FFA: t = 8.00, *p < .001*; and STS: t = 9.36, *p < .001*). Thus, viewpoint explained patterns of responses in all face-selective regions.

Experiment 1

Figure 4.5 *Regression analysis of the fMRI response patterns within face-selective regions.*
(a) Models used for the regression analysis of the fMRI response patterns. Binary models were defined representing the cases where the patterns of response are entirely predicted by either identity or viewpoint. Conditions labels code identity (BP = Brad Pitt, DB = David Beckham, and JT = Justin Timberlake) and viewpoint (left, frontal, or right), e.g. BP_L corresponds to Brad Pitt shown from the left. (b) Similarity of fMRI response patterns within face-selective regions measured using MVPA correlation for all conditions. These matrices were compared against binary regressor of identity and viewpoint using a multiple regression analysis. (c) The resulting beta coefficients for each regressor. Error bars represent 1 SE (*** p < .001).

4.4.2 Experiment 2: Expression versus viewpoint processing

Next, we compared the effects of two changeable aspects, expression and viewpoint. First, we measured the mean % signal change of each ROI to each expression at each viewpoint. Figure 4.6 shows similar levels of response to all conditions within each region. Repeated measures ANOVAs showed that there was no effect of expression in any face-selective region (OFA: $F(2,46) = 0.45$, $p = .639$; FFA: $F(2,46) = 0.36$, $p = .700$; STS: $F(2,46) = 0.22$, $p = .806$). However, there was an effect of viewpoint in all face-selective regions (OFA: $F(2,46) = 12.31$, $p < .001$; FFA$(2,46) = 8.03$, $p < .01$; STS: $F(2,46) = 8.04$, $p < .01$). These effects were driven by slightly higher responses to the right view in the OFA and FFA, and lower responses to the frontal view in the STS. Finally, none of the face-selective regions showed any significant interactions (OFA: $F(4,92) = 1.19$, $p = .322$; FFA: $F(4,92) = 0.15$, $p = .961$; STS: $F(4,92) = 1.76$, $p = 0.14$).
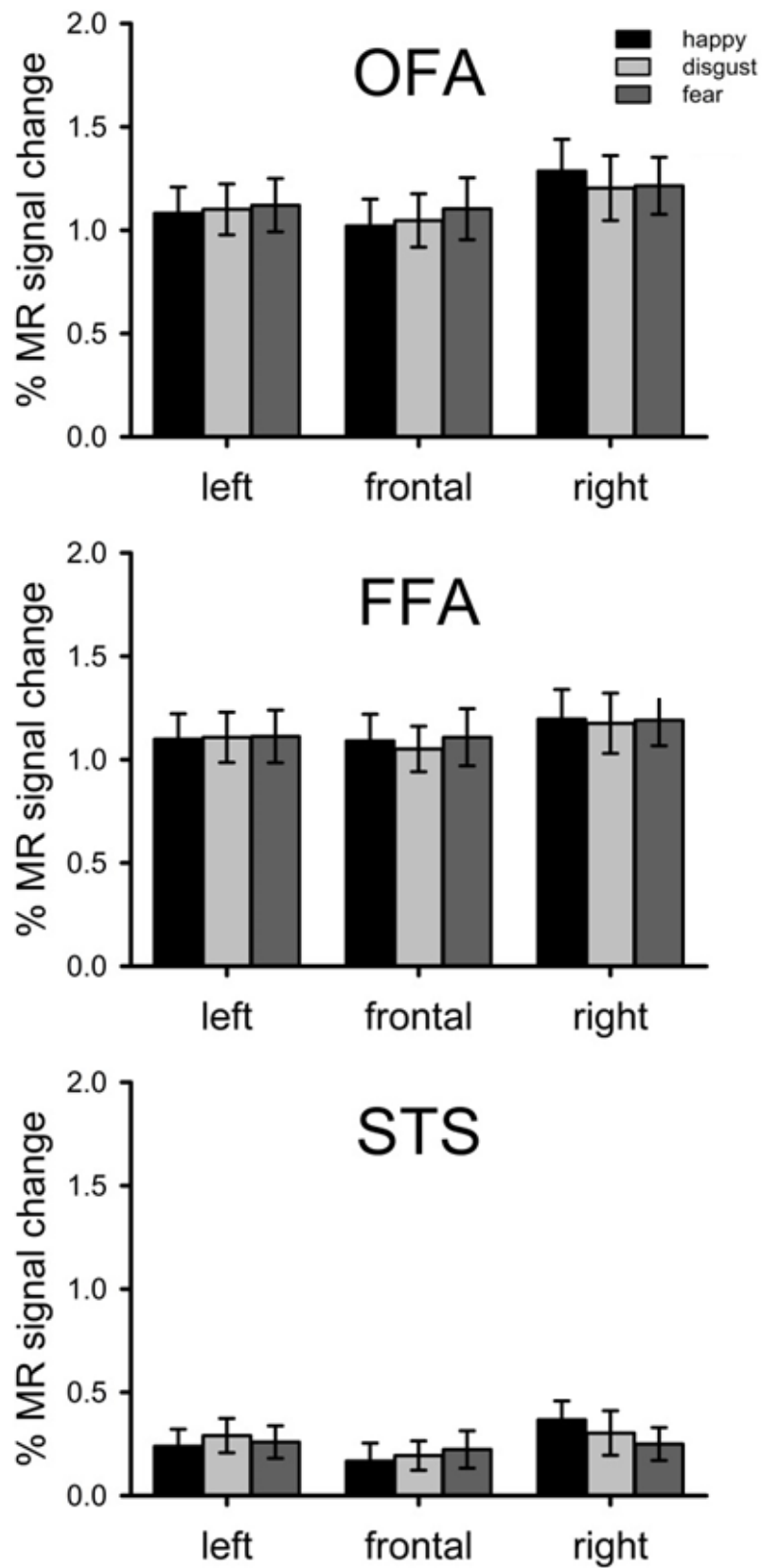
Figure 4.6 *Brain response within face-selective regions to each condition: Left, frontal, and right view of*

happy, disgust, and fear.

Next, we examined the pattern of brain response. For each ROI, we measured patterns of

brain response for each of our conditions and then conducted a correlation based MVPA

to measure their similarity (Fig. 4.7b). To test the effect of expression and viewpoint on

the brain responses, we used a representational similarity analysis. Model correlation

matrices were generated where the patterns of brain response are entirely predicted by

identity or by viewpoint (Fig. 4.7a). These were then used as regressors in a multiple

regression analysis of the fMRI data (Fig. 4.7c).

Expression did not predict patterns of response in any face-selective region (OFA:

β = 0.01, *p = .327*; FFA: β = 0.02, *p = .149*; and STS: β = 0.01, *p = .499*). However, just as in

Experiment 1 viewpoint explained brain responses in all face-selective regions

significantly better than expression (OFA: β = 0.31, p < .001 , t = 14.94, *p < .001*; FFA:

β = 0.14, *p < .001* , t = 5.69, *p < .001*; STS: β = 0.31, *p < .001* , t = 15.31, *p < .001*). Thus,

viewpoint explained patterns of responses in all face-selective regions in both
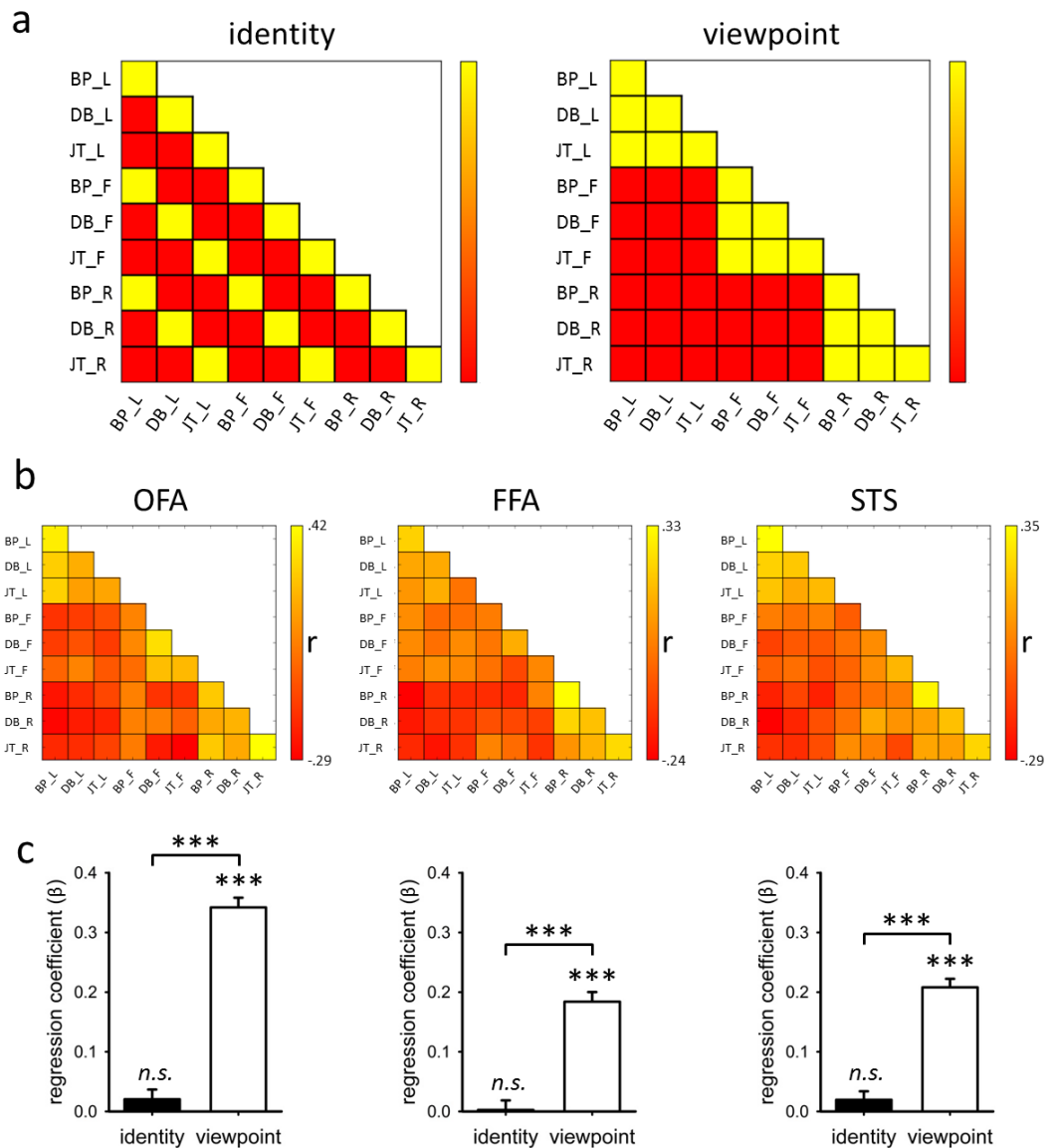
experiments.

Experiment 2
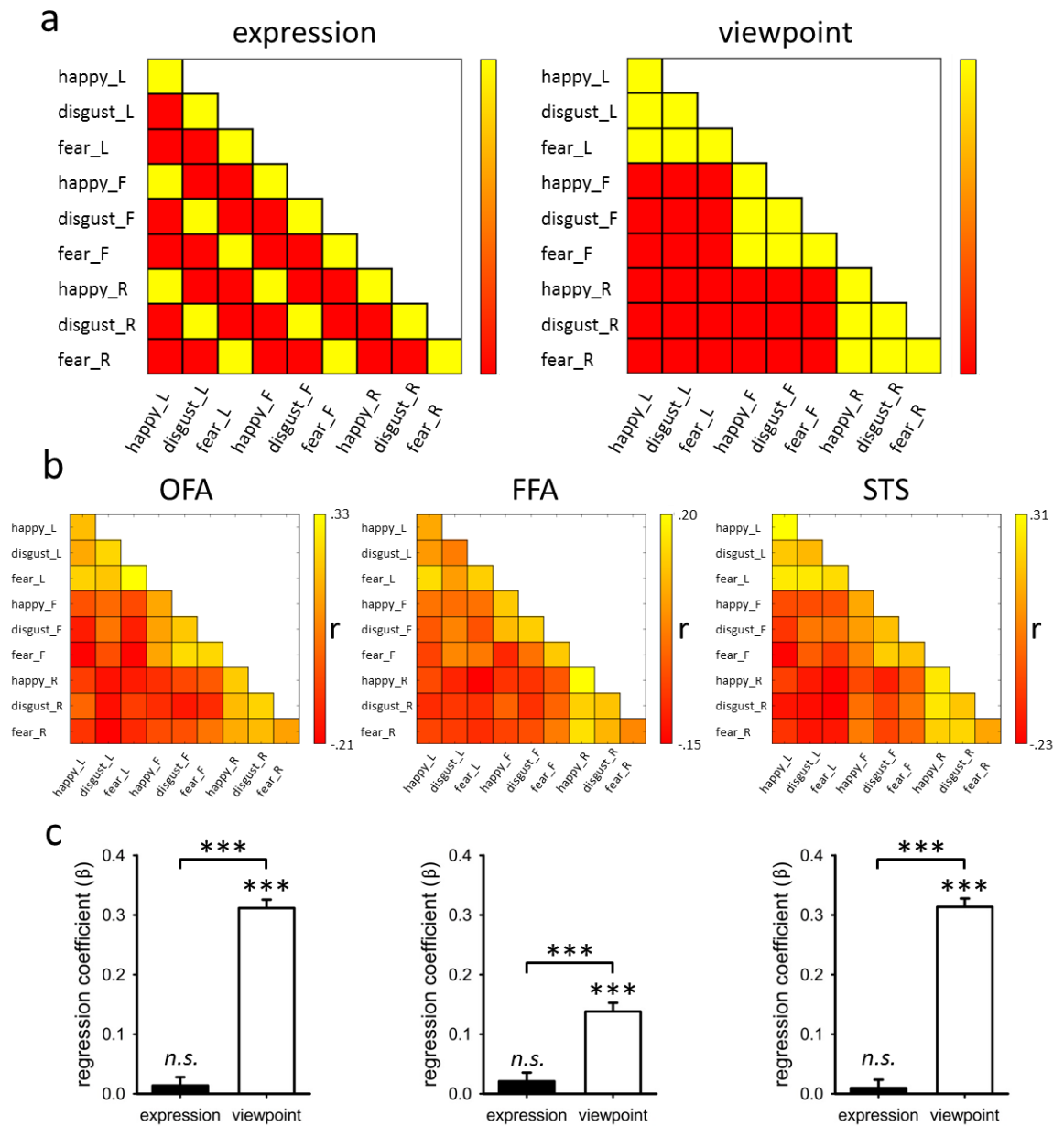
Figure 4.7 *Regression analysis of the fMRI response patterns within face-selective regions.*

(a) Models used for the regression analysis of the fMRI response patterns. Binary models were defined

representing the cases where the patterns of response are entirely predicted by either expression or

viewpoint. (b) Similarity of fMRI response patterns within face-selective regions measured using MVPA

correlation for all conditions. These matrices were compared against binary regressor of expression and

viewpoint using a multiple regression analysis. (c) The resulting beta coefficients for each regressor. Error

bars represent 1 SE (*** p < .001).

