The Importance of Image Properties in the Perception and Neural Representation of Faces.

Daniel Rogers

Doctor of Philosophy

University of York

Psychology

February 2024

Abstract

The overarching aim of this thesis, is to explore the role of visual information in the perception and recognition of faces. The experiments in this thesis use a combination of behavioural, computational and neuroimaging approaches to ask how facial representations are formed and what image properties are critical for perception and recognition. The thesis begins (Chapter 3) by investigating the process of generating a view-invariant representation from view-specific inputs. Using behavioural, neuroimaging and computational approaches, the results provide evidence of an intermediate view-symmetric representation. The emergence of view-symmetric representations from view-specific inputs was evident for canonical, but not non-canonical rotations of the face. Chapter 4 asked what visual information in the face (shape or texture) is important when making identity judgments. Here, it was shown that whilst texture properties are the dominant cue for familiar face recognition, shape properties provide unique and important contributions when making identity judgments, with the face-selective areas showing an equal sensitivity to both properties. The importance of shape and texture was further explored in Chapter 5, which showed that there was an intermediate band of image dimensions that were fundamental for familiar face recognition and for learning new faces. In contrast, earlier and later image dimensions of ambient face images were not important for recognition or learning. The final study of the thesis (Chapter 6) explored how image properties influence other aspects of face perception. The results showed that early image dimensions were critical for judgments of gaze, whereas different intermediate bands of image dimensions were important for judgements of gender and expression. Taken together, the findings presented in this thesis address the questions of how image properties are important for the perception and neural representation of face identity and whether the same or different image properties are important for other aspects of face perception.

List of Contents

Table of Contents

List of Tables

List of Figures

Acknowledgements

I express my deepest gratitude to Professor Tim Andrews, my supervisor, mentor, and friend throughout this challenging yet most incredibly rewarding journey. Your unwavering support, insightful feedback, and dedication to academic excellence have played a pivotal role in shaping the trajectory of my research and the production of this thesis. At every juncture, you guided me and provided the tools I did not even know I needed to overcome any challenges that arose. You have truly imparted invaluable lessons, both professionally and personally, which I will always carry with me and for which I will be forever indebted. Encountering a supervisor with your blend of intelligence, passion, and charisma is rare, and I feel blessed to have experienced it. Thank you, Tim.

A special thank you is also warranted to Dr Mila Mileva, for which without your expertise and willingness to collaborate, the experiments in Chapters 5 and 6 would have been much harder to complete.

This brief acknowledgment section unfortunately cannot adequately recognise the numerous remarkable individuals who have influenced my PhD experience. However, special thanks are needed for the future Dr. Noad and Dr. Chinzara. Kira, from even before day one of the PhD, you have been my go-to person for anything and everything. Your technical expertise, kindness, and understanding have been invaluable during the undertaking of my PhD, and you managed to navigate it all with humour and positivity, despite the many challenges we both faced along the way. Trish, from the moment you joined the office, you immediately made your wonderful presence known. Your attitude and tenacity for exploring new methods, knowledge, and opportunities have shown me how rewarding it can be to step outside of your comfort zone. I will always fondly look back on the experiences we have all had together, as I know that there are many more yet to come.

Finally, to my mum – none of this would have been possible without the years of sacrifice and selflessness you have demonstrated. Despite not fully understanding the specifics of my research field (even now), you have been right behind me every step of the way, ensuring that I moved forward with nothing but confidence in new endeavours. I could not have made it this far without your support, and I will continue to try and make you proud.

Author's Declaration

This thesis presents a collection of original work completed by the author, Daniel Rogers, under the supervision of Prof. Timothy Andrews. The empirical work presented in this thesis has been published in the following peer-reviewed journals:

Chapter 3- Experiments 1 & 2

Rogers, D., & Andrews, T. J. (2022). The emergence of view-symmetric neural responses to familiar and unfamiliar faces. *Neuropsychologia*, *172*, 108275.

Chapter 4- Experiments 1, 2 & 3

Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2022). The roles of shape and texture in the recognition of familiar faces. *Vision Research*, *194*, 108013.

Chapter 5- Experiments 1 & 2

Andrews, T. J., Rogers, D., Mileva, M., Watson, D. M., Wang, A., & Burton, A. M. (2023). A narrow band of image dimensions is critical for face recognition. *Vision Research*, *212*, 108297.

Results from multiple empirical chapters have been presented at the following conferences:

Talk- Rogers, D. (2023) What image properties contribute towards face recognition. *Invited talk to be presented at the Rank Symposium on the Visual Processing of Faces and Words,* October, 2023.

Talk- Rogers, D., T. Andrews., Mileva, M. (2023). A narrow band of image dimensions is critical for face recognition. *Talk presented at Vision Sciences Society- VSS, Florida,* May 2023.

Poster- Rogers, D., & Andrews, T. J. (2023) The emergence of view-symmetric neural responses to familiar and unfamiliar faces. *Poster presented at Faces and Deep Neural Networks conference, Notre Dame*, March 2023.

Talk- Rogers, D., T. Andrews., Mileva, M.(2023). A narrow band of image dimensions is critical for face recognition*. Talk presented at EPS* University College London, London January 2023.

Talk- Rogers, D. (June 2022). What image dimensions are important for familiar face recognition. *Talk presented at York and York St John PGR Conference,* York, June 2022.

Poster- Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2022). The contributions of shape and surface texture in the recognition of familiar faces. *Poster presented at British Cognitive Neuroscience Association (BACN) conference,* Birmingham May 2022.

Talk- Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2021). 'Does shape information contribute towards identity recognition?' *Talk presented at ESRC White Rose online conference*, November 2021.

Poster- Rogers, D., & Andrews, T. J. (2021). 'The emergence of view-symmetric neural responses to familiar and unfamiliar faces'. *Poster presented at ESRC AQM online conference*, November 2021.

Chapter 1- Literature Review- How are Faces Represented?

1.1 Information from Faces

Faces provide a range of socially relevant information (Bruce & Young, 2012). For example, the recognition of identity - detecting whether a face belongs to an individual we are familiar with or whether it belongs to a stranger - is critical in guiding social interactions (Bruce & Young, 1986). However, human faces consist of the same basic structure, two eyes above a nose above a mouth (Maurer, Le Grand & Mondloch, 2002). Thus, a fundamental challenge for human observers is the ability to rapidly and accurately process identity. This thesis will explore how image properties of the face contribute to the recognition of identity.

1.1.2 Faces as Images

When considering how we recognise identity from a face, it is important to consider what information is available in the image. At typical viewing distances, images of faces are composed of different spatial frequencies (SFs), each providing unique information about the image (Tian et al., 2018). Low spatial frequency (LSF) information depicts the global characteristics of the face, whereas high spatial frequency (HSF) information conveys the finer grained detail of the internal facial features (Kihara & Takeda 2019; Bar, 2004). This information from faces can be divided into two properties: shape and surface texture (Bruce & Young, 1998, 2012). For example, any facial image comprises of a set of edges generated by (higher spatial frequency) abrupt changes in reflectance due to the shapes and positions of facial features. These shape properties usually arise from how the 3D geometrical description of the face is projected onto a 2D image. Facial images also contain a broader pattern of (lower spatial frequency) reflectance based on the surface properties of the face. Texture 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.