4.4.3 Processing of lower-level image properties

Next, we asked whether the brain responses within face-selective regions could be explained by the lower-level properties of the image. To address this question, we measured lower-level image properties of each image using the GIST descriptor (Oliva & Torralba, 2001). We compared the images within and between conditions by correlating their GIST using a Leave-One-Image-Out (LOIO) analysis separately for each experiment. This yielded two average GIST similarity matrices (Fig. 4.8a and Fig. 4.8b). Next, we used this GIST similarity matrix as a regressor to predict brain response similarity within face-selective regions. GIST predicted brain response similarity within all face-selective regions significantly in Experiment 1 (OFA: β = 1.60, $p < .001$; FFA: β = 0.95, $p < .001$; STS: β = 1.09, $p < .001$) as well as in Experiment 2 (OFA: β = 1.20, $p < .001$; FFA: β = 0.55, $p < .001$; β = 1.23, $p < .001$) (Fig. 4.8). Indeed the observed effects were around five times as high as the previously found viewpoint effects.

Experiment 1
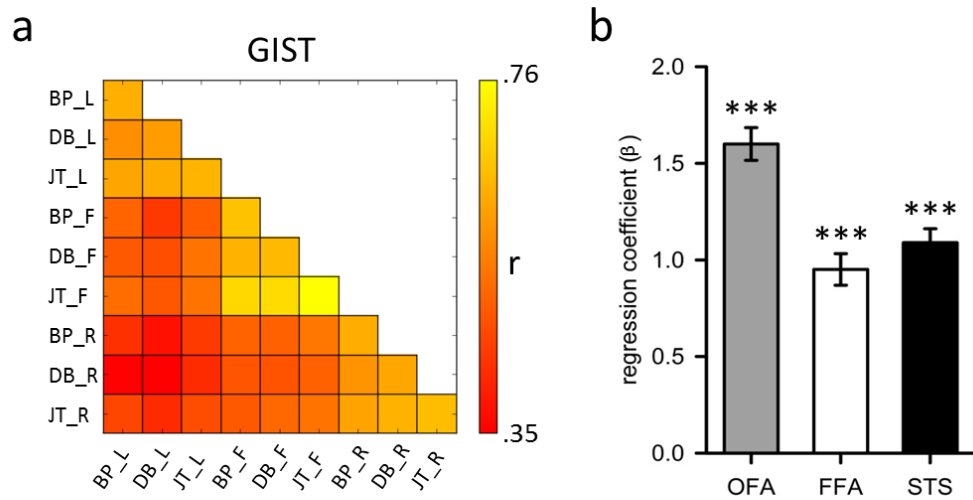
Experiment 2
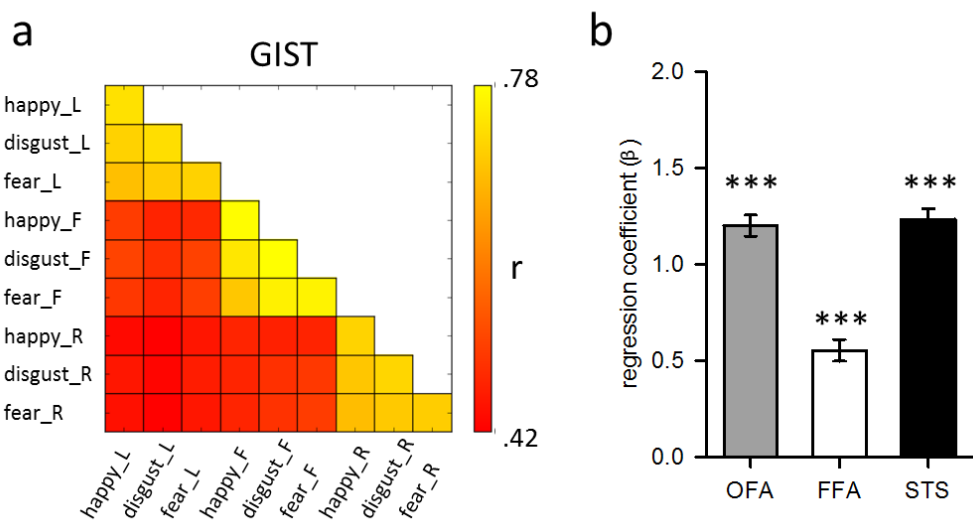
Figure 4.8 *Regression analysis of the lower-level image properties. a) Similarity of GIST measured using LOIO correlation for all conditions. b) Results of regression analysis using GIST similarity matrix to predict brain response in each face-selective region. The resulting beta coefficients with error bars represent 1 SEM (*** p < .001).*

4.5 Discussion

The first aim of this study was to investigate the relative role higher level properties of faces such as identity, viewpoint, and expression on brain response patterns in face-selective regions. In Experiment 1, participants viewed images from different identities at different views. Identity had no significant effect on patterns of response in face-selective regions. However, viewpoint significantly predicted response patterns in all face-selective regions. In Experiment 2, we showed images of faces with different expressions from different views. Again, only viewpoint significantly predicted response patterns in all face-selective regions. Next, we asked whether response within face-selective regions could be explained by lower-level image properties. We found that indeed lower-level image properties predicted response patterns in all face-selective regions.

Previous studies have reported distinct patterns of neural response to different identities independent of changes to the image, e.g. different views (Axelrod & Yovel, 2015; Guntupalli et al., 2016; Natu et al., 2010; Nestor et al., 2011). We did not find such an image-invariant representation of identity in any face-selective region. This could not be explained by difficulties to recognise the identities used, since all participants could name all our famous identities and successively recognised them across the different images used. A potential explanation could be our use of different images within each condition. Previous identity decoding studies measured brain response to the exactly same image (Anzellotti et al., 2013) or highly similar images, e.g. only varying slightly in size or position on screen (Guntupalli et al., 2016). We deliberately measured brain response to each condition across different images. This way we tried to ensure we are capturing brain response specific to an identity rather than to specific images.

Viewpoint significantly predicted response patterns in all face-selective regions. This finding was consistent across different analyses and is in line with studies reporting viewpoint representation across different identities within face-selective regions (Axelrod & Yovel, 2012; Guntupalli et al., 2016; Kietzmann et al., 2012). Consistent, with previous findings (Axelrod & Yovel, 2012; Guntupalli et al., 2016), the largest viewpoint effect was observed in the OFA, which is thought to be involved in processing the visual properties of face images. This suggests that rather than an effect of viewpoint, this finding might reflect visual image properties.

Recent studies have shown that lower-level image properties can be used to predict patterns of response in category-selective regions (Andrews et al., 2015; Rice et al., 2014; Watson et al., 2016). These findings suggest that response patterns within face-selective regions may reflect lower-level image properties that typically co-occur with different facial cues such as identity, viewpoint, or expression rather than these higher level properties per se. Hence, we measured lower-level image properties of our conditions and used these to explain response within face-selective regions. We found a significant effect of lower-level image properties on response patterns of all face-selective regions.

The link between the patterns of fMRI response to faces and their image properties has important implications for understanding how the ventral visual cortex is organized. The strong, linear relationship between low-level image properties and the spatial patterns of neural response suggests that more basic principles could underpin the topographic organization of face regions. These findings are consistent with previous studies that used PCA to show that neural responses to different object categories in

82

inferior temporal cortex can be predicted by variance in the principal components of the images (Baldassi et al., 2013; O'Toole, Jiang, Abdi, & Haxby, 2005).

The patterns of fMRI response generalized across participants. Using a modified cross-validation analysis (Haxby et al., 2001), we compared the pattern of response in one participant with the pattern from a group analysis in which that participant was left out (Shinkareva et al., 2008; Poldrack et al., 2009; Haxby et al., 2001). This LOPO approach showed that the topographic patterns of response to different object categories were consistent across individuals. These observations are significant in that they suggest that our findings reflect the operation of consistent, large-scale topographical organizing principles, rather than an arbitrarily distributed representation in each individual.

These results should not be taken to imply that higher level properties are not represented in face regions. Rather, these results show that the coarse, topographic organization is not dependent on these properties. It is entirely possible that higher level properties such as identity and expression could be represented at a finer spatial scale than can be detected by fMRI. For example, in early visual areas topographic maps of the visual field are clearly evident across the cortical surface (Engel, Glover, & Wandell, 1997). However, at a finer scale of cortical organization, topographic maps have also been found for orientation. For example, change in the preferred stimulus orientation of neurons has been shown to occur gradually across the cortical surface (Bonhoeffer & Grinvald, 1991; Hubel & Wiesel, 1968).

In conclusion, these results show that patterns of response in face regions are dominated by changes in the viewpoint of the face rather than changes in identity or expression. However, patterns of response are most clearly predicted by lower-level properties of the image.

Table 4.1 Mean image statistics (SD) of images in each condition.

A 3 x 3 repeated measured ANOVA with identity (BP = Brad Pitt, DB = David Beckham, JT = Justin Timberlake) and viewpoint (left, frontal, right) revealed that the mean grey value intensity and RMS contrast did not differ between conditions (grey value intensity: identity: $F_{(2,36)} = 0.07$, $p > .05$; viewpoint: $F_{(2,36)} = 0.09$, $p > .05$; identity x viewpoint: $F_{(4,36)} = 0.23$, $p > .05$; RMS contrast: identity: $F_{(2,36)} = 2.10$, $p > .05$; viewpoint: $F_{(2,36)} = 0.02$, $p > .05$; identity x viewpoint: $F_{(4,36)} = 0.88$, $p > .05$).

		left	frontal	right
Grey value intensity	BP	147.17 (6.96)	149.98 (7.09)	148.00 (5.15)
	DB	148.61 (11.90)	146.73 (5.53)	148.03 (11.09)
	JT	150.63 (9.07)	148.83 (3.68)	146.99 (3.24)
RMS contrast	BP	54.60 (4.97)	49.58 (4.04)	52.75 (2.89)
	DB	48.53 (4.65)	50.97 (4.54)	49.73 (4.64)
	JT	52.08 (9.28)	55.11 (5.47)	53.90 (5.47)

Chapter 5 – The neural representation of surface and shape in face-selective regions

This chapter is adapted from: Weibert, Kramer, Ritchie, Burton & Andrews (2016). The neural representation of surface and shape in face-selective regions. *(in preparation)*[3]

5.1 Abstract

A full understanding of face recognition will involve identifying the visual information that is used to discriminate different identities and how this is represented in the brain. Previous behavioural studies have shown that the recognition and matching of familiar faces is primarily based on their surface properties. The aim of this study was to explore the importance of surface and shape properties in the neural representation of familiar faces. We took a set of face images and measured variance in the surface and shape properties of the face. A principal component analysis (PCA) was then used to represent this variability in surface and shape properties. Next, we measured patterns of neural response using fMRI. We showed participants a subset of the images and measured the patterns of brain response. We then determined whether surface or shape properties could predict patterns of response using a representational similarity analysis. In contrast to the prediction from behavioural studies, patterns of response in the FFA could only be predicted by the shape of the face image. The dissociation between the neural and behavioural findings suggests that, although the FFA may play a role in the recognition of facial identity, this region is not solely responsible for this process.

---

[3] The author, Katja Weibert, designed the experiment, analysed the results, and wrote the article under the supervision of Prof. Timothy Andrews.

5.2 Introduction

Neuroimaging studies have revealed a network of regions in the occipital and temporal lobe which form a core system for the visual analysis of faces (Haxby et al., 2000; Kanwisher et al., 1997). These studies have consistently found regions that show stronger responses to faces than other visual stimuli in the fusiform gyrus (the fusiform face area, or FFA), occipital cortex (the occipital face area, or OFA) and the posterior superior temporal sulcus (pSTS). These three functionally localisable regions form a core system for the visual analysis of faces in the widely used neural model of Haxby and colleagues (2000). Here, we are interested in developing a more detailed analysis of the information that is represented in face regions, and in particular whether it corresponds for the information that is critical to familiar face recognition.

A distinction between *shape* and *surface* properties is widely used in face perception research (Bruce & Young, 1998, 2012). Any facial image consists of a set of edges created by abrupt changes in reflectance due to the shapes and positions of facial features and a broader pattern of reflectance based on the surface properties of the face – also known as texture or albedo (Bruce & Young, 1998, 2012). Surface properties result from the pattern of reflectance of light due to the combination of ambient illumination, the face's pigmentation, and shape from shading cues. Shape properties arise from the 3D geometrical description of a face, and how that is projected onto a 2D image.

A number of studies have shown that the surface properties of faces play a critical role in the invariant representation that is used for the recognition of familiar faces (Burton, Jenkins, Hancock, & White, 2005; Hole, George, Eaves, & Rasek, 2002; Russell, Biederman, Nederhouser, & Sinha, 2007; Russell & Sinha, 2007). For example, familiar

face recognition is not substantially affected if the surface properties are presented on a standardized shape (Burton et al., 2005), or when a face's shape is distorted by stretching the image (Hole et al., 2002). In contrast, line drawings of faces, which lack any surface properties, are not usually sufficient for recognition (Davies, Ellis, & Shepherd, 1978; Leder, 1999). The reason shape information may not be a reliable cue for the recognition of familiar face identity is thought to be that shape cues (particularly from the internal features of the face) vary across different images of the same face (Burton, 2013). Together, these studies suggest that surface properties of the face are the dominant cue in the recognition of familiar faces.

The aim of this study was to investigate the relative importance of shape and surface properties in the neural representation of familiar faces. Within the core system of face-selective regions, the fusiform face area (FFA) is thought to be particularly important for the representation of invariant facial characteristics that are important for face recognition (Haxby et al., 2000; Grill-Spector et al., 2004; Rotshtein et al., 2004). To address whether patterns of response in brain regions such as the FFA are more sensitive to surface or shape properties, we measured the variance in a set of face images. The variance in the surface and shape properties was then analysed using PCA. This allowed us to determine the similarity of two face images in terms of either surface or shape properties. Next, we asked whether the similarity in either surface or shape properties could predict patterns of response in the FFA. Based on behavioural studies, our prediction was that patterns of response should be predicted by surface rather than shape.

5.3 Methods

5.3.1 Subjects

Twenty-three right-handed subjects with normal or corrected to normal vision participated in the experiment (9 males; mean age: 23.57 ± 0.86 years). All subjects gave their written informed consent. The study was approved by the York Neuroimaging Centre Ethics Committee.

5.3.2 Stimuli

Different face images of Leonardo DiCaprio were used. Using a famous identity ensured participants' recognition (all participants could name Leonardo DiCaprio from an array of 10 celebrities). Since a PCA analysis requires a large number of images using a famous identity also ensured that many different images were readily available. First, face images were gathered using the same technique as in Burton, Kramer, Ritchie, and Jenkins (2016). "Leonardo DiCaprio" was used in Google Images as a search term, along with criteria (full-color, large, face images only). Then the first 50 images meeting the following criteria were chosen: (a) no part of the face obscured (e.g., by clothing, glasses, or a hand); (b) pose very broadly full-face in order to allow the placement of landmarks for the PCA; and (c) pose standing or sitting, but not lying down, to limit head angle to relatively upright. Finally, the images were converted to greyscale (Fig. 5.1).