Previous research shows that texture properties play a critical role when perceiving the identity of a familiar face (Burton, 2013). Early studies investigating the effect of contrast negation suggest that texture properties (in particular pigmentation and surface brightness) are critical when making judgements regarding identity (Bruce & Langton, 1994; Kemp, Pike, White, & Musselman, 1996). Negation reverses the contrast polarities of an image, making black areas white, light grey areas dark grey, and so forth. It is a reversible manipulation that does not remove any information from the image. Despite no information being lost, the ability to process texture information is affected (White, 2001). Russell and Sinha (2006), used two sets of faces, where the individual faces differed in either shape or texture. During a face matching task, they found that performance was impaired by contrast negation only when faces varied in texture. Moreover, Sandford, Sarker and Bernier (2018) found that this effect of contrast negation affecting recognition performance did not generalise to other classes of familiar objects (in this case company logos).

In contrast to the role of texture, simple spatial transformations of a face suggest that shape information may not be critical for identity recognition. For example, when linearly stretching a face horizontally or vertically, recognition rates of familiar faces are unaffected (Hole, George, Eaves & Rasek, 2002). However, when keeping the shape of the face constant but altering its texture (by blurring neighbouring pixels together), successful recognition declines by half (Hole, George, Eaves & Rasek, 2002). Furthermore, familiar face recognition is not considerably affected when texture properties are presented onto a standardised shape (Burton, Jenkins, Hancock & White, 2005). In contrast, when presented with a facial line-drawing (thus, removing all texture information) and a photograph, face matching becomes significantly worse than when using two photographs (Leder, 1999). Together, these studies imply that firstly, texture properties unique to an identity, provide the dominant cue to distinguish and recognise identities familiar to us. Secondly, shape information is seen as a less reliable indicator of identity, as shape cues (mainly from internal facial features) can vary quite dramatically across images of the same identity (Burton, 2013; Andrews, Baseler, Jenkins & Burton, 2016).

Although the behavioural studies outlined above show that texture may be a more important cue, this does not mean that shape properties do not contribute to the recognition of identity recognition. Hole, George, Eaves and Rasek (2002) operationalised successful recognition as participants identifying if a face was famous or not. In natural viewing, however, face recognition relies on activating the stored mental representations of an identity (synonymous to a facial-recognition-unit, Bruce & Young, 1986), which can then trigger further processing such as activating a name code for a recognised identity. So, tasks involving judgements of familiarity may not reveal how successful recognition is naturally expressed. Interestingly, Hole and colleagues (2002) found that non-linear vertical stretching of the face impaired recognition, implying that configural information from the stretched part of the face was influencing the process of recognition. The authors concluded that the configural information used as a basis for face recognition is unlikely to involve information about the absolute position of facial features relative to each other. However, an alternative possibility is that this finding implies that both shape and texture information are important for recognition and that the two properties are used in tandem when making recognition judgements. Furthermore, Leder (1999) used a face-matching task in which line-drawings and photographs of unfamiliar faces were presented in pairs and recognition was calculated based on hits and false-alarm rates. Again, this paradigm does not allow for the activation of a stored mental representation of identities as the images used were unfamiliar. Additionally, whilst recognition was worse when matching a line-drawing to a photograph compared to matching two photographs, performance using line-drawings containing purely shape information was still high. For example, *A′* was 0.97 when using two photographs and dropped to 0.84 when matching a line-drawing to a photograph and remained stable when even matching a line drawing to a novel viewpoint of the face. This therefore would imply that the most basic shape information generated in a line-drawing can be used for making recognition judgements. Lastly, Galper (1970) revealed that identity recognition reduced to 50% when the images were contrast negated. Whilst this is a significant decrease in performance, the fact faces can still be recognised to some extent, implies there is good potential that shape information does have a role in the representation of facial identity. Taken together, it is important to note that whilst texture properties are important for recognising identity, the contribution of shape properties might have been overlooked due to methodological paradigms and conflating shape from shading as a purely texture component. Therefore, it is of current interest to explore the roles of shape and texture in familiar face recognition and how these properties are represented in the brain.

1.2 Models of Face Recognition

1.2.1 Bruce and Young Model (1986)

Bruce and Young (1986) reported an influential framework for the perception of faces. The framework details that after the initial encoding of the structure of a face, there is a functional and spatial separation of processing, for changeable and unchangeable facial properties. This distinction originated from neuropsychological case studies documenting the experiences of patients with specific localised deficits within identity recognition (prosopagnosia), whilst having intact expression recognition (a changeable aspect of the face). This processing independence of expression and identity is further supported by individuals who report impairments with expression identification whilst performing similarly to control participants in identity recognition, evidencing a face processing double dissociation (Humphreys, Donnelly, & Riddoch, 1993; Parry, Young, Saul & Moss, 1991; Young, Newcombe, De Haan, Small, & Hay, 1993).

Under this model, the recognition of faces is based on an initial image-based representation followed by the activation of a view-invariant structural code. When seeing a face, it is proposed that an image-based or pictorial representation is generated irrespective of familiarity. Using this representation, basic processing occurs to determine a variety of face signals such as age, gender and gaze, regardless of familiarity to the face stimulus. It is proposed that the recognition of a familiar face relies on a match between the initial image-based representation of the face and previously stored structural representations located within face recognition units (FRUs). Following this match, a cascade of responses occurs, including the activation of a personal identity node (if the face is sufficiently similar) to access identity-specific semantic codes from which name codes are subsequently recovered. If a face is unfamiliar or has low familiarity then only an initial viewer-centred pictorial representation can be accessed, reflecting the limited identity specific information obtainable.

Evidence in support of this model stems from seminal research that has explored the key differences in the processing of familiar and unfamiliar faces. For example, when judging whether two faces belong to the same identity, participants are highly accurate for familiar faces, but often have difficulty deciding whether two unfamiliar identities are from the same or different identity (Bruce et al., 1999; Davies-Thompson, Gouws & Andrews, 2009). This familiar face advantage is also exacerbated when viewpoint, orientation and expression are manipulated (Bruce, Valentine & Baddeley, 1987; Bruce et al., 2001). The Bruce and Young (1986) model accounts for these effects by suggesting that faces are first encoded pictorially, with this representation allowing for the recognition of both familiar and unfamiliar faces when the same images are used. However, when image conditions are manipulated or different images of the same identity are used, this initial pictorial code is not sufficient for recognition. For this, a more abstract and flexible representation is needed which encodes identity-image-invariantly.

Whilst the function of the FRUs has provided a key theoretical underpinning to the recognition of identity from faces, the underlying image dimensions remain unclear. Understanding the nature of these facial representations, specifically how they are triggered by the many possible variations of someone's face exclusively, as well as what image properties are important for recognition, will allow for a fuller understanding of facial recognition.

1.2.2 Interactive Activation Competition (IAC) Model (1999)

Although classical models of face recognition (Bruce & Young, 1986) recognise the importance of FRUs, the nature of the visual properties that are used in this structural code are not clearly defined. Burton and colleagues address this issue in the Interactive Activation Competition model (IAC, Burton, Bruce & Hancock, 1999). This model combines a perceptual 'front end' based on principal component analysis (PCA) of face images and a cognitive back-end based on a simple interactive activation and competition processes. This model was the first to consider what image dimensions underlie the structural representation of identity.

PCA aims to derive statistical descriptions of image sets (Moon & Phillips, 2001; Gong et al., 2000) and has been used to explain the variance in face images that can be related to perception (Turk & Pentland, 1991; O'Toole et al., 1993; Hancock, Burton & Bruce, 1996; Calder et al., 2001; Jozwik et al., 2022). When PCA is applied to a set of faces, it delivers novel dimensions (eigenfaces), which together form a multi-dimensional framework, within which to characterise any face image (Scheuchenpflug, 1999; Tredoux et al., 2002; Nestor et al., 2013). Within this framework, 'early' dimensions capture the most variance within a learning set, and tend to be associated with coarse-scale image variation (for example coding changes in head orientation or whether an image is brighter on one side or the other). Later components, capturing progressively smaller variance, tend to capture finer-scale information. When PCA is applied to a particular set of faces, it is important to note that, given a sufficiently large sample, the resulting space generalises well. So, components derived from one set of faces, tend to capture the variance of novel sets well – particularly if training sets incorporate a range of variation in face images.