Figure 5.1 *Stimuli for PCA.*

5.3.3 Image analysis using PCA

PCA was performed as in Burton et al. (2016) using 50 different images of Leonardo

DiCaprio. Each image was rescaled to 190 pixels wide x 285 pixels high, and converted to

greyscale. Face shape was determined using anatomical landmarks. A standard grid

comprised 82 xy-coordinates was used, resulting in a vector of 164 numbers (82 points x 2

coordinates; Fig. 5.2a). Shape PCA was based on these shape vectors (Fig. 5.2b). Next, the

surface PCA was computed. To do so, first the average shape was generated by

computing the mean coordinates for each landmark of all 50 images. The surface for each

image was then morphed to the average shape, yielding a vector of 54,150 numbers (190

width x 285 height x 1 grey value intensity). Surface PCA was based on these surface

vectors (Fig. 5.2b).
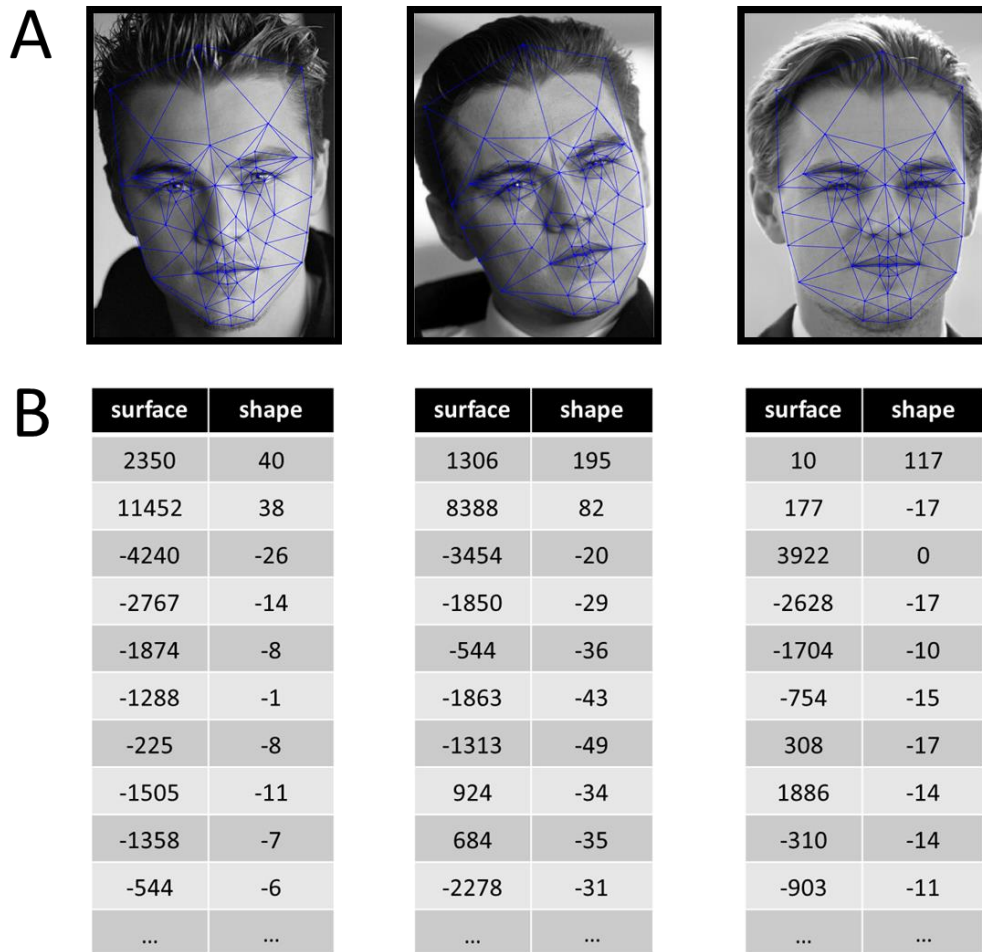
Figure 5.2 *Example images with anatomical landmarks and corresponding PCA vectors. A) Example images, superimposed in blue are the 82 defined anatomical landmarks. B) Corresponding PCA-values of surface and shape. Images were described along 49 dimensions in surface and shape space using PCA (values of first 10 dimensions shown). To compare images the PCA vectors were correlated. This was done separately for surface and shape information. Left and middle images have similar surface information (r = .93) and similar shape information (r = .87). Left and right images have similar shape information (r = .59, p = .072), but different surface information (r = -.07). Middle and right images have similar shape information (r = .86), but different surface information (r = -.03).*

5.3.4 fMRI Experiment

To investigate the neural representation of faces in face-selective regions, 10 images of Leonardo DiCaprio were used. These were chosen from the original images maintaining

the original variation in surface as well as shape. Each of the 10 stimuli was presented in a block resulting in 10 conditions. A red-spot task was used to maintain attention during the scan, where participants responded with a button press whenever they saw the red spot (one target per block). Each stimulus block lasted for 3s and contained 3 identical images. Within each stimulus block each image was presented for 800ms followed by a 200ms blank screen. Blocks were separated by a 6s fixation screen (a white fixation cross on a black background). The experiment consisted of two experimental runs, each with 6 repetitions of each of the 10 stimuli conditions (total of 60 blocks per run) presented in a counterbalanced order.

5.3.4.1 Imaging parameters

Data from the fMRI experiment were collected using a GE 3 Tesla HD Excite MRI scanner at the York Neuroimaging Centre at the University of York. A T1-weighted structural MRI (1x1x 1mm voxel) and a gradient-echo EPI were acquired for each subject. Functional data was collected using a gradient-echo EPI sequence with a radio-frequency coil tuned to 127.4 MHz was used to acquire 38 axial slices (TR = 3s, TE = 33ms, flip angle = 90°, FOV = 260mm, matrix size = 128 x 128, slice thickness = 3mm, voxel size: 2.25 x 2.25 x 3mm).

5.3.4.2 fMRI analysis

Statistical analysis of the fMRI data was carried out using FEAT version 5.0 in the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). The first 3 volumes (9 s) of each scan were removed to minimize the effects of magnetic saturation, and slice-timing correction was applied. Motion correction was followed by spatial smoothing (Gaussian, full width at half maximum 6 mm) and temporal high-pass filtering (cut-off, 0.01 Hz). Separately for each run, parameter estimates for each condition in the general linear model (GLM) were generated by regressing the hemodynamic response of each voxel against a box-car

regressor convolved with a single-gamma hemodynamic response function. Then, a fixed-effects analysis was used to compute the average parameter estimate for each condition across the two runs.

Analyses were restricted to face-selective regions FFA, OFA and STS using masks from a previous study (Sormaz, Watson, Smith, Young, & Andrews, 2016). Face-selective regions of interest (ROIs) were defined at the group level using the contrast faces > scrambled faces. First, the peak voxels for the OFA, FFA and STS in each hemisphere were localised. Then the 500 voxels with the highest z-scores within each region were selected and finally masks were combined across hemispheres generating 3 masks of equal size (1000 voxels) for the OFA, FFA and posterior STS forming the core face-selective regions in Haxby et al.'s (2000) neural model.

Next, we measured patterns of response to each stimulus condition within these face-selective regions. Parameter estimates were generated for each condition. The reliability of response patterns was tested using a leave-one-participant-out (LOPO) cross-validation paradigm (Poldrack et al.,2009; Shinkareva et al., 2008). First parameter estimates were determined using a group analysis of all participants except one. This generated parameter estimates for each condition in each voxel. This LOPO process was repeated such that every participant was left out of a group analysis once. These data were then submitted to correlation-based pattern analyses (Haxby et al., 2014; Haxby et al., 2001) implemented using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates were normalised by subtracting the mean response per voxel across all experimental conditions (see Haxby et al., 2001). For each iteration of the LOPO cross-validation, the normalised patterns of response to each stimulus condition were correlated between the group and the left-out participant. The final correlation

matrix provides a measure of the similarity in the pattern of response across different combinations of conditions. These neural correlations were transformed using Fisher's Z-transform and then converted into z-scores.

5.3.4.3 Regression analyses: Effect of PCA information on brain response

We next used a representational similarity analysis (RSA; Kriegeskorte et al., 2008) utilising multiple regression to assess the relative contributions of surface and shape PCA-information to the neural response patterns. For each factor (surface and shape) a regressor was generated by correlating the PCA vector of two images. This was repeated until all images were compared yielding two correlation matrices, one for surface and one for shape, representing the PCA image similarities. These regressors were then used to predict measured brain pattern similarities separately for the OFA, the FFA, and the STS, using multiple regression. This analysis yielded a beta value and associated standard error for each regressor which would be expected to differ significantly from zero if that regressor were able to explain a significant amount of the variance in the MVPA correlations. A t-contrast was used to assess the significance of the differences between the betas.

5.3.5 Analysis of lower-level image properties measured by GIST

To investigate general effects (unspecific to faces) of low-level image properties on patterns of neural response in face-selective regions, the image statistics of each object were computed using the GIST descriptor (http://people.csail.mit.edu/torralba/code/spatialenvelope/) (Oliva, & Torralba, 2001). For each image, a vector of 2048 values was obtained by passing the image through a series of Gabor filters across eight orientations and four spatial frequencies, and windowing the filtered images along a 8x8 grid (Fig. 5.3).
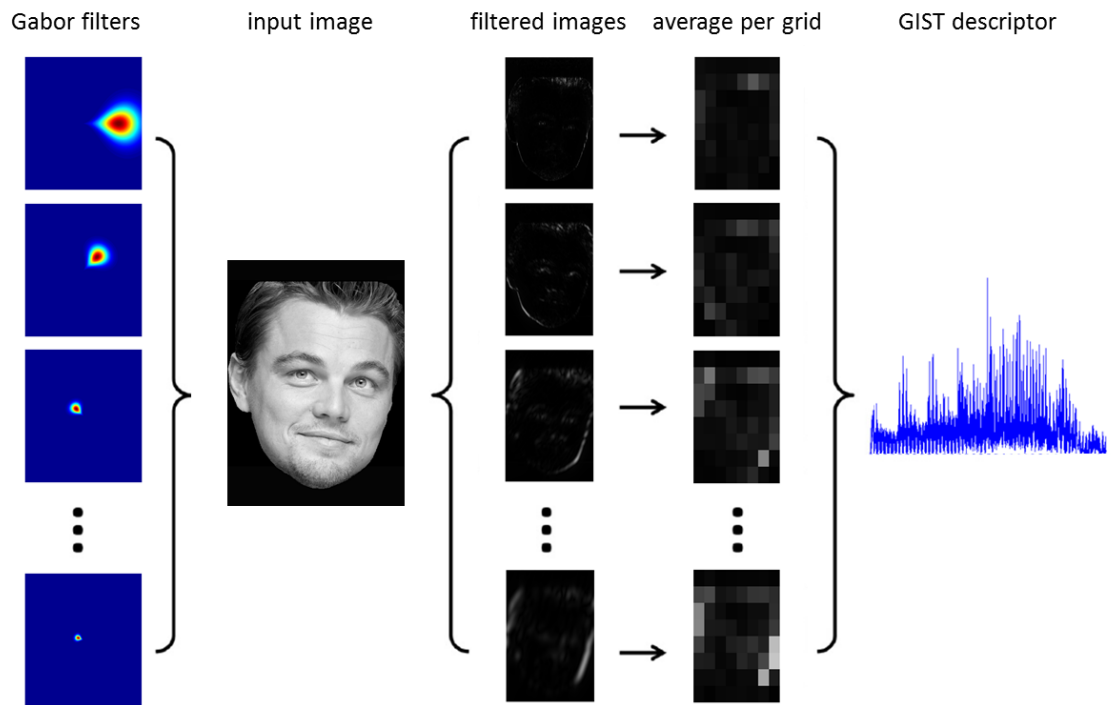
Figure 5.3 *Schematic illustration of the calculation of a GIST descriptor for an example image. A series of Gabor filters across eight orientations and four spatial frequencies are applied to the image. Each of the resulting 32 filtered images is then windowed along a 8x8 grid to give a final GIST descriptor of 2048 values (right).*

Each vector represents the image in terms of the spatial frequencies and orientations present at different positions across the image. To assess how similar individual images were, the GIST descriptors of two images were correlated. This was performed for each possible image pair resulting in a similarity matrix based on lower-level image properties. This similarity matrix of the correlation values for the GIST descriptor across all pairwise combinations of conditions was then used as a regressor in a multiple regression analysis to predict the fMRI pattern of response within face-selective regions. This analysis yielded whether beta value and associated standard error differed significantly from zero and thus being able to explain a significant amount of the variance in the MVPA correlations.

5.4 Results

First, we measured the mean % signal change of each ROI to each face image (Fig. 5.4A). Fig. 5.4B shows similar levels of response to all face images within each region. A repeated measures ANOVA showed that there was an effect of region ($F(2,44) = 40.82$, $p < .001$), with lower response in the STS. There was also a small effect of face image (Greenhouse-Geisser corrected: $F(4.85, 106.71) = 2.55$, $p < .05$). This was due to a slightly higher reponse to face image 6 compared to image 3 and 8. However, there was no interaction between region and face image (Greenhouse-Geisser corrected: $F(5.76, 126.60) = 1.69$, $p = .132$).
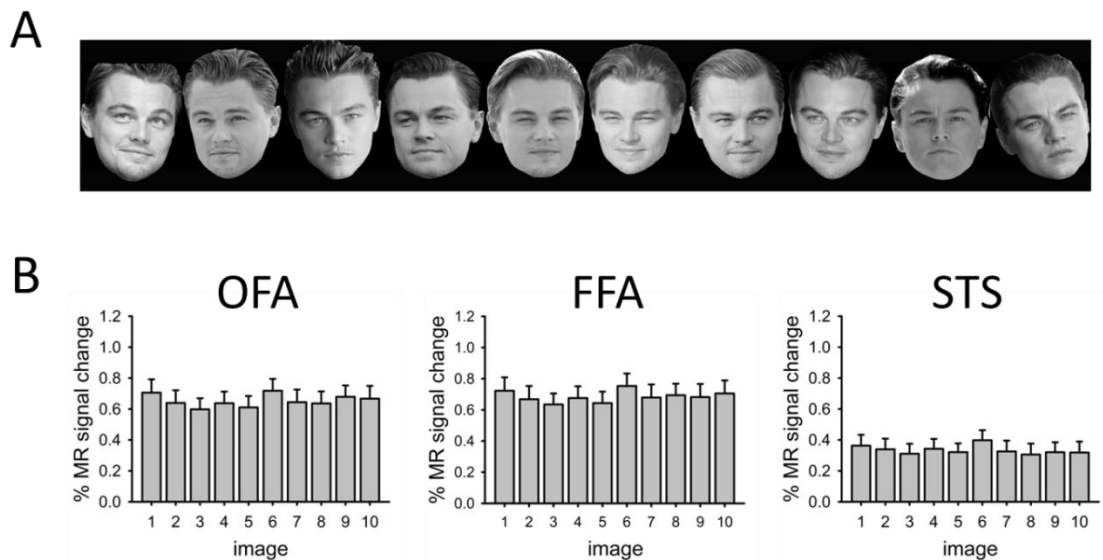
Figure 5.4 *Brain response within face-selective regions to each face image. A) Face images used in the fMRI experiment. B) Brain response within face-selective regions to each face image. All face images elicited comparable brain responses in all face-selective regions.*

Next, we examined the pattern of brain response. For each ROI, we measured patterns of brain response for each of our face images and then conducted a correlation-based MVPA to measure their similarity (Fig. 5.5B). To test the effect of PCA information (surface and shape) on the brain responses, we used a representational similarity analysis. Model

correlation matrices were computed by correlating PCA information of two images (separately for surface and shape) until all images were compared. Two model correlation matrices resulted, one describing PCA surface image similarities and one PCA shape image similarities (Fig. 5.5A). These were then used as regressors in a multiple regression analysis of the fMRI data (Fig. 5.5C).
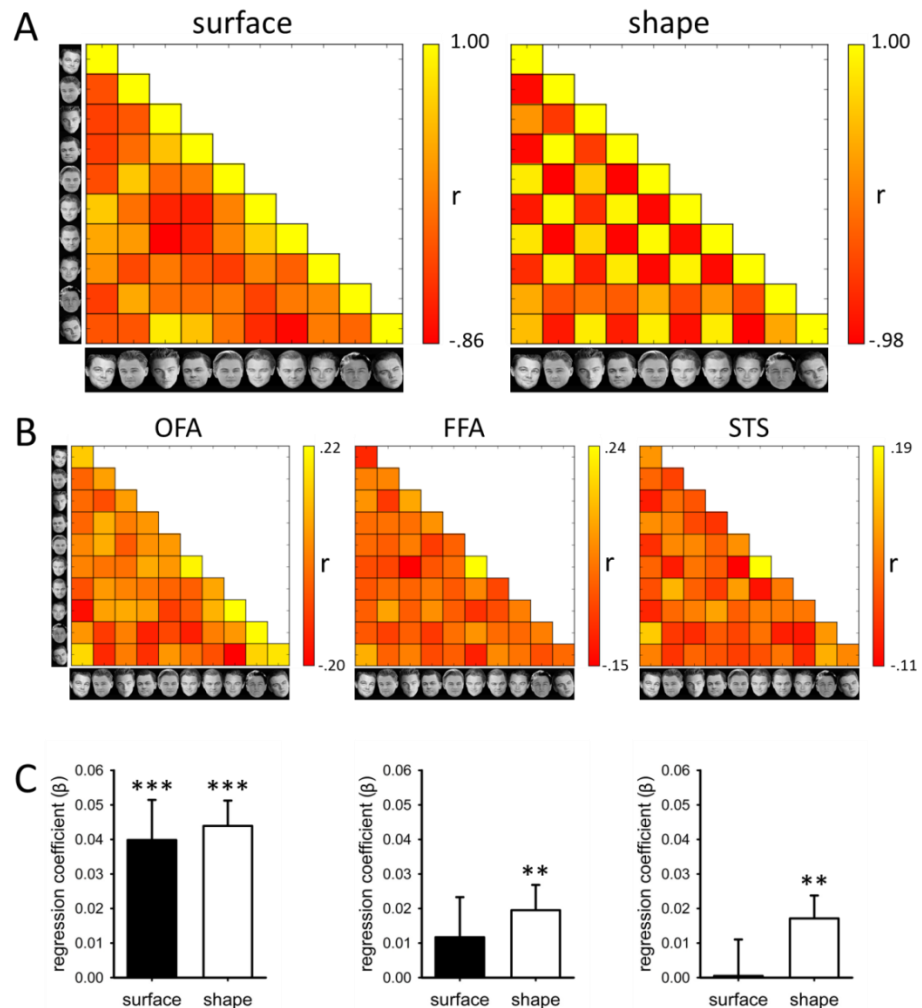
Figure 5.5 *Regression analysis of the fMRI response patterns within face-selective regions. A) Models used for the regression analysis of the fMRI response patterns. To assess image similarity, PCA vectors of two images were correlated. This was performed for all pairwise image comparisons separately for surface and shape. B) Similarity of fMRI response patterns within face-selective regions measured using MVPA correlation for all face images. These matrices were compared against the regressors of surface and shape using a multiple regression analysis. C) The resulting beta coefficients for each regressor. Error bars represent 1 SEM (*** p < .001, ** p < .01).*

Surface predicted patterns of response only in the early face processing OFA (β = 0.04, $p < .001$), but not in the FFA (β = 0.01, $p = .314$) and not in the STS (β = 0.00, $p = .981$). Shape predicted patterns of response in all face-selective regions (OFA: β = 0.04, $p < .001$; FFA: β = 0.02, $p < .01$; and STS: β = 0.01, $p < .01$). Although shape was a better predictor (higher β-value) of brain response in all face-selective regions, this difference did not reach significance (OFA: t = 0.28, $p = .781$; FFA: t = 0.53, $p = .594$; and STS: t = 1.27, $p = .204$).

Next we asked whether the observed effects were specific to the face-measures (surface and shape) or reflect general lower-level image properties. To address this question, we used another measure of image properties, GIST, which captures image properties of scenes (Oliva & Torralba, 2001). Analogue to the PCA surface and shape analysis, we computed a model correlation matrix this time using the GIST information of two images, which was used as regressors in a multiple regression analysis of the fMRI data (Fig. 5.6A). GIST could not predict patterns of response in any face-selective region (OFA: β = -0.01, $p = .970$; FFA: β = -0.01, $p = .823$; STS: β = 0.01, $p = .791$) (Fig. 5.6B). Overall our results suggest patterns of response in face-selective regions are predicted by face shape.
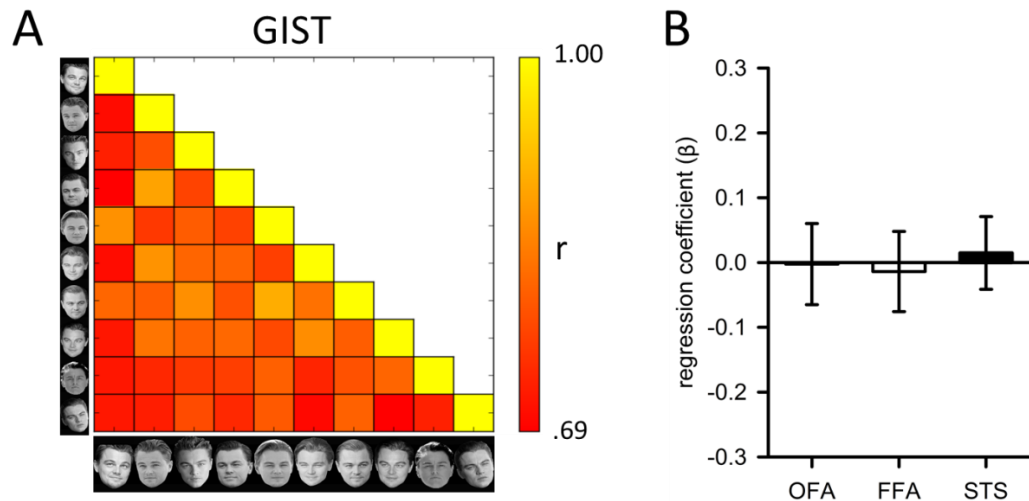
Figure 5.6 *Regression analysis of the non-face image properties using GIST. A) Similarity of all face pairings using GIST and correlation. B) Results of regression analysis using GIST similarity matrix to predict brain response in each face-selective region. The resulting beta coefficients with error bars represent 1 SE.*

5.5 Discussion

The aim of this study was to investigate how the surface and shape properties of face images are represented in the brain. Variance in the surface and shape properties of a large set of images were described using PCA. We then measured the response to these faces in FFA. The similarity in the patterns of response in the FFA could be predicted by shape information, but not by surface properties.

Thinking of facial images as broadly consisting of surface (pigmentation, shading patterns) and shape (feature positions) properties has helped our understanding of facial identity recognition, where it is clear that although surface and shape cues can contribute (Russell, Sinha, Biederman, & Nederhouser, 2006; Troje, & Bülthoff, 1996), the role of surface cues becomes more salient for familiar faces (Burton et al., 2005 ; Russell & Sinha, 2007). A useful way of testing the importance of surface- and edge-based cues in face perception is with contrast reversal (as in a photo negative). In a contrast-reversed image

the edges that define feature shapes remain in the same positions, despite the huge change in overall surface properties. Recognition of facial identity is severely disrupted by contrast-reversal, showing the importance of surface patterns to the recognition of facial identity (Bruce, & Langton, 1994; Burton et al., 2005; Russell et al., 2006). Further support for the importance of surface properties comes from studies showing that line drawings of faces are not usually sufficient for the accurate recognition or discrimination of identity (Davies et al., 1978; Leder, 1999).

Our fMRI experiment built on these behavioural results by using the logic that a neural region that is directly responsible for the recognition of facial identity should show a corresponding differential sensitivity to surface over shape information. This was achieved with a paradigm in which surface and shape properties were used to predict patterns of neural response in the face regions. In contrast to our hypothesis, we found that the neural responses in the FFA could be predicted by shape but not by surface properties of the faces. The difference between previous behavioural findings using familiar faces and the pattern of neural responses in FFA implies that the spatial pattern of response is not directly linked to the recognition of identity.

How do these findings sit with the idea that the FFA is involved in processing invariant characteristics of faces (Haxby et al., 2000; Gobbini, & Haxby, 2007)? Identity is obviously central to these invariant characteristics, making it tempting to equate the FFA with an area primarily responsible for face recognition. However, previous neuroimaging studies have reported mixed results about whether the FFA has an image-invariant representation of identity. Some studies have reported image-invariance (Loffler et al., 2005; Rotshtein et al., 2004; Eger et al., 2005; Ewbank, & Andrews, 2008; Davies-

Thompson et al., 2013), whereas others have reported image dependence (Grill-Spector et al. 1999; Andrews, & Ewbank, 2004; Pourtois et al., 2005b; Davies-Thompson et al., 2009; Xu et al., 2009). More generally, these results are consistent with previous studies that have shown patterns of response in face regions are sensitive to the image properties (Yue, Tjan, & Biederman, 2006; Xu al., 2009; Rice et al., 2014; Watson et al., 2016). For example, patterns of response in the fusiform gyrus to faces can be predicted by their image properties (Rice et al., 2014).

Lower-level image properties have been shown to predict patterns of response in category-selective regions (Andrews et al., 2015; Rice, et al., 2014; Watson et al., 2016). The observed effects in face-selective regions could therefore reflect processing of general lower-level image properties rather than face-specific image information. To investigate this, we used the GIST-descriptor, a measure of lower-level image properties developed for scenes (Oliva & Torralba, 2001). In contrast to chapter 4, we found no effect of GIST on patterns of response in any face-selective region. The difference between these findings could be due to the stimuli used. In this chapter, all stimuli depicted the same identity and thus much less variation in low-level image properties. Our findings suggest, that only a face-specific measure of lower-level image properties, e.g. via PCA, allows this finer discrimination between different images of the same identity.

Models of face processing suggest that information about faces is initially processed in the OFA (Kanwisher et al., 1997; Haxby et al., 2000). The OFA is thought to be involved in the visual analysis of the face image. Consistently, we found that patterns of response in the OFA could be predicted by both surface and shape properties of the face, supporting the suggested role of the OFA in early face processing.

Processing of changeable aspects of faces such as gaze and expression facial expressions is thought to occur in the posterior STS (Haxby et al., 2000). Previous studies have shown that judgements of changeable aspects of faces such as expression rely on shape information (Bruce & Young, 1998; Etcoff & Magee, 1992; McKelviet, 1973). Consistent with these findings, our results show that the patterns of response in the STS were sensitive to the shape of the face.

The perception of facial identity is proposed to involve the FFA (Haxby et al., 2000). Behavioural studies have shown that surface information plays the primary role in identity processing (Bruce, & Langton, 1994; Burton et al., 2005; Davies et al., 1978; Leder, 1999; Russell et al., 2006). This suggests that the FFA should represent the surface information of faces. However, contrary to these predictions, we found that variation in surface information of faces did not predict patterns of response in the FFA. In fact, patterns of response in the FFA reflected shape information.

In conclusion, the aim of this experiment was to determine the relative contribution of surface and shape properties to the neural representation of familiar faces. Consistent with behavioural findings and the neural model of face processing, we found that surface and shape properties predicted patterns of response in the early face processing OFA. Consistent with the models of face processing and behavioural findings, only shape properties influenced the pattern of response in the STS. Although surface properties have been shown to be important in identity processing, the response in the FFA did not represent surface properties. This implies that the spatial pattern of response in the FFA is not directly linked to the recognition of identity.

Chapter 6 – An image-invariant neural response to familiar faces in the human medial temporal lobe

This chapter is adapted from: Weibert, Harris, Mitchell, Byrne, Young, & Andrews (2016). An image-invariant neural response to familiar faces in the human medial temporal lobe. (*submitted*)[4]

6.1 Abstract

The ability to recognize familiar faces with ease across different viewing conditions contrasts with the inherent difficulty in the perception of unfamiliar faces across similar image manipulations. Models of face processing suggest that this difference is based on the neural representation for familiar faces being more invariant to changes in the image, than it is for unfamiliar faces. Here, we used an fMR-adaptation paradigm to investigate neural correlates of image-invariant face recognition in face-selective regions of the human brain. Participants viewed faces presented in a blocked design. Each block contained different images of the same identity or different images from different identities. Faces in each block were either familiar or unfamiliar to the participants. First, we defined face-selective regions by comparing the response to faces with the response to scenes and scrambled faces. Next, we asked whether any of these face-selective regions showed image-invariant adaptation to the identity of a face. The core face-selective regions showed image-invariant adaptation to familiar and unfamiliar faces. However, there was no difference in the adaptation to familiar compared to unfamiliar faces. In contrast, image-invariant adaptation for familiar faces, but not for unfamiliar

faces, was found in face-selective regions of the medial temporal lobe (MTL). Taken together, our results suggest that the marked differences in the perception of familiar and unfamiliar faces may depend critically on neural processes in the medial temporal lobe.

6.2 Introduction

As members of a highly social primate species, our everyday lives depend critically on being able to recognise people we know, so that we can interact with them appropriately based on our knowledge of their characteristics and personal histories. Recognising the faces of familiar individuals is often central to this process, and this has led to a great deal of interest in the neural underpinnings of face recognition.

The distinction between often seen familiar faces and unfamiliar faces that have not been previously encountered is central to understanding face recognition. While photographs of unfamiliar faces can be remembered and later recognised remarkably well, recognition performance with unfamiliar faces breaks down as soon as any changes are made between studied and test images (Bruce, 1982; Longmore, Liu & Young, 2008). Remarkably, the same problems arise in perceiving unfamiliar faces, where the perceptual matching of unfamiliar faces is severely hampered by image changes (Hancock et al., 2000; Kemp, Towell, & Pike, 1997). In striking contrast, the behavioural hallmark of familiar face recognition is that it is remarkably successful across substantial changes in expression, viewing angle, and lighting conditions (Bruce, 1994; Bruce & Young, 2012; Burton, 2013).

These findings have been incorporated into cognitive models of face processing which propose that familiar faces are represented differently from unfamiliar faces (Bruce

& Young, 1986; Burton et al., 1999). These models propose that all faces are initially encoded in an image-dependent representation, which is sufficient to recognise identical images of faces. However, our ability to recognise familiar faces across changes relies on representations that are relatively invariant to changes in the image, which are often referred to as face recognition units (FRU) in models of face processing (Bruce & Young, 1986). These FRUs interact with person identity nodes (PINs), which are involved in the retrieval of names, and other semantic information associated with the face (Bruce & Young, 1986).

In terms of how faces are represented in the brain, many studies have followed Kanwisher and colleagues' (1997) procedure of localising face-selective regions by contrasting neural responses to faces and other visual stimuli. These have revealed a network of posterior regions now usually designated the occipital face area (OFA), the fusiform face area (FFA) and the posterior superior temporal sulcus (STS) which form a core system for the visual analysis of faces in the widely used neural model proposed by Haxby and colleagues (2000). Within this core system of face-selective regions, the FFA is thought to be particularly important to the representation of invariant facial characteristics necessary for face recognition (Grill-Spector et al., 2004; Haxby et al., 2000; Ishai, 2008). The Haxby et al. model explicitly acknowledges that other brain regions will contribute to the recognition of faces. One of these regions is the anterior temporal lobe which is suggested to represent biographical semantic knowledge associated with a face, e.g. the name of the person (Collins & Olson, 2014).