In the original version of the IAC model, Burton, Bruce and Hancock (1999) tested three variations of a PCA analysis: 1) raw images (not subjected to a shape-free transformation), 2) shape-free (texture PCA) and 3) shape-free plus shape, to ascertain which model best captured recognition of neutral and expressive faces. Here they found that the latter two models best captured identity, with a very small increase in expressive face recognition when shape information was added into the PCA. Thus, after subsequent testing of the shape-free model, it was concluded that this model, best captured the underlying structural representation that could drive the process of face recognition, as the additional shape information was seen as less reliable. However, it has been shown that in certain cases, shape information can be useful for the purpose of identification (McKone & Yovel, 2009; Tanaka & Gordon, 2011; Piepers & Robbins, 2012) and that some texture information is not diagnostic for recognition, for example, ambient lighting information. Taken together, it is therefore appropriate to ask how information about identity is carried in the components derived from both a shape and texture PCA, that underlies the structural representation of identity. For example, it remains unclear whether and to what extent certain image dimensions are more useful than others for capturing identity information.

The IAC model also extends the Bruce and Young (1986) model by explaining how FRUs can be activated. Here, it is suggested that there are interconnected nodes representing different concepts or features, and the activation spreads through these nodes based on their connections and interactions. Nodes compete with each other for activation, leading to dynamic patterns of activation that can simulate cognitive processes. In addition to FRUs and PINs (personal identity nodes), the IAC model suggests that following PINs, there are pools of interconnecting nodes labelled Semantic Information Units (SIUs) and Name Information Units (NUIs) which code information about known individuals and their names (respectively). Information about a person is coded in the form of a link between the person's PIN and the relevant SIU. Note that many SIUs will be shared (e.g. there may be many people represented with occupation "singer" or with nationality "American"). The notion is that activation of any of these units to a common threshold allows retrieval of that piece of information. Finally, there is a pool of units labelled "lexical output" which are intended to capture the first stage of processes involved in speech and other output modalities. Thus, this model offers a more distributed account of face processing whereby many factors can impact one’s ability to recognise identity from many sources of information.

1.2.3 Multidimensional Face Space Model

The Multidimensional Face Space Model also provides a framework to explain how faces are represented in memory (Valentine, 1991). Here, each face is represented by a single location within the multidimensional space, where each dimension maps onto either a specific parameter or global property of a face that varies from one face to another. Image based properties (such as the distance between the eyes) or more abstract properties of a face (such as trustworthiness), have all been considered possible dimensions of face-space. It is proposed that faces are normally distributed within each dimension, forming a multivariate Gaussian distribution within the space. There are two key models of face-space that differ in terms of their explanations of a face’s location within the multidimensional space. In the original account put forward by Valentine (1991) faces were represented as individual points within the space. Under this logic, faces near the origin are typical (also known as exemplar) having values near the central tendency on all dimensions. Here, faces that are similar in appearance will have a closer proximity within the face-space. The second model however suggests that faces are encoded relative to the central face at the origin of the multidimensional space, and that faces are arranged using vectors from this norm, with the vector’s parameters of length and direction determined by the distinctiveness and features of the face respectively. These variations of the face-space model have been put forward, each providing subtle differences and predictions about human face processing, that aim to account to the behavioural findings within face recognition and perception such as, distinctiveness, inversion, caricaturing and the other race effect.

A key concept within face-space models is the notion of an encoding error. When a face is encoded into the face-space there is a degree of error associated with the encoding, whereby if the encoding conditions are difficult such as brief presentation time or when faces are contrast negated, a relatively large encoding error will occur. In practice encoding error is likely to result in greater difficulty when recognising typical faces as opposed to face that are more distinctive in face space dimensions. This stems from exemplar-based face space models, where it is posited that faces that are closer to the central tendency across multiple dimensions (i.e. typical faces) are more densely clustered towards the centre of face-space. Thus, an increase in encoding error is more likely to lead to confusion and longer processing times for typical faces versus distinctive faces that have less neighbours in the face space. In contrast, for a distinctive face, when presented under conditions that provide a high encoding error (such as low-quality images), the target identity is more likely to be the nearest face in face-space. Using the example of face inversion, Valentine (1991a) found that when participants were shown inverted faces (increasing the encoding error) there was a smaller recognition impairment for distinctive faces in comparison to typical faces. This effect was seen for familiar and unfamiliar faces during a recognition memory task and was taken to reflect the notion that distinctive faces lie further from the exemplar cluster, thus supporting the assumptions of a multidimensional face space model where the centre of the space reflects the mean of each dimension.

However, as shown by Burton and Vokey (1998) the key assumption of this exemplar-based face space model being that the relationship between the distance from the centre and local density, does not hold true for a multidimensional space that contains more than two dimensions. In other words, the centre of face space does not contain the highest number of points (faces). This was shown through a series of calculations by plotting the frequency of faces as a function of squared distance from the centre with more than two dimensions. When doing this, it was observed that the majority of faces were not at the centre of the space but are rather found at some distance from the centre, whereby that distance increases as a function of the number of dimensions sampled. This observation is also consistent with the finding that most faces are not rated as highly typical (Burton & Vokey, 1998).

The face space framework described above (exemplar-based face space) was the preferred model described in later work by Valentine and Endo (1992). However, an alternative face-space model was also developed from the legacy of schema theory (Goldstein & Chance, 1980), known as the norm-based face space. Whilst similar to the exemplar-based model, it differs in that the facial similarity metric is proposed to be based upon a prototypical norm face contained at the centre of face space. Here, when a deviation from the norm is encoded, the direction of the vector from the norm face is described as more important than the magnitude of said vector (Stevenage, 1995). This is in contrast to the exemplar-based face space described above, whereby an aberration is encoded as simple-point vector. The distinctiveness of a face is represented by the length of this vector with the direction defining the identity (Valentine, Lewis, Hills, 2016). This model gained popularity over the exemplar-based face space model by being able to better account for existing behavioural effects, in comparison to the exemplar-based face space framework. It has previously been shown that there is an advantage for recognising faces that have been caricatured. Faces are found to be more recognisable when the features are exaggerated in directions away from an average face, in comparison to the veridical image (Rhodes et al., 1987). The norm-based model explains this effect under the notion that the caricature's representation subtends in the same angle of direction as the norm face but its magnitude is larger (Rhodes, 1997). However, a problem with this original explanation is that it does not account for the finding that once the caricature level surpasses a certain level, recognition becomes worse in comparison to the veridical image. For example, Rhodes et al (1987) found that exaggeration beyond 16% for line drawings leads to a reduction in recognition.

To account for these caricature effects, an adapted version of the exemplar-based model was developed- referred to as the absolute coding face space (Valentine, 1991; Byatt & Rhodes, 1998). The key effect of caricaturing is that the face representation of an identity is shifted into a region of the face space that has a lower exemplar density. The argument put forward is that this lower exemplar density leads to caricatured images being easier to recognise compared to the veridical image as there is less competition from other face representations. However, when caricatured to a higher degree the absolute distance increases much more, and the caricatured image is now encoded much further away from the stored exemplar vector of that identity, thus is now no longer seen as being that individual. Under this framework, the ease to which we recognise faces is not based on just the two nearest exemplars (exemplar-based model) but instead on all exemplars within a certain range. The parameters of this range and the manner in which exemplars interact with one another is however unspecified.