Although functional localisers can be used to identify face-selective brain regions, this in itself gives only limited information about what such regions do. A powerful

complementary method for understanding the functional properties of a region is fMR-adaptation, as it offers insight into the underlying neural mechanisms (Grill-Spector, Henson, & Martin, 2006). Consistent with Haxby et al.'s (2000) idea of FFA involvement in processing invariant aspects of faces (such as identity), fMRI studies have shown a reduced response (adaptation) to repeated images of the same face in the FFA (Andrews & Ewbank, 2004; Grill-Spector et al., 1999; Harris, Rice, Young, & Andrews, 2015; Loffler et al., 2005; Rotshtein et al., 2004; Yovel & Kanwisher, 2005). Such findings imply that the identity of the face is represented at some level in the FFA and it is being adapted by repeated presentations. However, given our discussion of the behavioural evidence, a much stronger test for a link between neural activity and the recognition of facial identity is needed to determine whether this adaptation is still evident when different images of the same identity are shown (i.e. image-invariant adaptation). It turns out that fMR-adaptation studies that have used different images of the same identity have shown mixed results. Some studies show a complete absence of adaptation to different images in the FFA (Andrews & Ewbank, 2004; Grill-Spector et al., 1999; Pourtois, Schwartz, Seghier, Lazeyras, & Vuilleumier, 2005a; Xu, Yue, Lescroart, Biederman, & Kim, 2009), whereas other studies show continued adaptation (Loffler et al., 2005; Winston, Henson, Fine-Goulden, & Dolan, 2004).

A limitation of previous studies using fMR-adaptation to probe the neural correlates of face recognition is that they often fail to provide a direct comparison of familiar and unfamiliar faces (Johnston & Edmonds, 2009; Natu & O'Toole, 2011). This is a key limitation since, as we have noted, cognitive models only propose an image-invariant representation for familiar and not for unfamiliar faces (Bruce & Young, 1986; Burton, et al., 1999). However, studies that have compared familiar and unfamiliar faces also report

mixed results. Some studies have found image-invariant identity adaptation in the FFA for familiar but not unfamiliar faces (Eger et al., 2005; Ewbank & Andrews, 2008), whereas other studies fail to find any difference in adaptation to familiar and unfamiliar faces (Davies-Thompson et al., 2009; Pourtois et al., 2005b). Together, these findings show at best limited evidence that the marked behavioural differences in the perception of familiar and unfamiliar faces are linked to differences in the way faces are represented in core face-selective regions of the human brain. One problem with identifying image-invariant responses to faces is that the representation of identity may involve a sparse code (Quiroga et al., 2005) involving only a limited number of neurons and thus lie beyond detection in conventional fMRI designs. To address this issue, we performed an fMR-adaptation experiment with a large sample of participants (N = 80). Our aim was to use the combination of the sensitive adaptation method and the statistical power of a large participant sample to reveal regions in either the core or extended face processing network that show an image-invariant response to familiar faces.

6.3 Methods

6.3.1 Participants

Eighty right-handed participants with normal or corrected to normal vision participated in the experiment (45 females; mean age: 23.8 ± 4.2 years). All participants gave their written informed consent. The study was approved by the York Neuroimaging Centre Ethics Committee.

6.3.2 fMRI Experiment

Face stimuli were taken from previous studies (Davies-Thompson et al., 2013; Weibert & Andrews, 2015) and included male and female identities. All images showed frontal views and neutral facial expressions, but varied in other aspects of appearance such as hairstyle and viewpoint. Our aim was to capture naturally occurring variation in images. Faces were either familiar or unfamiliar. Our familiar face stimuli depicted famous identities. An advantage of using famous faces was that it allowed us to use the same stimuli for each participant. This contrasts with the use of personally familiar faces in which different stimuli are used for each participant. To ensure familiarity with our familiar faces, prior to taking part in the experiment, participants were tested on their recognition of the familiar faces. Participants could name over 85% of the identities used (M = 85.5%, SD = 10.0%). Crucially, they could name all (100%) of the identities from the familiar same condition. This way we ensured, that they could recognise when the same famous identity was repeated across different images. We also measured low-level image properties of the familiar and unfamiliar faces. The mean change in image intensity was calculated by taking the average of the absolute differences in grey value at each pixel for successive

pairs of images within a block. Table 6.1 shows that the mean intensity change was similar across all face conditions (F(3,108) = 0.488, *p > .5*).

Table 6.1 Mean change in intensity (S.D.) between successive images in each condition of the adaptation scan.		
	Same Identity	Different Identity
Familiar	16.1 (7.6)	17.2 (8.0)
Unfamiliar	18.0 (8.7)	18.7 (8.7)

There were 6 stimulus conditions in a block design fMRI experiment:

    I.   familiar different
    II.   familiar same
   III.   unfamiliar different
   IV.   unfamiliar same
    V.   scenes
   VI.   scrambled faces

The different condition blocks included face images from different identities. The same condition blocks involved sequences of different images from the same identity (Fig. 6.1).
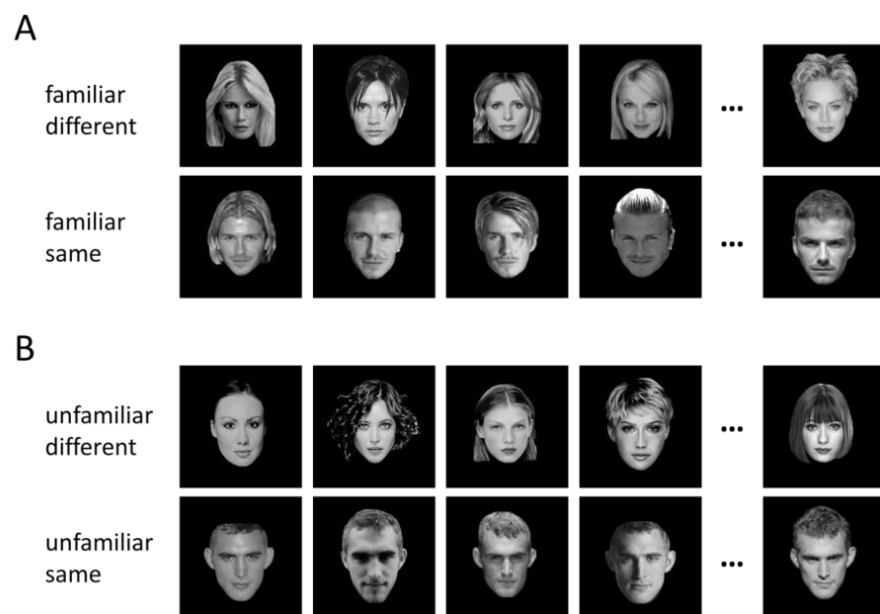
Figure 6.1   Examples of stimuli. During each face condition different face images were shown of either different identities or the same identity of (a) familiar faces and (b) unfamiliar faces.

Scene stimuli were taken from the LabelMe scene database (Oliva & Torralba, 2001). Scrambled face images were created by scrambling the phase of the face images.

All images were presented in greyscale. Stimuli from each condition were presented in a blocked design. Each stimulus block lasted for 9s and contained 8 images. Within each stimulus block each image was presented for 950ms followed by a 200ms blank screen. Each of the six stimulus conditions was repeated 4 times. This gave a total of 24 blocks, which were presented in a counterbalanced order. Blocks were separated by a 9s fixation screen (a white fixation cross on a black background). To maintain attention during the scan, participants performed a one-back task in which they responded with a button press every time an identical image was directly repeated (one target per block). This allowed us to use a common behavioural task across all conditions and minimised the influence of task on the neural response.

Data from the fMRI experiment were collected using a GE 3 Tesla HD Excite MRI scanner at the York Neuroimaging Centre at the University of York. A T1-weighted structural MRI (1 x 1 x 1mm voxel) and a gradient-echo EPI were acquired for each participant. Functional data was collected using a gradient-echo EPI sequence with a radio-frequency coil tuned to 127.4 MHz was used to acquire 38 axial slices (TR = 3s, TE = 33ms, flip angle = 90°, FOV = 260mm, matrix size = 128 x 128, slice thickness = 3mm, voxel size: 2.25 x 2.25 x 3mm).

Statistical analysis of the fMRI data was carried out using FEAT version 4.1 in the FSL toolbox (http://www.fmrib.ox.ac.uk/fsl). The first 3 volumes (9 s) of each scan were removed to minimize the effects of magnetic saturation, and slice-timing correction was applied. Motion correction was followed by spatial smoothing (Gaussian, full width at half maximum 6 mm) and temporal high-pass filtering (cut-off, 0.01 Hz). Regressors for each

condition in the general linear model (GLM) were convolved with a gamma hemodynamic response function. Individual participant data of all subjects were then aligned into MNI 152 space and combined using a higher-level mixed effects group analysis of the whole-brain to generate statistical maps across participants (FLAME, http://www.fmrib.ox.ac.uk/fsl).

Face-selective regions were defined by contrasting each face condition with scenes and scrambled faces and then computing the average of these contrasts. This average contrast was used as a mask for all further analyses. Next, we determined voxels that responded more to familiar than unfamiliar faces (familiar different & familiar same > unfamiliar different & unfamiliar same). Finally, we measured adaptation to familiar and unfamiliar faces. Adaptation was defined by the contrast different > same. This was done separately for familiar and unfamiliar faces. To correct for multiple comparisons in all analyses, the resulting statistical maps were thresholded at $z > 3.5$, corresponding to $p < 0.05$ (voxel-wise corrected).

Additionally, we performed a Region of Interest (ROI) analysis. ROIs were individually defined in each participant as in Weibert and Andrews (2015). First, an individual's statistical map was divided into individual face-selective clusters corresponding to core face-selective regions FFA, OFA, and STS by increasing the z-score threshold. Next, the peak voxel within each face-selective cluster was identified using the cluster function in FSL. Then, a flood-fill algorithm was used to grow a ROI from each peak voxel of a fixed size (50 voxel chosen based on Weibert & Andrews, 2015). This was accomplished by increasing the z-score threshold until only 50 voxels lie above. This resulted in ROIs with the same number of voxels across individuals. Only ROIs where all voxels were above the minimum threshold of $Z = 2.3$ were included in the analysis. ROIs

were merged across hemispheres except in case of the STS, where it was not possible to reliably identify a face-selective region in the lSTS, which was not included in further analysis.

6.4 Results

6.4.1 Face-selective regions

Using a whole-brain group analysis, we localized face-selective voxels across the 80 participants. Significant face-selectivity was evident in the core face-selective regions: fusiform face area (FFA), occipital face area (OFA), and posterior superior temporal sulcus (STS) in both hemispheres (Fig. 6.2A). We also found significant face-selective responses with the whole-brain group analysis in the right inferior frontal gyrus (rIFG) and in the medial temporal lobes (MTL) of both hemispheres (Fig. 6.2B). The MTL region overlapped with hippocampus and amygdala (http://www.cma.mgh.harvard.edu/fsl_atlas.html). The peak coordinates of these regions were also comparable to previous studies (Table 6.2; Davies-Thompson & Andrews, 2012; Kanwisher et al., 1997; Weibert & Andrews, 2015).
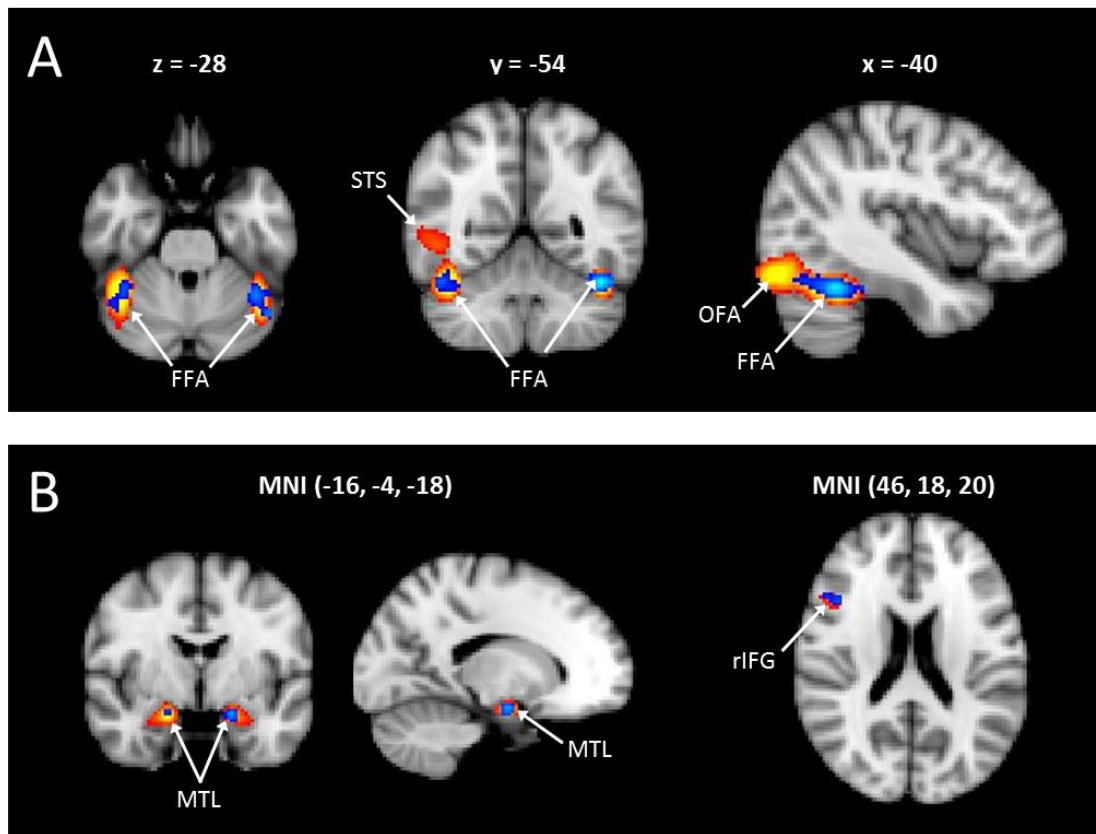
Figure 6.2   Face-selective regions (red) in the (A) core face-selective regions (FFA: fusiform face area, OFA: occipital face area, posterior STS: superior temporal sulcus) and (B) extended face-selective regions (MTL: medial temporal lobe, rIFG: right inferior frontal gyrus). Voxels showing selectivity for familiarity (familiar > unfamiliar) are shown in blue and were found in the FFA, rIFG and MTL. Images are shown in radiological convention reporting MNI coordinates.

6.4.2 Selectivity for familiarity

Next, we performed a voxel-wise analysis within the face-selective regions to locate voxels that were more selective to familiar compared to unfamiliar faces. We found that 19.5% of voxels within core face-selective regions responded more to familiar faces than to unfamiliar faces. These voxels were located in the FFA, but not in the OFA or STS (Fig. 6.2A, Table 6.2). This familiarity effect was more pronounced in the left hemisphere (LH: 57.1%, RH: 42.9%). There were no voxels that responded more to unfamiliar than familiar faces.

We then asked whether voxels in extended face-selective regions responded more to familiar than to unfamiliar faces. We found that 31.7% of voxels within extended face-selective regions responded more to familiar faces than to unfamiliar faces. These voxels overlapped with both the rIFG (30.8%) and MTL (69.2%) regions (Fig. 6.2B, Table 6.2). In the MTL, this familiarity effect was more pronounced in the left hemisphere (LH: 88.3%, RH: 11.7%). Again, there were no voxels that responded more to unfamiliar than familiar faces.

6.4.3 Selectivity for identity

Next, we investigated image-invariant adaptation to identity within core face-selective regions. This analysis was performed separately for familiar and unfamiliar faces by contrasting different identities > same identity. First, we investigated the core face-selective regions. There were no voxels in the core face-selective regions that showed significant adaptation to familiar or unfamiliar faces. There were also no voxels that showed a significantly higher response to the inverse contrast (same identity > different identities) (Table 6.2).

Table 6.2 Responses within face-selective clusters to familiarity and image-invariant identity adaptation separately for familiar and unfamiliar faces significant at $p < .05$ (voxel-corrected for multiple comparisons).

	size (mm³)	MNI coordinates			Familiarity (mm³)		Adaptation (mm³)	
		x	y	z	familiar > unfamiliar	unfamiliar > familiar	familiar different > familiar same	unfamiliar different > unfamiliar same
OFA	6328	44	-82	-18	-	-	-	-
	3952	-40	-86	-20	-	-	-	-
STS	7096	50	-62	4	-	-	-	-
	1488	-48	-74	4	-	-	-	-
FFA	7016	42	-52	-24	2544	-	-	-
	4536	-40	-56	-24	3384	-	-	-
IFG	688	48	20	18	456	-	-	-
MTL	2088	20	-6	-18	120	-	456	-
	1888	-18	-6	-18	904	-	320	-

It is possible that, although there were no significant differences in the independent voxels across the core face-selective regions, there was a difference in the mean response of a region. To address this possibility, we performed a ROI analysis in which the responses to each condition were averaged across voxels in each core face-selective ROI. The results of this analysis are shown in Figure 6.3. To analyse these results a 2x2 ANOVA was performed with familiarity and identity as factors. There was a significant effect of familiarity in the FFA ($F(1,73) = 33.17$, $p < .001$) and STS ($F(1,46) = 30.41$, $p < .001$) but not in the OFA ($F(1,57) = 0.43$, $p = .516$). This was due to a higher response to familiar faces compared to unfamiliar faces. There was a significant effect of identity in all regions (FFA: $F(1,73) = 35.89$, $p < .001$; OFA: $F(1,57) = 20.05$, $p < .001$; STS: $F(1,46) = 26.18$, $p < .001$). This was due to higher responses to different identity compared to same identity faces. Critically, however, none of the ROIs showed a significant interaction between familiarity and identity: FFA: $F(1,73) = 2.64$, $p = .109$; OFA: $F(1,57) = 0.81$, $p = .372$; STS: $F(1,46) = 2.76$, $p = .104$. The absence of any difference in adaptation to familiar and unfamiliar faces suggests that the core face regions do not explain the behavioural advantage for familiar face recognition.
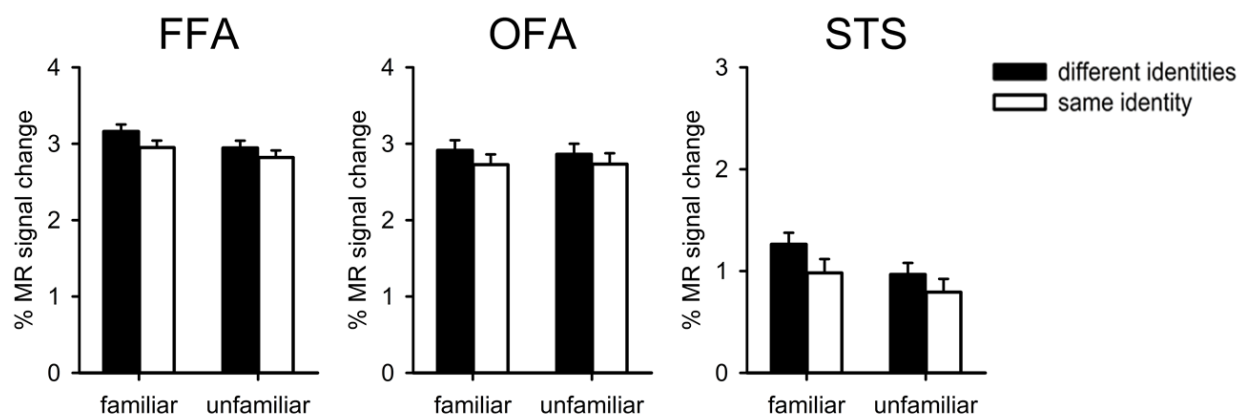

Figure 6.3 *ROI analysis of core face-selective regions. A significantly larger response to familiar compared to unfamiliar faces was evident in the FFA (N = 74) and STS (N = 47), but not in the OFA (N = 58). All regions showed adaptation to facial identity. However, there was no interaction between familiarity and identity.*

Next, we investigated image-invariant adaptation in the extended face-selective regions separately for familiar and unfamiliar faces. Within the extended face-selective regions, 16.6% of voxels showed an image-invariant adaptation to familiar faces. These voxels showed a smaller response (adaptation) to different images of the same familiar identity compared to images of different familiar identities (Table 6.2). Voxels showing image-invariant adaptation to familiar faces were found bilaterally in the MTL (Fig. 6.4A, Table 6.2). These voxels overlapped with the anatomical location of the hippocampus (59.8%) and the amygdala (40.2%) regions (Fig. 6.4B). No voxels showed image-invariant adaptation to unfamiliar faces. This analysis suggests that these voxels show image-invariant adaptation to familiar but not unfamiliar faces.
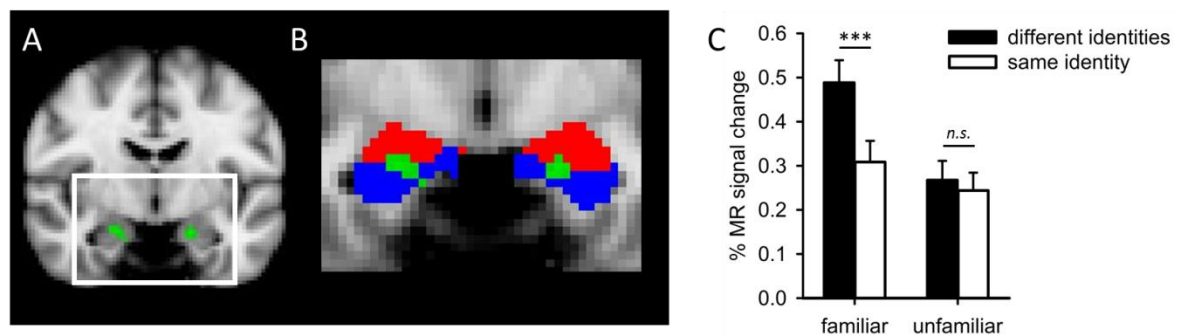
Figure 6.4 *Adaptation to familiar face identity in the medial temporal lobe (MTL). Images are shown in radiological convention. (A) Voxels within the MTL that showed adaptation to familiar faces. This adaptation was only evident for familiar faces, but not unfamiliar faces (familiar different identities > familiar same identity, depicted in green). B) Region of image-invariance to familiar faces superimposed onto Harvard Oxford structural masks (red: amygdala, blue: hippocampus). C) Mean % MR signal change within this region averaged across participants. Error bars depict ± 1SE.*

However, although adaptation was evident to familiar faces but not to unfamiliar faces, this does not necessarily show that adaptation was greater to familiar than unfamiliar faces. To directly compare image-invariant adaptation to familiar and unfamiliar faces, we performed a 2x2 repeated measures ANOVA with familiarity and

identity as factors. There was a significant effect of familiarity with higher response to familiar than unfamiliar faces (F(1,79) = 19.09, *p < .001).* There was also a significant effect of identity with higher response to different identity conditions than same identity conditions (F(1,79) = 8.76, *p < .01*). These effects were driven by a significant interaction between familiarity and identity (F(1,79) = 9.24, *p < .01*). Pairwise t-test revealed that there was no difference between the different identity and same identity conditions with unfamiliar faces (t(79) = 0.60, *p = .548*). However, brain response significantly decreased for the same familiar identity compared to different familiar identities (t(79) = 3.83, *p <.001*). The adaptation to familiar faces was significantly larger than to unfamiliar faces (t(79) = 3.04, *p < .01*; Fig. 6.4C).

Finally, we asked whether there was any adaptation to faces in non face-selective regions. Table 6.3 shows the location of adaptation to faces outside the core and extended face-selective regions using a whole brain analysis. We found significant adaptation to familiar and unfamiliar faces in the precuneus. Adaptation to familiar faces was found in the middle temporal lobe and the insula. Adaptation to unfamiliar faces was found in the frontal pole.

Table 6.3 Adaptation to familiar and unfamiliar faces in non face-selective regions. Voxel-wise analysis was thresholded at *p < .05* (voxel-corrected for multiple comparisons).

	MNI coordinates			Adaptation (voxel)	
				familiar different > familiar same	familiar different > familiar same
	x	y	z		
precuneus	-2	-60	20	952	560
	4	-58	18	672	568
middle temporal	-58	-12	-10	616	-
	62	-8	-18	280	-
insula	34	-20	16	328	-
frontal pole	8	64	12	-	1016
	-2	62	14	-	464

6.5 Discussion

The aim of this study was to determine the neural correlates for an image-invariant representation of familiar faces, analogue to face recognition units (Bruce and Young, 1986). Using an fMR-adaptation paradigm, we found some evidence for an image-invariant representation of facial identity in the core face-selective regions. However, this adaptation did not differ between familiar and unfamiliar faces. In contrast to the core face-selective regions, image-invariant adaptation was found for familiar, but not

unfamiliar faces in face-selective regions of the medial temporal lobe (MTL) potentially underlying the behavioural familiar face recognition advantage.

Consistent with previous studies that have shown a bias towards familiar faces in face-selective regions, familiar faces generated a higher response compared to unfamiliar faces in the FFA, rSTS and MTL (Eger et al., 2005; Elfgren et al., 2006; Leveroni et al., 2000, Sergent et al., 1992). Overall, the familiarity effect was more pronounced in the left than right hemisphere, which is also consistent with previous studies (Elfgren et al., 2006; Gobbini & Haxby, 2007; Pierce et al., 2004). However, we did not find any selectivity for unfamiliar compared to familiar faces. This contrasts with previous studies that have found higher responses to unfamiliar faces in regions of the occipital and temporal lobes (Dubois et al., 1999; Gobbini et al., 2004; Mur, Ruff, Bodurka, Bandettini & Kriegeskorte, 2010). It is not clear why we did not find selectivity for unfamiliar faces in these regions, but is should be noted that this selectivity was not consistent across these studies.

Next, we investigated how face-selective regions represent identity information. Here, we found some image-invariant adaptation to facial identity in the core face-selective regions. This result is consistent with previous studies, which have shown adaptation to facial identity across different images (Rotshtein et al., 2005; Eger et al., 2005; Pourtois et al., 2005a; Ewbank & Andrews, 2008). This contrasts with other studies that have reported no adaptation to facial identity across different images within these regions (Davies-Thompson et al., 2009; Pourtois et al., 2005b). The ability to detect image-invariant adaptation in this study may reflect the increased power in our design. Although we found image-invariant identity adaptation, we found no interaction between identity processing and familiarity. The observed adaptation was equally pronounced for familiar compared to unfamiliar faces. This is in line with a recent study, where the

amount of image variation was systematically varied within a block and a similar release from adaptation was found for familiar as well as unfamiliar faces (Davies-Thompson et al., 2013). Together, it seems that the pattern of response in face-selective regions such as the FFA does not reflect the familiar face bias in face recognition.

In contrast to the core face-selective regions, image-invariant adaptation was found in the MTL. Consistently with Bruce and Young's model (1986), MTL adaptation was observed only for familiar but not unfamiliar faces. The MTL has previously been reported to respond more to familiar than to unfamiliar faces (Bar, Aminoff, & Ishai, 2008; Barense, Henson, & Graham, 2011; Eger et al., 2005; Elfgren et al., 2006; Ishai et al., 2002; Leveroni et al., 2000; Nielson et al., 2010; Pierce et al., 2004; Sergent et al., 1992). Additionally, it has been suggested to process semantic information associated with a face rather than the image itself (Haxby et al., 2000; Ranganath, Cohen, Dam, D'Esposito, 2004; Todorov, Gobbini, Evans, & Haxby, 2007). Our findings suggest that the involvement of MTL in familiar face recognition might reflect the activation of an image-invariant representation of person identity information (Burton et al., 1999).

The reduction in response evident in fMR-adaptation can be explained by a variety of different neural mechanisms (Grill-Spector et al., 2006; Krekelberg, Boynton, & van Wezel, 2006; Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). Models differ in terms of whether adaptation reflects a change in bottom-up processing (for example, neuronal fatigue or sharpening of the representation) or whether it reflects top-down control (for example, predictive processing or expectation). It is not possible to discriminate between these possibilities in this data. However, the role of the MTL in the recognition of familiar identity is consistent with studies that have recorded neural responses in the MTL. These electrophysiological studies have revealed that neurons in

the MTL respond to familiar identities across different stimuli such as different face images or even across modalities using face images and corresponding names (Quiroga et al., 2005; Nielson et al., 2010; Quiroga, Kraskov, Koch, & Fried, 2009). For example, Quiroga et al. (2005) investigated responses in the MTL in patients who had been implanted with depth electrodes. Subsets of neurons within the MTL seemed to respond selectively to a familiar identity across different stimuli, e.g. the actress Halle Berry. These neurons increased their firing rate for different photos of Halle Berry, her name, line drawings of her and even images depicting her masked as cat woman, a character she played. Such findings, suggest that the MTL might play a role in linking perception of a familiar face image with the knowledge associated with it (Ranganath et al., 2004; Todorov et al., 2007). Activity in the MTL might therefore reflect automatic retrieval of person-identity information from long-term memory.

These findings are consistent with neuropsychological studies of prosopagnosia. Lesions to the fusiform gyrus often lead to a selective deficit in the ability to accurately perceive a face (apperceptive prosopagnosia) rather than in the matching of an image with a memory of a facial identity (associative prosopagnosia; Barton et al., 2002; Barton, 2008). In contrast, lesions to anterior regions of the temporal lobe leave the perception of the face intact but impair recognition and semantic memory of people (Ellis, Young, & Critchley, 1989; Patterson, Nestor, & Rogers, 2007; Barton et al., 2008; Gainotti, 2014; Collins & Olson, 2014). Similar findings have been reported in people with developmental prosopagnosia. These studies reveal largely intact activity in the core face-selective regions, but reduced activity in anterior regions that may reflect a disruption in the connectivity between the core and extended face regions (Thomas et al., 2009; Avidan et al., 2014).

Similar conclusions have been drawn from neuroimaging studies using multivariate analyses that have implicated the ventral anterior temporal lobes in the generation of image-invariant representations of facial identity (Kriegeskorte et al., 2007; Nestor et al., 2011; Anzellotti et al., 2014). We did not find face-selective responses in this anterior regions. However, this may reflect signal distortion and drop out in the MRI signal in our data due to the proximity to the sinuses (Axelrod, & Yovel, 2013; Visser, Jefferies, & Ralph, 2010).

Finally, we asked whether adaptation was evident outside the face-selective regions. We found adaptation to familiar faces in the precuneus, insula and middle temporal lobe, consistent with a previous study by Pourtois and colleagues (2005b). We also found adaptation to unfamiliar faces in the precuneus and frontal pole consistent with previous studies (Davies-Thompson et al., 2009; Pourtois et al., 2005b).

In conclusion, we failed to find an image-invariant response to familiar but not unfamiliar faces in the core face-selective regions, such as the FFA. In contrast, we found image-invariant adaptation to specific to familiar faces in the MTL. This suggests that the recognition of familiar faces across different images relies on neural representations in the MTL.

Chapter 7 - Discussion

The ease with which we recognise a familiar identity from face images that vary in illumination, expression, viewing angle, and appearance contrasts with the inherent difficulty in the recognition of unfamiliar identities (Bruce et al., 1987; Davies-Thompson et al., 2009; Hancock et al., 2000; Jenkins et al., 2011; Megreya & Burton, 2006; Johnston & Edmonds, 2009). Based on these findings, cognitive models of face processing have suggested that there are different neural representations of familiar and unfamiliar faces (Bruce & Young, 1986; Burton et al.,1999). These models propose that all faces are initially represented by image-dependent neural code. However, it is only at later stages of processing that familiar faces are represented by an image-invariant code. This image-invariant identity representation has been suggested to underlie the familiar identity recognition advantage. Nevertheless, the neural mechanisms underlying this advantage remain unclear (Natu & O'Toole, 2011).