Support for this model of face space also stems from being able to account for the other race effect (ORE). This effect describes the notion that faces of a viewer's own race are recognised more accurately and quickly than faces of a different race, in addition to further perceptual advantages (O'Toole et al., 1994). In a key study by Byatt and Rhodes (1998) they compared the assumptions of both a norm-based coding and an absolute-based coding face space in explaining the ORE and caricature effects. According to a norm-based coding system, faces are represented in terms of deviations from a prototypical face, with caricatured faces being more recognisable because they exaggerate this norm deviation information and explain the ORE with other race faces being coded relative to an own race prototype. Contrastingly, an absolute-based face space suggests that faces are encoded as absolute values on a set of shared dimensions irrespective of a norm, with caricatures being effective as they minimise the exemplar density and other race faces being less recognisable as they are more tightly clustered. In this study, European participants' identification of European and Chinese faces caricatured for both race norms was tested. Here, norm-based coding, would predict that caricatures of Chinese faces made by distorting differences from the European norm would be more effective than caricatures made relative to the Chinese norm. Whereas, an absolute based framework predicts that all faces would be recognised more accurately when caricatured against their own-race norm. It was revealed that the assumptions of the latter were more supported.

Taken together, despite support for face-space models in general a number of questions still remain. For example, the models do not specify the overall number of dimensions within face space, nor is the number of image properties that might contribute to different aspects of a face posited. For instance, numerous image properties might be used to encode a facial expression and some of these components might also be utilised during the processing of other facial signals such as age, gender etc. However, with techniques such as principal components analysis-PCA (see Chapter 2, for a comprehensive review of this method) it is possible to explore multidimensional face spaces, whose dimensions (or principal components) are derived from a set of images. Thus, allowing for a data driven approach to help uncover what image properties are useful for the recognition of identity and other facial signals, and probing how faces are represented.

1.2.4 Deep Convolutional Neural Networks (DCNNs)

In more recent years, the development of deep convolution neural networks (DCNNs) has made substantial progress on the complex problem of recognising faces across variations of viewpoint, illumination and appearance (O'Toole, Castillo, Parde, Hill & Chellappa, 2018). Deep convolutional neural networks mimic the neural processing of the primate visual system to generate a face space that preserves details regarding the categorical identity of faces and relevant image characteristics. Faces are represented in a DCNN through a hierarchical extraction of features, from simple patterns in early layers to complex facial structures in deeper layers, eventually leading to a classification after the fully connected layers. DCNNs initiate their operations with raw images that first undergo feature extraction within the early convolutional layers. Here, these layers scan the input image for simple features like edges, textures, and basic shapes. In the context of faces, these layers might detect simple patterns like curves or the edges of facial features. As you move deeper into the network, the convolutional layers build a hierarchy of features. Higher layers combine low-level features to represent more complex structures, possibly capturing facial parts like eyes, nose, and mouth. After each convolutional layer there is a process of pooling. Pooling layers are often used to downsample the spatial dimensions of the features, retaining the most essential information. This helps in creating a spatial hierarchy of features and reducing computational complexity. The output from the convolutional and pooling layers is often flattened to be fed into fully connected layers. Fully connected layers analyse the high-level abstractions of the features, combining information from different parts of the image. In the case of faces, these layers might learn to recognise holistic facial patterns and configurations. For a DCNN to be successful at face recognition, it must be trained on many identities where each identity is depicted using multiple variable images.

Despite the widespread use of DCNNs as models of face processing in humans, the representation that emerges within the final layer of the DCNN is not fully understood (O'Toole, Castillo, Parde, Hill & Chellappa, 2018). Recent studies have found that that the top-layer DCNN representations across image variation, retain surprisingly accurate information about the original input image. For example, it has been found that within the top layers of a DCNN it is possible to predict the viewpoint of the face (Parde et al., 2017). This result is consistent with the electrophysiological recordings within IT of the macaque which similarly show explicit coding of viewpoint, as well as fMRI findings in the human face selective regions which show representations of viewpoint. However, it is yet to be determined how viewpoint is represented throughout the many layers of a DCNN and if these representations mirror the processes and representations seen in human observers.

1.3 The Neural Representation of Faces

Numerous techniques and methods have been used to probe and classify face selective neural populations in the brain including single-cell recordings (Baylis, Rolls, & Leonard, 1985; Leonard, Rolls, Wilson, & Baylis, 1985; Allison, Puce, Spencer & McCarthy, 1999) and methods aiming to record event related potentials (ERPs) specific to faces such as EEG and MEG (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Bötzel, Schulze, & Stodieck, 1995). However, the most commonly employed neuroimaging method is now functional magnetic resonance imaging (fMRI) which tracks the measurement of blood flow assumed to reflect neural responses in face selective neural populations in vivo in healthy participants.

The discovery of different face-selective regions with fMRI has led to the development of neural models of face perception, such as Haxby, Hoffman and Gobbini’s (2000) distributed human neural system for face perception. The model aims to provide an account of how different face regions contribute to different aspects of face perception. Here, face processing involves a hierarchical system consisting of several interconnected brain regions, including the fusiform face area (FFA), the occipital face area (OFA), and the superior temporal sulcus (STS). Each of these regions plays a distinct role in different aspects of face perception, such as facial feature analysis, holistic face processing, and the interpretation of facial expressions. They also connect to an extended network of regions that are not exclusively involved in processing faces, but are involved in processing important semantic, episodic and affective information from the face.

Haxby, Hoffman and Gobbini's (2000) model emphasises the dynamic interactions between the core and extended regions, highlighting the importance of both bottom-up (sensory-driven) and top-down (cognitive-driven) processes in face recognition. In the following sections the components of this core hierarchical network for the visual analysis of faces will be discussed with reference to our current understanding of what image properties are represented within the core brain regions that are critical to the perceptual recognition of familiar faces.

1.3.1 The Role of the Occipital Face Area (OFA)

The Occipital Face Area (OFA), forms a core part of most neural models of face processing and is the first implicated region within Haxby's (2000) model of face processing. It is suggested to be involved in the preliminary processing of facial features, generating an initial representation of a facial image. The location of the OFA being posterior to the FFA and STS implies that OFA performs an initial analysis of faced information that is then passed on to other regions for higher level facial analysis.

A number of neuroimaging studies have found that the OFA is sensitive to the image properties of the face. Rotshtein, Henson, Treves, Driver and Dolan (2005) used an fMR-adaptation paradigm to investigate the properties of the OFA. In fMR-adaptation studies, participants are presented with a face stimulus multiple times, leading to a decrease in neural response as a function of exposure (Grill-Spector & Malach 2001). When a new stimulus is introduced or manipulated the neural response then increases providing that, that region encodes the stimulus change. Rotshtein and colleagues (2005) found that adaptation in the OFA only occurred when the same image was repeated; there was a release from adaptation when the face changed in physical appearance, but retained the same identity or when the face changed in physical appearance but the identity changed. This suggests that the OFA is sensitive to the physical properties of the image. In contrast, adaptation in the FFA occurred when the face changed in physical appearance, but retained the same identity, only releasing from adaptation when the face changed in physical appearance and the identity changed. Fox, Moon, Iaria and Barton (2009), investigated the effects of changes in expression and identity on the responses of the OFA. They found that the OFA was sensitive to changes in physical appearance when structural changes occurred along both an axis of identity or expression. It was also found that this effect in the OFA was independent of task demands or the perceptual experience of the individual participant. It has long been suggested that sensitivity to spatial relations is important for successful recognition (Maurer et al., 2002), specifically, the variation between the second-order relational information. Rhodes, Michie, Hughes and Byatt (2009) reported that the OFA is sensitive to spatial encoding of facial features. Taken together, these studies support the notion that the OFA has a functional purpose regarding the initial encoding of the face before further more higher-level analyses begin (such as identity or expression analysis).