A wide range of studies have investigated the neural basis of face processing. Functional imaging studies have consistently found regions in the occipital and temporal lobes that respond selectively to faces (Kanwisher et al., 1997; McCarthy et al., 1997). Haxby and colleagues' (2000) neural model of face processing suggests that different regions process different aspects of faces. One region — the fusiform face area (FFA) — is suggested to represent invariant facial characteristics, such as identity, that are necessary for familiar face recognition (Grill-Spector et al., 2004; Haxby et al., 2000; Ishai, 2008). Consistent with these neuroimaging studies, neuropsychological studies show that lesions in the approximate location of the FFA can lead to selective impairments in face

recognition (Barton et al., 2002; Damasio et al., 1982). These findings make the FFA a promising candidate for the neural correlate of the familiar face recognition advantage.

However, neuroimaging studies directly comparing the response in the FFA to familiar and unfamiliar faces show mixed results. Some studies report stronger FFA activity for familiar compared to unfamiliar faces (Elfgren et al., 2006; Gobbini et al., 2004; Pierce et al., 2004; Sergent et al., 1992). However, others find no difference in response (Gorno-Tempini & Price, 2001; Leveroni et al., 2000) and one study found stronger response to unfamiliar than familiar faces (Gobbini et al., 2004).

Thus, it remains unclear from previous research which role the FFA plays in the recognition of identity, particularly whether activity in this area can explain the familiar face recognition advantage. Therefore the aims of this thesis were:

- To locate the the neural correlates of the familiar face recognition.

- To explore the representation of familiar faces within face-selective regions.

7.1 The neural correlate of the familiar face recognition advantage

If the familiar face recognition advantage arises from differential processing of familiar and unfamiliar faces, we would expect to find a brain region (or regions), which responds differently to familiar and unfamiliar faces. This difference in brain response should account for the difference in behavioural face recognition between familiar and unfamiliar faces. Chapter 3 combined behavioural and fMRI techniques to address this question using an individual differences approach. First, brain activity was measured to familiar and unfamiliar faces. Next, participants face recognition of familiar and unfamiliar

faces was assessed. A correlational analysis of brain activity and behavioural recognition data revealed that activity in the rFFA predicts the behavioural familiar face recognition advantage. That is, the higher the difference in response to familiar compared to unfamiliar faces in the rFFA, the better familiar compared to unfamiliar face recognition.

7.2 Representation of higher- and lower-level properties in face-selective regions.

Although the findings presented in Chapter 3 show the importance of FFA in familiar face recognition, the nature of the face representation within the FFA remains unclear. To address this, Chapter 4 investigated the organizing principles that underpin spatial patterns of response in the FFA. Previous studies have reported that there are distinct patterns of response to facial identities in the FFA that are independent of changes in viewpoint or expression (Anzellotti et al., 2013; Guntupalli et al., 2016; Natu et al., 2010; Nestor et al., 2011). However, other studies have shown distinct patterns of response to viewpoint or expression independent of identity (Axelrod & Yovel, 2012; Kietzmann et al., 2012; Said et al., 2010; Wegrzyn et al., 2015; Zhang et al., 2016). To explore the relative role of these higher-level properties of faces, I compared the relative role of identity and viewpoint. The results clearly showed that patterns of response in the FFA and other face regions were dominated by changes in viewpoint. A similar pattern of results was evident for the effects of viewpoint and expression – a clear dominance for viewpoint in all regions including the STS, which previous studies have shown to be selective for expression (Allison et al., 2000; Engell & Haxby, 2007, Harris et al., 2012; Pelphrey et al., 2004; Winston et al., 2004).

125

It is clear that changes in viewpoint give rise to a much larger change in the image than changes in identity or expression. Indeed, previous studies have shown that patterns of response in the ventral visual pathway can be predicted by low-level image properties (Andrews et al. 2015, Rice et al. 2014, Watson et al., 2016). So, it is possible that these patterns of response could be explained by sensitivity to lower-level properties of the image. To test this possibility, we measured image properties using the GIST descriptor and used the output to explain the observed brain responses. The GIST descriptor significantly explained patterns of brain response in all face-selective regions. Moreover, the effects were much larger than observed for viewpoint. Together, these findings suggest that the topographic organization within the FFA reflects lower-level image properties rather than higher-level face properties such as identity.

7.3 How is visual information about faces represented in face-selective regions?

Chapter 4 showed that patterns of response in face-selective regions could be predicted by lower-level properties of the image rather than by higher-level properties such as identity. The aim of the Chapter 5 was to understand how visual information about faces is represented in these regions.

There are many ways of thinking about the visual information conveyed by a photograph of a face, but one that has proved very useful is in terms of its shape and surface properties. Any facial image consists of a set of edges created by abrupt changes in reflectance due to the shapes and positions of facial features and a broader pattern of reflectance based on the surface properties of the face – also known as texture or albedo (Bruce & Young, 1998, 2012). Shape properties can be operationally defined by the spatial

locations of fiducial points that correspond to facial features; note that in this sense 'shape' properties will include both the feature shapes and their positions. In contrast, surface properties result from the pattern of reflectance of light due to the combination of ambient illumination, the face's pigmentation, and shape from shading cues.

Thinking of facial images as broadly consisting of shape (feature positions) and surface (pigmentation, shading patterns) properties has also helped our understanding of facial identity recognition, where both cues can contribute (Russell, Sinha, Biederman, & Nederhouser, 2006; Troje & Bülthoff, 1996). However, the role of surface cues becomes more salient for familiar faces (Burton et al., 2005; Russell & Sinha, 2007). The fMRI experiment in Chapter 5 measured patterns of response to different face images. I then asked whether the shape or surface properties could best explain these patterns of response. Contrary to our prediction, I found that shape but not surface properties could explain patterns of response in the FFA. This difference between the neural and perceptual responses to facial identity suggests that while the FFA may contribute to the early stages of analysis of invariant characteristics of faces, it is not itself responsible for the recognition of facial identity.

7.4 Image-invariant representations of familiar faces

Cognitive models of face processing suggest an image-invariant representation of familiar faces to underlie the familiar face recognition advantage. However, the findings reported in Chapters 4 and 5 did not provide any evidence for such a representation in the spatial pattern of response. Nevertheless, this does not mean that there is no image-invariant identity representation in the FFA. The representation could be finer spatial scale than is detected by MVPA. To address this issue, we used fMR-adaptation to investigate identity

representations within voxels (Grill-Spector & Malach, 2001). The principle behind fMRI adaptation is that repetition of a stimulus causes a reduction or habituation in the neural response, which leads to a lower fMRI signal. The sensitivity of the neural representation can then be determined for different changes to the stimulus. If the underlying neural representation is insensitive to a particular type of change in the stimulus, the reduction in fMRI signal for this type of change will be similar to the overall reduction produced by repetitions of identical stimuli. However, if the underlying neural representation is sensitive to this change, the fMRI signal will remain at its original (non-adapted) level.

The aim of Chapter 6 was to use fMR-adaptation to investigate image-invariant adaptation in face regions. Participants viewed either different identities or different images of the same identity. Additionally, the identities were either familiar or unfamiliar. If the representation in the FFA is linked to the recognition of facial identity, we expected a lower response (adaptation) when different images of the same familiar identity are repeated compared to when images of different familiar identities are shown. Moreover, similar adaptation should not be evident for unfamiliar faces.

We found image-invariant adaptation in the FFA for both familiar and for unfamiliar faces. However, there was also no difference in the magnitude of adaptation to familiar and unfamiliar faces. This does not fit with models of familiar face processing that predict adaptation only for familiar faces. None of the other core face regions showed any image-invariant adaptation to familiar or unfamiliar faces. However, we also found face-selective activity in the MTL. This region showed adaptation to familiar faces but not adaptation to unfamiliar faces. The MTL has been previously reported to store identity representations across different images or even different modalities, e.g. different images

of one identity and their spoken and written name (Quiroga, Kraskov, Koch, & Fried, 2009; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005; Nielson et al., 2010). Moreover, activity in the MTL has been suggested to reflect knowledge associated with an identity automatically activated by the face image (Haxby et al., 2000; Ranganath et al., 2004; Todorov et al., 2007). These results suggest that an image-invariant representation that could potentially underpin the behavioural familiar face recognition advantage seems to occur outside the FFA. A potential candidate for such a neural correlate seems to be the MTL, which shows an image-invariant representation for familiar but not unfamiliar identities.

One potential avenue of future research that may help understand the familiar face recognition advantage would be to examine how the spatial patterns of response observed in these studies fits temporal patterns of response shown in EEG and MEG studies. Event-related potential (ERP) and magnetoencephalography (MEG) studies have also shown that faces and other objects can be distinguished by the pattern of electrical activity across the occipitotemporal lobe (for review see Schweinberger & Neumann, 2016). Studies have investigated the importance of peaks in activity at 170msec and 250msec in the processing of facial identity (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Schweinberger,& Burton, 2003; Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002). Overall, these findings suggest that that the occipitotemporal N250r is highly sensitive to the familiarity of faces, and reflects the transient activation of individual representations of repeated familiar faces in repetition priming experiments. A combination of MEG/EEG and fMRI could help provide a more complete spatial and temporal account of how familiar faces are represented in the brain.

7.5 Conclusion

This thesis aimed to investigate the neural correlate of the behavioural familiar face recognition advantage. The first experiment provides some support for a link between response in the FFA and face recognition. The second, third, and fourth experiment do not provide much evidence for an image-invariant identity coding in the FFA, but rather coding of image properties important for face recognition. Finally, the fifth experiment shows some image-invariant adaptation in the FFA, but to both familiar and unfamiliar faces. However, we did find an image-invariant identity representation restricted to familiar faces outside the FFA, namely in the MTL.

According to Haxby and colleagues' model of face processing, the FFA represents the identity of faces image-invariantly. Although we did find an image-invariant adaptation in brain response intensity, our findings also show that the pattern of response in the FFA is to some degree image-dependent reflecting different viewpoints of faces or different face images. Thus it seems that the FFA contains image-invariant but also image-dependent information of faces, potentially forming a link between representing the specific image of a face and the abstract image-invariant representation of the identity of the face.

In conclusion, these results show that the FFA plays a role in familiar face recognition, such as in the encoding of shape information. However, this FFA representation seems not to be sufficient for recognition. Instead, activation of face representations in regions such as the MTL appear to be important for the retrieval of person knowledge and successful face recognition.

Bibliography

Aharon, I., Etcoff, N., Ariely, D., Chabris, C. F., O'Connor, E., & Breiter, H. C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, 32(3), 537-551.

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267-278.

Andrews, T. J. (2005). Visual cortex: How are objects and faces represented. *Current Biology*, 15(12), 451-453.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *NeuroImage*, 23(3), 905-913.

Andrews, T. J., Watson, D. M., Rice, G. E., & Hartley, T. (2015). Low-level properties of natural images predict topographic patterns of neural response in the ventral visual pathway. *Journal of Vision*, 15(7):3, 1-12.

Anzellotti, S., Fairhall, S. L., & Caramazza, A. (2013). Decoding representations of face identity that are tolerant to rotation. *Cerebral Cortex*, 24(8), 1988-1995.

Avidan, G., Tanzer, M., Hadj-Bouziane, F., Liu, N., Ungerleider, L. G., & Behrmann, M. (2014). Selective dissociation between core and extended regions of the face processing network in congenital prosopagnosia. *Cerebral Cortex*, 24(6), 1565-1578.

Axelrod, V., & Yovel, G. (2012). Hierarchical processing of face viewpoint in human visual cortex. *The Journal of Neuroscience*, 32(7), 2442-2452.

Axelrod, V., & Yovel, G. (2013). The challenge of localizing the anterior temporal face area: a possible solution. *NeuroImage*, 81, 371-380.

Axelrod, V., & Yovel, G. (2015). Successful Decoding of Famous Faces in the Fusiform Face Area. *PloS one*, 10(2), e0117126.

Baldassi, C., Alemi-Neissi, A., Pagan, M., DiCarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology*, 9(8), e1003167.

Bar, M., Aminoff, E., & Ishai, A. (2008). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex*, 18(6), 1233-1238.

Barense, M. D., Henson, R. N., & Graham, K. S. (2011). Perception and conception: temporal lobe activity during complex discriminations of familiar and novel faces and objects. *Journal of Cognitive Neuroscience*, 23(10), 3052-3067.

Barton, J. J. (2008). Structure and function in acquired prosopagnosia: lessons from a series of 10 patients with brain damage. *Journal of Neuropsychology*, 2(1), 197-225.

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, 58(1), 71-78.

Baseler, H., Harris, R., Young, A., & Andrews, T. (2014). Neural Responses to Expression and Gaze in the Posterior Superior Temporal Sulcus Interact with Facial Identity. *Cerebral Cortex*, 24, 737-744.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, 8(6), 551-565.

Bentin, S., & Deouell, L. Y. (2000). Structural encoding and identification in face processing: ERP evidence for separate mechanisms. *Cognitive Neuropsychology*, 17(1-3), 35-55.

Benton, A. L. (1980). The neuropsychology of facial recognition. *American Psychologist*, 35(2), 176-186.

Bi, T., Chen, J., Zhou, T., He, Y., & Fang, F. (2014). Function and Structure of Human Left Fusiform Cortex Are Closely Associated with Perceptual Learning of Faces. *Current Biology*, 24(2), 222-227.

Bilalić, M., Langner, R., Ulrich, R., & Grodd, W. (2011). Many faces of expertise: fusiform face area in chess experts and novices. *The Journal of Neuroscience*, 31(28), 10206-10214.

Bindemann, M., Burton, A. M., Leuthold, H., & Schweinberger, S. R. (2008). Brain potential correlates of face recognition: geometric distortions and the N250r brain response to stimulus repetitions. *Psychophysiology*, 45(4), 535e544.

Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353(6343), 429-431.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology*, 73(1), 105-116.

Bruce, V. (1994). Stability from variation: The case of face recognition. *Quarterly Journal of Experimental Psychology*, 47, 5-28.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P. J., Burton, A. M., & Miller, P. (1999). Verification of face identities from images captured on video. *Journal of Experimental Psychology: Applied*, 5(4), 339-360.

Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7(3), 207-218.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception*, 23(7), 803-822.

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1(2), 109-120.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, 77(3), 305-327.

Bruce, V., & Young, A. (1998). In the eye of the beholder: The science of face perception. Oxford: Oxford University Press.

Bruce, V., & Young, A. (2012). Face perception. Hove, East Sussex: Psychology Press.

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology*, 66, 1467-1485.

Burton, A. M., Bruce, V., & Hancock, P. J. (1999). From pixels to people: A model of familiar face recognition. *Cognitive Science*, 23(1), 1-31.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D. (2005). Robust representations for face recognition: the power of averages. *Cognitive Psychology*, 51, 256-84.

Burton, A. M., Kramer, R. S. S., Ritchie, K. L., & Jenkins, R. (2016). Identity from variation: Representations of faces derived from multiple instances. *Cognitive Science*, 40(1), 202–223.

Caharel, S., Poiroux, S., Bernard, C., Thibaut, F., Lalonde, R., & Rebai, M. (2002). ERPs associated with familiarity and degree of familiarity during face recognition. *International Journal of Neuroscience*, 112(12), 1499-1512.