When thinking about what visual information within a face is being processed by the OFA, a number of related findings emerge. For example, the patterns of BOLD activity in the face selective regions to images of individual facial features and combinations of different features has been explored. Here, it has been reported that the OFA shows a preference for single features of the face over various combinations of features and stimulus types (objects), whereas the FFA shows equal sensitivity to both individual facial parts and combinations of features (Arcurio, Gold & James, 2012). More recently, representational similarity analysis was used to investigate what type of identity distinguishing information is encoded within the OFA when participants viewed naturalistic videos of famous faces (Tsantani et al., 2021). It was revealed that the representational distances in the OFA were mainly driven by differences in low-level image-based properties (pixel-wise and Gabor-Jet dissimilarities). Contrastingly, dissimilarities between face identities in FFA were accounted for by differences in perceived similarity of the faces, social traits and gender, but not differences in the low-level image properties of the faces. This suggests that the OFA and FFA utilise and process different types of identity information to discriminate between facial identities. Moreover, it implies that the FFA representation is further removed from the facial stimuli one is perceiving, encoding higher-level perceptual face information, in comparison to the OFA which has a stronger affinity for image specific properties. These findings strengthen the notion that the OFA is positioned spatially and functionally within a hierarchy of visual areas critical for facial processing.

1.3.2 The Role of the Fusiform Face Area (FFA)

The most well-known region in the human brain that responds preferentially to faces compared to other classes of objects is the fusiform face area (FFA). In their seminal paper, Kanwisher, McDermott and Chun (1997) showed participants face and non-face images. They found a region located in the right fusiform gyrus (FFA) that produced a significantly higher signal intensity during epochs in which faces were presented, rather than during epochs in which objects were displayed. The face selectivity of this area has been replicated across a number of viewing conditions, including; whether the face is actively or passively viewed (Berman et al., 2010) as well as the area being shown to be modulated by task difficulty (increasing degradation of images), whereby performance in the face recognition task was linearly correlated to the activation in the FFA (Bokde et al., 2005; Weibert et al., 2015).

Evidence supporting the role of the FFA during identity recognition, stems from Grill-Spector, Knouf and Kanwisher (2004) who measured the correlation between FFA activity (using fMRI) and behavioural outcomes during perceptual tasks designed to test face detection and within category identification of faces and other objects. It was found that the FFA was involved with the detection and identification of faces but showed little involvement in within-category identification of non-face objects (such as types of car and animals). Furthermore, research suggests that the FFA is invariant to the size, (Andrews & Ewbank, 2004), viewpoint (Pourtois et al., 2005; although see Andrews and Ewbank, 2004) and emotional expression (Winston, Henson, Fine-Goulden & Dolan, 2004) of faces. Taken together, these neuroimaging studies show some degree of invariance in the neural representation of the FFA.

The neural code underlying the FFA and its functions is of particular interest when discerning the role of how low-level properties of facial images may interact with identity recognition. Central to this topic is whether the FFA is sensitive to identity changes or whether it is sensitive to changes in image properties (independent of identity). Xu et al (2009), scaled the physical similarity of view changes of the same person (using Gabor-jets) to be equivalent to that produced by an identity change. They found that both identity and orientation changes led to equivalent releases from adaptation in the FFA (relative to identical faces) thus implying that the FFA is sensitive to the image properties of faces rather than identity. As stated previously, behavioural data often finds that texture information (as opposed to shape properties) underpins perceptual judgments of identity. However, within the FFA similar releases in adaptation are observed when an adapted familiar face displays a texture change or a shape change (Baseler, Andrews, Burton & Young, 2011). Thus, taken together with Xu et al (2009), it is arguable a low-level image property code, reliant on various low-level image properties may underlie identity processing within the FFA, and across the face selective network. This findings contrast with the results found from Tsantani et al (2021), who found that pixel-wise and Gabor-Jet models (low-level image properties) did not underpin the patterns of activation within the FFA, which was instead found to encoding higher-level perceptual face information. Therefore, unpacking the neural code of how faces are represented within the FFA is of current interest within the field.

1.3.3 The role of the Superior Temporal Sulcus (STS)

Located under the lateral fissure which separates the temporal, parietal and frontal lobes, the STS has been implicated to serve functions related to higher level vision as well as playing a role within the perception of social cues (Haxby et al., 2000; Akiyama et al., 2006). Within the distributed neural system for face perception (Haxby et al., 2000) the STS is suggested to process changeable aspects of a face including but not limited to the perception of eye gaze, lip movements and most notably expression changes. It is further theorised that the STS has reciprocal links to an extended system including the intraparietal cortex (for spatially directed attention), the auditory cortex (for initial speech perception) and the amygdala (for the processing of emotion from facial expression).

The STS has been shown to respond more to changes in the viewpoint of a face when identity is kept constant compared to when the identity changes (Andrews & Ewbank, 2004; Baseler et al., 2014). This implies that unlike more ventral face regions (e.g. FFA) the pSTS processes changeable aspects of faces with similar meaning across individuals, unrelated to their identity. However, there was increased functional connectivity between the pSTS and FFA when participants viewed same identity faces compared with different identity faces (when expression was varied), implying that there are distinct neural pathways involved in the analysis of identity and expression, but there is a level of interaction to process changeable aspects of the face. Similarly, Harris, Young and Andrews (2012) first found a release in adaptation in the pSTS for changes in expression but not for identity. In a follow up study, fMR-adaptation was used to investigate whether the coding of expression in the pSTS was categorical or continuous, using expression continua generated by morphing between two expressions. They found an equivalent release from adaptation for within-category compared to between-category changes in expression, suggesting that the coding of expression in the pSTS is continuous rather than categorical.

Similar to the OFA the STS has been found to show sensitivity to individual facial features regardless of typical feature configuration (Liu, Harris & Kanwisher, 2010). This sensitivity has been linked to the role the wider role of the STS during eye gaze perception (Pelphrey et al., 2005). For example, Puce, Allison, Bentin, Gore and McCarthy (1998) showed participants faces that were superimposed onto a radial background that moved inward. It was found that the STS did not show an increase in activation when the eyes were static, but did when the eyes were averted to the left or right, suggesting the STS is involved in the perception of eye gaze but not to motion in general. Furthermore, it has been shown that using deviated and frontal faces with averted and direct gaze in a combined EEG and MEG paradigm there is an interaction between gaze direction and head orientation between 134 and 162ms (in MEG) and a main effect of gaze direction between 171 and 186ms (in EEG) (Burra, Baker & George, 2017). Importantly the locus of these effects was centred in the posterior and anterior regions of the STS respectively.

1.4 Image Invariant Representations

Image invariant representations are thought to be a crucial component of face recognition. The face of a person can generate a large number of different images due to changes in pose, illumination, expression, and occlusion. Image invariant representations are thought to be crucial for capturing the invariant features of the face that are reflect identity. This robustness ensures that the recognition system can identify individuals under various viewing conditions. It is important that these representations can be used to tell different faces belong to the same identity (within-person variability) as well as telling faces apart.

Models of face perception and recognition often focus on the differences between identities and how a recognition system tells identities apart. However, they seldom comment on the natural variability of the face within an identity, often referred to as within-person variability (Burton, 2013). Empirically, research shows that the facial images belonging to one identity can often be perceived as more dissimilar than facial images of different identities, when these identities are unfamiliar to us (Jenkins, White, Van Montfort & Burton, 2011). In this study, participants were asked to freely sort two sets (one familiar and one unfamiliar) of 40 facial photos into piles based on identity. The image sets each contained only 2 identities and the images were ambient in nature, referring to the fact that they were not chosen to fulfil any experimental requirements (such as having the face images taken with the same camera at the same time). It was revealed that the average number of piles (7.5) was far greater than the correct answer (2) for the unfamiliar faces. However, the piles rarely contained different identities. This suggests that participants could tell the identities apart but often perceived that two images of the same person were too different to belong to the same identity. Thus, it is of importance to understand how within person variability is learned and adopted into an invariant representation allowing for a familiar observer to recognise an identity over a range of viewing conditions.

1.4.1 The Problem of Within-Person Variability when Learning New Faces

One recognition model which considers how within person variability can be used to create an image-invariant representation, is the averaging model (Burton, Jenkins, Hancock, & White, 2005; Jenkins & Burton, 2008). Here, it is proposed that a structural representation is created by filtering out image specific information through averaging. It has been found that average faces were better recognised by humans and face recognition algorithms than individual images of a person (Burton, Jenkins, Hancock & White, 2005). Moreover, variability in exposure has been shown to be fundamental to learning new faces, implying that an average facial representation can become more stable with a greater number of inputs (Devue & de Sena, 2023). Several studies have shown that, when learning new faces, participants are better able to recognise previously unseen faces when the learning images contain high variability (Murphy et al., 2015; Baker et al., 2017; Ritchie and Burton, 2017; Kramer et al., 2018). For example, when participants were taught to associate names with faces of unfamiliar individuals, using learning images that had a high degree of variability or a low degree of variability, participants were more accurate and quicker in verifying the names of identities they had learned with high variability compared to those learned with low variability (Ritchie & Burton, 2017). Taken together, these results are consistent with the idea that an average face representation could underpin our recognition of faces. Similarly, if this average representation is based on face images that are more representative of the natural variation of a person, then there will be an advantage in the recognition of novel views of a person.

However, Burton, Jenkins and Schweinberger (2011) suggest that this averaging model may not be sufficient. Treating facial information that is less common across a series of images of an identity as noise and thus filtering this out of a view-invariant representation, may be unhelpful when recognising a face. For example, in real life abrupt changes in appearance due to stylistic choices (e.g. dramatic haircuts or changes in surface properties e.g. tanning) do not usually impact one's ability to recognise a familiar face. This variation in different images of a face could form part of one's representation of that identity allowing for a more flexible and informative representation. Supporting this, Ritchie, Mireku and Kramer (2020), investigated the use of average images during a live face matching paradigm. They found when attempting to match a live target face to a single image, a 4-image array or a face average image, performance accuracy was comparable across all conditions. This suggests that whilst face averages have produced improvements in performance when image matching is performed using stimuli presented on a screen, this does not generalise to a real-world setting.

1.4.2 Achieving View-Invariance from View-Symmetry

Understanding the process of generating a view-invariant representation from view-specific inputs is thought to be critical for unravelling the intricacies of face recognition. It had been assumed that this process involved the convergence of multiple view-specific representations. However, a more recent hypothesis posits a two-step process for achieving view-invariance, instead of a one stage process whereby view-specific inputs converge onto a view-invariant representation. Here it is suggested that there are two stages of convergence, whereby view-specific representations first converge into view-symmetrical representations which then further convolve into a view-invariant representation of identity (Freiwald & Tsao, 2010). Support for the role of view-symmetric representations in face recognition is elucidated by behavioural studies. These studies have demonstrated that faces with symmetrical viewpoints (e.g. a full left and full right profile) exhibit greater perceptual similarity than those with non-symmetrical viewpoints (e.g. full left profile and ¾ left view). Additionally, recognition accuracy is enhanced when the test viewpoint is symmetrical with the learned viewpoint (Troje & Bulthoff, 1998; Busey & Zaki, 2004; Flack et al., 2019).

Parallel to these findings, neuroimaging studies have also revealed a similar representational hierarchy for viewpoint in face-selective regions. Firstly, fMRI studies have shown view-selective responses to faces in the OFA (Grill-Spector et al., 1999; Andrews & Ewbank, 2004; Fang et al., 2007; Carlin et al., 2011; Guntupalli et al., 2017; Weibert et al., 2018). Moreover, studies have also reported view-symmetric representations and view-invariant representations in regions such as the FFA (Axelrod & Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019). Interestingly, these view-symmetric neural responses were predicted by the perceptual similarity of faces from different viewpoints, suggesting that they might play an important role in face recognition (Flack et al., 2019).

Collectively, these findings suggest a functional hierarchy in facial representation within these regions that could underlie the process of recognition. This raises important theoretical questions regarding how image properties of faces contribute to the representation of faces that underlie the recognition of identity. For example, images of the same identity that are symmetrical in nature have a degree of difference in the image properties they possess, thus a key question remains regarding how this potential intermediate symmetrical representation is generated and subsequently utilised in the formation of a view-invariant representation.

1.5. Thesis Aims

The overarching aim of this thesis, is to further explore the role of visual information in the perception and recognition of faces. The experiments in this thesis use a combination of behavioural, computational and neuroimaging approaches to ask how facial representations are formed and what facial information is critical for the recognition and perception of faces.

In order to address what image properties underpin facial representations, it is first important to consider the process of generating a view-invariant representation that is utilised for making identity judgements. The first experimental chapter (Chapter 3) investigates how view-invariant representations are generated. A key unresolved question within the literature, is how we generate view-invariant representations from view-specific inputs. This chapter explores the possibility of having an intermediate representation of view-symmetry, prior to achieving view-invariance. To achieve this, we compared behavioural and neural responses to canonical (yaw) and noncanonical (roll) rotations of the face, to interrogate how view-invariance emerges. Finally, we measured responses to viewpoint in a deep convolutional neural network trained on faces, to examine whether humans and DCNNs process viewpoint in similar ways.

Chapter 4, builds upon the previous chapter to explore what information within a view-invariant representation is critical for familiar face recognition. As suggested previously, surface texture properties of the face are proposed as the dominant cue for recognition. However, in this chapter we explore the roles of both shape and texture information in the perception and neural representation of identity. This chapter uses hybrid faces in which the shape properties of one individual are combined with the texture properties of a different identity. Previous behavioural research investigating the contributions of shape and texture properties in recognition, have often employed tasks that can rely on perceptual matching. Here, we offer a different approach in which participants have to rely on their stored mental representations (as we do in everyday life). In the final experiment, we measured the relative sensitivity to shape and texture in face-selective regions of the human brain, to assess the similarity in the contributions of shape and texture information behaviourally and in the neural response.

Upon establishing the overall relative contributions of shape and texture properties within familiar face recognition and the neural representation of faces, Chapter 5 aims to explore what specific image properties underlie this view-invariant representation. We used a behavioural approach in combination with principal components analysis to reveal the critical image dimensions for face recognition. We did this to answer two key questions using tasks that again, only rely on the stored mental representations of faces. Firstly, we asked which image dimensions were important for familiar face recognition. Secondly, we wanted to establish using a novel face learning paradigm, are the same image dimensions also important for becoming familiar to a new identity.