Caharel, S., Ramon, M., & Rossion, B. (2014). Face familiarity decisions take 200 msec in the human brain: electrophysiological evidence from a go/no-go speeded task. *Journal of Cognitive Neuroscience*, 26(1), 81-95.

Calder, A. J., & Young, A. W. (2005). Understanding the recognition of facial identity and facial expression. *Nature Reviews Neuroscience*, 6(8), 641-651.

Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*, 83(1), 1-29.

Collins, J. A., & Olson, I. R. (2014). Beyond the FFA: the role of the ventral anterior temporal lobes in face processing. *Neuropsychologia*, 61, 65-79.

Damasio, A. R., Damasio, H., & Van Hoesen, G. W. (1982). Prosopagnosia Anatomic basis and behavioral mechanisms. *Neurology*, 32(4), 331-341.

Davies, G., Ellis, H. D., & Shepherd, J. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 63, 180–187.

Davies-Thompson, J., & Andrews, T. J. (2012). Intra-and interhemispheric connectivity between face-selective regions in the human brain. *Journal of Neurophysiology*, 108(11), 3087-3095.

Davies-Thompson, J., Gouws, A., & Andrews, T. J. (2009). An image-dependent representation of familiar and unfamiliar faces in the human ventral stream. *Neuropsychologia*, 47(6), 1627-1635.

Davies-Thompson, J., Newling, K., & Andrews, T. J. (2013). Image-Invariant responses in face-selective regions do not explain the perceptual advantage for familiar face recognition. *Cerebral Cortex*, 23, 370-377.

Doerr, P., Herzmann, G., & Sommer, W. (2011). Multiple contributions to priming effects for familiar faces: analyses with backward masking and event-related potentials. *British Journal of Psychology*, 102, 765e782.

Dubois, S., Rossion, B., Schiltz, C., Bodart, J.-M., Michel, C., Bruyer, R., & Crommelinck, M. (1999). Effect of familiarity on the processing of human faces. *NeuroImage*, 9(3), 278-289.

Eger, E., Schweinberger, S., Dolan, R., & Henson, R. (2005). Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. *NeuroImage*, 26(4), 1128-1139.

Elfgren, C., van Westen, D., Passant, U., Larsson, E.-M., Mannfolk, P., & Fransson, P. (2006). fMRI activity in the medial temporal lobe during famous face processing. *NeuroImage*, 30(2), 609-616.

Ellis, A. W., Young, A. W., & Critchley, E. M. (1989). Loss of memory for people following temporal lobe damage. *Brain*, 112(6), 1469-1483.

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7(2), 181-192.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, 45(14), 3234-3241.

Epstein, R. A., Higgins, J. S., Jablonski, K., & Feiler, A. M. (2007). Visual scene processing in familiar and unfamiliar environments. *Journal of Neurophysiology*, 97(5), 3670-3683.

Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44(3), 227-240.

Ewbank, M. P., & Andrews, T. J. (2008). Differential sensitivity for viewpoint between familiar and unfamiliar faces in human visual cortex. *NeuroImage*, 40(4), 1857-1870.

Farah, M. J. (1996). Is face recognition 'special'? Evidence from neuropsychology. *Behavioural Brain Research*, 76(1), 181-189.

Farah, M. J., Levinson, K. L., & Klein, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia*, 33(6), 661-674.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science*, 330(6005), 845-851.

Friston, K. J., Frith, C. D., Turner, R., & Frackowiak, R. S. (1995). Characterizing evoked hemodynamics with fMRI. *NeuroImage*, 2(2), 157-165.

Furl, N., Garrido, L., Dolan, R. J., Driver, J., & Duchaine, B. (2011). Fusiform gyrus face selectivity relates to individual differences in facial recognition ability. *Journal of Cognitive Neuroscience*, 23(7), 1723-1740.

Gainotti, G. (2014). The Neuropsychology of Familiar Person Recognition from Face and Voice. *Psychologica Belgica*, 54(3), 298-309, DOI: http://dx.doi.org/10.5334/pb.at

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3(2), 191-197.

Gobbini, I. M., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, 45(1), 32-41.

Gobbini, I. M., Leibenluft, E., Santiago, N., & Haxby, J. V. (2004). Social and emotional attachment in the neural representation of faces. *NeuroImage*, 22(4), 1628-1635.

Goesaert, E., & Op de Beeck, H. P. (2013). Representations of facial identity information in the ventral visual stream investigated with multivoxel pattern analyses. *The Journal of Neuroscience*, 33(19), 8549-8558.

Gorno-Tempini, M., & Price, C. (2001). Identification of famous faces and buildings A functional neuroimaging study of semantically unique items. *Brain*, 124(10), 2087-2097.

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia*, 49(9), 2736-2745.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetition and the brain: neural models of stimulus-specific effects. *Trends in Cognitive Sciences*, 10(1), 14-23.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, 7(5), 555-562.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychologica*, 107(1), 293-321.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24(1), 187-203.

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2016). Disentangling the representation of identity from head view along the human face processing pathway. *bioRxiv*, 045823.

Haber, S. N., & Knutson, B. (2010). The reward circuit: linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35(1), 4-26.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330-337.

Hanke, M., Halchenko, Y. O., Sederberg, P. B., Hanson, S. J., Haxby, J. V., & Pollmann, S. (2009). PyMVPA: A python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*, 7(1), 37-53.

Harley, E. M., Pope, W. B., Villablanca, J. P., Mumford, J., Suh, R., Mazziotta, J. C., ... & Engel, S. A. (2009). Engagement of fusiform cortex and disengagement of lateral occipital cortex in the acquisition of radiological expertise. *Cerebral Cortex*, 19(11), 2746-2754.

Harris, R. J., Rice, G. E., Young, A. W., & Andrews, T. J. (2015). Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. *Cerebral Cortex*, 1-8.

Harris, R. J., Young, A. W., & Andrews, T. J. (2012). Morphing between expressions dissociates continuous from categorical representations of facial expression in the human brain. *Proceedings of the National Academy of Sciences*, 109(51), 21164-21169.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, 37, 435-456.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539), 2425-2430.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223-233.

Hay, D., & Young, A. W. (1982). The human face. *Normality and Pathology in Cognitive Functions*, 173-202.

Henson, R., Shallice, T., & Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, 287(5456), 1269-1272.

Herzmann, G., Schweinberger, S. R., Sommer, W., & Jentzsch, I. (2004). What's special about personally familiar faces? A multimodal approach. *Psychophysiology*, 41, 688e701.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, 3, 80-84.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31(10), 1221–1240.

Huang, L., Song, Y., Li, J., Zhen, Z., Yang, Z., & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Frontiers in Human Neuroscience*, 8, 483-492.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1), 215-243.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). Functional magnetic resonance imaging (Vol. 1): Sinauer Associates Sunderland.

Ishai, A. (2008). Let's face it: it's a cortical network. *NeuroImage*, 40(2), 415-419.

Ishai, A., Haxby, J. V., & Ungerleider, L. G. (2002). Visual imagery of famous faces: effects of memory and attention revealed by fMRI. *NeuroImage*, 17(4), 1729-1741.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, 96(16), 9379-9384.

Jenkins, R., White, D., Van Montfort, X., & Burton, M. A. (2011). Variability in photos of the same face. *Cognition*, 121(3), 313-323.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory*, *17*, 577-596.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302-4311.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: photographs, credit cards and fraud. *Applied Cognitive Psychology*, 11, 211-222.

Kietzmann, T. C., Swisher, J. D., König, P., & Tong, F. (2012). Prevalence of selectivity for mirror-symmetric views of faces in the ventral and dorsal visual pathways. *The Journal of Neuroscience*, 32(34), 11763-11772.

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, 29(5), 250-256.

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences*, 104(51), 20600-20605.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2(4), 1-28.

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, 24(8), 1377-1388.

Leder, H. (1999). Matching person identity from facial line drawings. *Perception*, 28, 1171–1175.

Leveroni, C. L., Seidenberg, M., Mayer, A. R., Mead, L. A., Binder, J. R., & Rao, S. M. (2000). Neural systems underlying the recognition of familiar and newly learned faces. *The Journal of Neuroscience*, 20(2), 878-886.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature Neuroscience*, 8(10), 1386-1391.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412(6843), 150-157.

Longmore, C. A., Liu, C. H., & Young, A. W. (2008). Learning faces from photographs. *Journal of Experimental Psychology: Human Perception & Performance*, 34, 77-100.

Malone, D. R., Morris, H. H., Kay, M. C., & Levin, H. S. (1982). Prosopagnosia: a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery & Psychiatry*, 45(9), 820-822.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, 9(5), 605-610.

McGugin, R. W., Gatenby, J. C., Gore, J. C., & Gauthier, I. (2012). High-resolution imaging of expertise reveals reliable object selectivity in the fusiform face area related to perceptual performance. *Proceedings of the National Academy of Sciences*, 109(42), 17063-17068.

Mckelviet, S. J. (1973). The meaningfulness and meaning of schematic faces. *Perception & Psychophysics*, 14(2), 343-348.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *The Quarterly Journal of Experimental Psychology*, 46(1), 1-10.

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, 34(4), 865-876.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience*, 9(5), 555-604.

Mur, M., Ruff, D. A., Bodurka, J., Bandettini, P. A., & Kriegeskorte, N. (2010). Face-identity change activation outside the face system: "release from adaptation" may not always indicate neuronal selectivity. *Cerebral Cortex*, 20, 2027–2042.

Natu, V. S., Jiang, F., Narvekar, A., Keshvari, S., Blanz, V., & O'Toole, A. J. (2010). Dissociable neural patterns of facial identity across changes in viewpoint. *Journal of Cognitive Neuroscience*, 22(7), 1570-1582.

Natu, V., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology*, 102(4), 726-747.

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proceedings of the National Academy of Sciences*, 108(24), 9998-10003.

Nielson, K. A., Seidenberg, M., Woodard, J. L., Durgerian, S., Zhang, Q., Gross, W. L., . . . Rao, S. M. (2010). Common neural systems associated with the recognition of famous faces and names: an event-related fMRI study. *Brain and Cognition*, 72(3), 491-498.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, 17(4), 580-590.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24), 9868-9872.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, 47(6), 419-427.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3), 145-175.

Paller, K. A., Gonsalves, B., Grabowecky, M., Bozic, V. S., & Yamada, S. (2000). Electrophysiological correlates of recollecting faces of known and unknown individuals. *NeuroImage*, 11(2), 98-110.

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8(12), 976-987.

Pelphrey, K. A., Viola, R. J., & McCarthy, G. (2004). When strangers pass processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science*, 15(9), 598-603.

Pfütze, E. M., Sommer, W., & Schweinberger, S. R. (2002). Age-related slowing in face and name recognition: evidence from event-related brain potentials. *Psychology and aging*, 17(1), 140.

Pierce, K., Haist, F., Sedaghat, F., & Courchesne, E. (2004). The brain response to personally familiar faces in autism: findings of fusiform activity and beyond. *Brain*, 127(12), 2703-2716.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology*, 17(18), 1568-1573.

Poldrack, R. A. (2007). Region of interest analysis for fMRI. *Social Cognitive and Affective Neuroscience*, 2(1), 67-70.

Poldrack, R. A., Halchenko, Y. O., & Hanson, S. J. (2009). Decoding the large-scale structure of brain function by classifying mental states across individuals. *Psychological Science*, 20(11), 1364-1372.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005a). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of Cognitive Neuroscience*, 17(7), 1043-1057.

Pourtois, G., Schwartz, S., Seghier, M. L., Lazeyras, F., & Vuilleumier, P. (2005b). View-independent coding of face identity in frontal and temporal cortices is modulated by familiarity: an event-related fMRI study. *NeuroImage*, 24(4), 1214-1224.

Quiroga, R. Q., Kraskov, A., Koch, C., & Fried, I. (2009). Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology*, 19(15), 1308-1313.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102-1107.

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *The Journal of Neuroscience*, 24(16), 3917-3925.

Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *The Journal of Neuroscience*, 34(26), 8837-8844.

Rhodes, G., Byatt, G., Michie, P., & Puce, A. (2004). Is the fusiform face area specialized for faces, individuation, or expert individuation? *Journal of Cognitive Neuroscience*, 16(2), 189-203.

Rotshtein, P., Henson, R. N. A., Treves, A., Driver, J., & Dolan, R. J. (2004). Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nature Neuroscience*, 8(1), 107-113.

Russell, R., Biederman, I., Nederhouser, M., & Sinha, P. (2007). The utility of surface reflectance for the recognition of upright and inverted faces. *Vision Research*, 47, 157–165.

Russell, R., & Sinha, P. (2007). Real-world face recognition: The importance of surface reflectance properties. *Perception*, 36(9), 1368–1374.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important for face recognition? Evidence from contrast negation. *Perception*, 35(6), 749-759.

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision*, 10(5): 11, 1-12.

Schweinberger, S. R., & Burton, A. M. (2003). Covert recognition and the neural system for face processing. *Cortex*, 39(1), 9-30.

Schweinberger, S. R., Pfütze, E. M., & Sommer, W. (1995). Repetition priming and associative priming of face recognition: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(3), 722.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research*, 14(3), 398-409.

Schweinberger, S. R., & Neumann, M. F. (2016). Repetition effects in human ERPs to faces. *Cortex*, 80, 141-153.

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing A positron emission tomography study. *Brain*, 115(1), 15-36.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., Mitchell, T. M., & Just, M. A. (2008). Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. *PLoS One*, 3(1), e1394.

Sormaz, M., Watson, D. M., Smith, W. A., Young, A. W., & Andrews, T. J. (2016). Modelling the perceptual similarity of facial expressions from image statistics and neural responses. *NeuroImage*, 129, 64-71.

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, 11(9), 1004-1006.

Tarr, M. J., & Gauthier, I. (2000). FFA: a flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, 3, 764-770.

Thomas, C., Avidan, G., Humphreys, K., Jung, K. J., Gao, F., & Behrmann, M. (2008). Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nature Neuroscience*, 12(1), 29-31.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, 45(1), 163-173.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, 36(12), 1761-1771.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79(4), 471-491.

Visser, M., Jefferies, E., & Ralph, M. L. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083-1094.

Watson, D. M., Young, A. W., & Andrews, T. J. (2016). Spatial properties of objects predict patterns of neural response in the ventral visual pathway. *NeuroImage*, 126, 173-183.

Wegrzyn, M., Riehle, M., Labudda, K., Woermann, F., Baumgartner, F., Pollmann, S., . . . Kissler, J. (2015). Investigating the brain basis of facial expression perception using multi-voxel pattern analysis. *Cortex*, 69, 131-140.

Weibert, K., & Andrews, T. J. (2015). Activity in the right fusiform face area predicts the behavioural advantage for the perception of familiar faces. *Neuropsychologia*, 75, 588-596.

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology*, 92(3), 1830-1839.

Xu, X., Yue, X., Lescroart, M. D., Biederman, I., & Kim, J. G. (2009). Adaptation in the fusiform face area (FFA): image or person? *Vision research*, 49(23), 2800-2807.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81, 141-145.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research*, 46, 3802-3811.

Young, A. W., McWeeny, K. H., Hay, D. C., & Ellis, A. W. (1986). Matching familiar and unfamiliar faces on identity and expression. *Psychological Research*, 48(2), 63-68.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology*, 15(24), 2256-2262.

Zhang, H., Japee, S., Nolan, R., Chu, C., Liu, N., & Ungerleider, L. G. (2016). Face-selective regions differ in their ability to classify facial expressions. *NeuroImage*, 130, 77-90.