Finally, experiments in Chapter 6 investigated further aspects of face processing, the perception of gaze, gender and emotional expression. It remains unclear within the literature, what image properties underlie the perception of the many signals of information available within a face, and whether the same image properties are critical for the perception of different signals. A similar approach to Chapter 5 was used, in which principal components were removed from faces that varied in either gaze, gender or expression, to explore the contributions of different combinations of image dimensions to the perception of these categories. We did this to examine whether a small set of image dimensions provide a unique or overlapping contribution to the perception of different facial signals.

Taken together, the experiments presented in this thesis aim to provide a cohesive account of how facial representations are formed, and what properties of a facial image are critical for the processes of face learning, face recognition and for aspects of face perception. Findings reported here, suggest that there are three stages of facial representation, ending with full view-invariance, sufficient for familiar face recognition. This latter representation relies on both shape and texture information from a face, with texture being the dominant property utilised. Furthermore, there is a critical band of shape and texture image dimensions that underlies both face learning and face recognition, with distinct (but overlapping to a small degree) bands of other image dimensions underlying the processing of different facial signals.

Chapter 2- General Methods Review

2.1 The fMRI BOLD Signal

Functional Magnetic Resonance Imaging (fMRI) is a widely popular technique aimed to measure the neural responses in the brain by tracking the changes in blood flow that are then associated with neural activity. As neurons increase their rate of fire, their energy reserves deplete and need to be replenished. Consequently, there is an increase in the transfer of oxygen to those neurons through the bloodstream leading to local changes in blood oxygenation. It has been shown that Blood Oxygenation Level-Dependent (BOLD) change causes a measurable change in local magnetic signal which can be detected by MRI (Ogawa, Lee, Kay & Tank, 1990), and used to infer the underlying brain response. Despite fMRI being an indirect measure of brain responses limited by the temporal resolution of the haemodynamic response, changes in the BOLD response do correlate well with changes in local field potentials and action potentials (measured using multi-unit activity) which occur within milliseconds (Logothetis, Pauls, Augath, Trinath & Oeltermann, 2001). Moreover, with fMRI having a high spatial resolution allowing for the resolution of differences in brain response within millimetres (voxel level), the measurement of the BOLD signal is a fundamental technique that can be analysed using a number of statistical techniques, including univariate and multivariate analyses.

2.2 Univariate Analysis

2.2.1 Cognitive Subtraction

In order to analyse the BOLD signal, classical fMRI analyses typically employ a univariate general linear model (GLM) approach. A set of regressors are used to model the brain responses within voxels or a region of interest (ROI) to different stimulus conditions. A box-car model can then be used to predict zero response when the stimulus is absent and a non-zero response when the stimulus condition is present. This model can then be convolved with the hemodynamic response function (HRF) to produce an expected timeseries response that considers the non-abrupt changes seen in blood flow. Regressing this model against the collected fMRI BOLD signal on a voxel by voxel basis, results in a whole-brain statistical map of regression coefficients (parameter estimates) which reflect the fit of the model at the voxel level. In this way, larger coefficient values reflect voxels that are activated by the stimulus condition; conversely, voxels that are not are not responsive to a stimulus will be predicted poorly and assigned a smaller coefficient value. This process can be seen in Figure 2.1.

This analysis is typically performed on multiple stimulus conditions and can be used to infer the underlying properties of the brain or ROIs. For instance, parameter estimates for different experimental stimulus conditions (e.g. a control/baseline) can be contrasted against one another. In this way, the statistical significance at each voxel can be estimated (e.g. as a p-value or a z-score) and be used to infer the underlying response (Huettel, Song, & McCarthy, 2004).

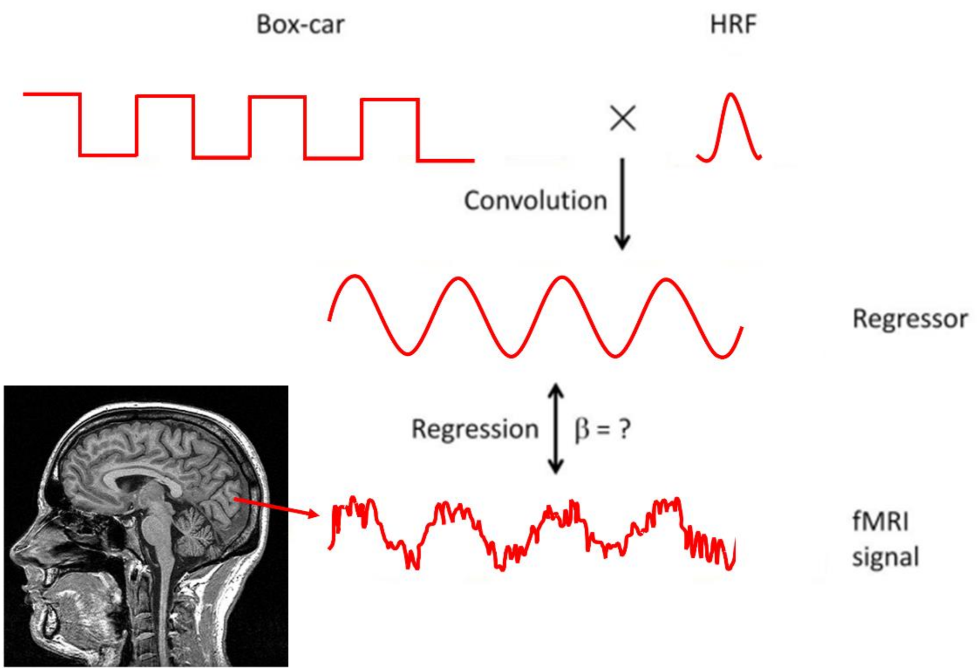

Figure 2.1. *Univariate GLM analysis. A box-car function is defined that corresponds to the timing of the stimulus presentation. A hemodynamic-response function is used to convolve the box-car generating a hemodynamic regressor, which can then be regressed against the fMRI signal independently for each voxel.*

Whilst the univariate GLM approach is an informative technique allowing for the analysis of the BOLD signal there are limitations to this method. One such limitation is that when using GLM contrasts, only a single signal is obtained per voxel, however even voxels of 1mm^3 contain several hundred thousand neurons. This, in tandem with the assumption that responses which deviate significantly from zero reflect stimulus related neural activity, can lead to less informative and less complex analysis. For instance, differences in response within a single voxel may either reflect response of all neurons within that population or greater change in only a subset of neurons. One resolution to differentiate these potential underlying sources of the BOLD signal is fMR-adaptation.

Similarly, coherent patterns of neural response may be observed across multiple voxels which may include voxels showing both super- and sub-threshold positive and negative responses. Net neural patterns such as these may be found to differ reliably between stimulus conditions. In such cases it may even be that the aggregate response across voxels in a given brain region is near zero, but crucially this does not mean that this region does not contain information about the stimulus. Standard univariate analyses such as GLM, however, will not be sensitive to information represented in distributed neural patterns. It is for these reasons that it is now becoming ubiquitous to employ a variety of analyses including the use of multivariate methods within neuroimaging research, aiming to capture distributed neural patterns of activity.

2.2.2 fMR Adaptation

fMR adaptation is based on the principle that repeated stimulus presentation habituates the neurons which are responsive to that type of stimuli. However, if the stimulus or its properties are changed, only neurons which are sensitive to the change will recover from adaptation and increase their response. What this technique allows us to do is address whether or not in one voxel we have two populations of neurons coding different stimuli types (or stimulus properties), or whether we have one population of neurons coding both stimuli types (or stimulus properties) (Grill-Spector & Malach 2001).

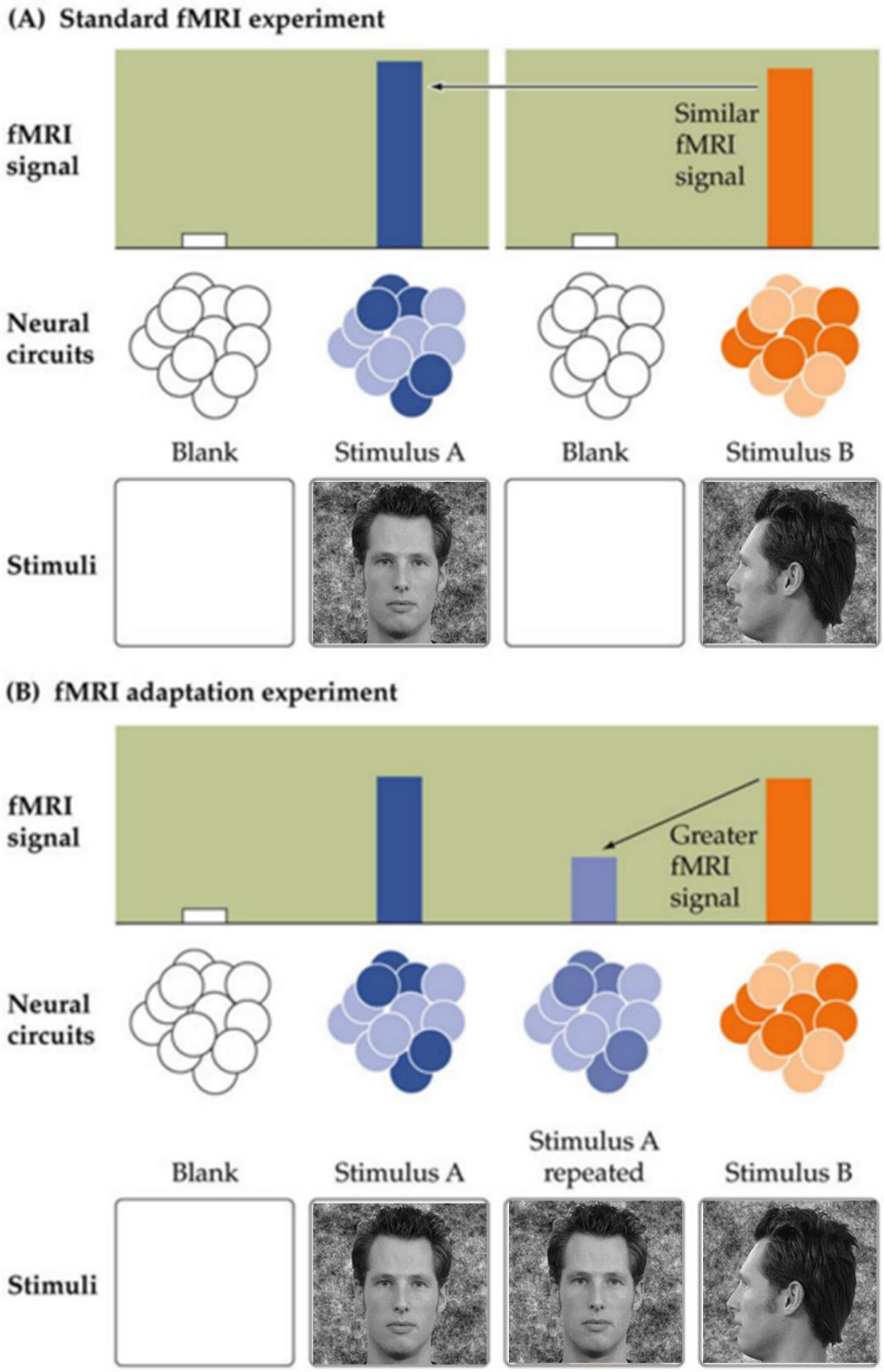

Figure 2.2. *Examples of a standard fMRI experiment (A) versus an fMR adaptation experiment (B). Adapted from Principles in Cognitive Neuroscience, Box 15B. Purves et al., (2013). Faces taken from the Radboud Face Database- Langner et al., (2010).*

In a typical fMRI experiment (Figure 2.2A) we could have a voxel which contains a population of neurons which are selective for a particular viewpoint of a face (e.g. a frontal view). If a participant is shown an image of a face that is at this viewpoint, then you will generate a response from these neurons. Within this voxel, there could also be a population of neurons selective for another viewpoint of a face (e.g. left profile). If the participant is then shown an image depicting a left profile of a face, this population of neurons could give a similar response to that of a front on face. Using a standard contrast approach, we would compare the response of the front view and the left profile. This would give rise to no difference. One interpretation is that there is no selectivity to viewpoint. However, this would hide the real interpretation which is that there are equal numbers of neurons selective to frontal and left profile faces.

However, in an adaptation paradigm (Figure 2.2B), repeated presentations of one viewpoint (frontal) cause the neural response to decrease because the neurons selective to frontal view to habituate. However, if a new viewpoint (e.g. left profile), the response increases because there is a separate population of neurons which are selective for this and have not been habituated. Therefore, using an adaptation paradigm means the underlying neural coding for (in this example) the processing of facial viewpoints, can be more directly addressed.

2.3 Multivariate Pattern Analysis (MVPA)

To combat the limitations of traditional univariate GLM based analyses, the use of multivariate methods allows for the patterns of response across multiple voxels to be interrogated simultaneously. One such method is multi-voxel pattern analysis (MVPA) which can be advantageous when two conditions may produce similar overall mean responses within a ROI, but crucially have different underlying neural patterns. In this case, the difference between the conditions would be masked using a GLM univariate analysis due to contrasting within an ROI leading to no significant differences. Therefore, analysing the patterns of neural response across voxels within an ROI can reveal differences (if present) between conditions.

2.3.1 Correlational MVPA

A fairly straightforward form of MVPA is the correlational method which was applied in the original Haxby et al (2001) study, to demonstrate that neural patterns of response can be used to distinguish between object categories. To perform this method the data is first split into odd and even runs of the stimulus presentation for each condition (known as cross-validation). Next, parameter estimates are generated for each condition, independently for each run using a univariate GLM analysis in each voxel within the ROI. The data is then normalised to reduce the shared variance between across the conditions. For instance, if the stimulus conditions were facial expressions (e.g. happy and sad faces), the parameter estimates generated for these conditions are expected to contain shared variance that is explained by all conditions using faces, attentional effects and generic responses to visual stimulation. Through a process of normalisation such as subtracting the mean response across all experimental conditions, from each individual condition on a voxel by voxel basis, we leave behind the pattern of response that is exclusive to the experimental conditions. Pairwise correlations can then be calculated between the neural response patterns for each possible combination of conditions across the splits of the data. This is performed for both within-condition comparisons (e.g. happy faces-even with happy faces-odd) and between-condition comparisons (e.g. happy faces-even with sad faces-odd). MVPA outputs are often represented in correlational matrices, with the prediction that if response patterns can distinguish the stimulus conditions then, the within-condition correlations will be higher than the corresponding between-condition correlations. Haxby et al (2001), revealed higher within than between-category correlations for multiple object classes, indicating that different object categories could be discriminated through distinct patterns of response. Crucially, when this analysis was restricted to only those voxels which did not show a selective univariate response to any condition, the results remained. Thus, it was argued that while standard univariate analysis did not produce distinct brain responses to different categories, multivariate analysis was sensitive enough to reveal category-specific brain response patterns.

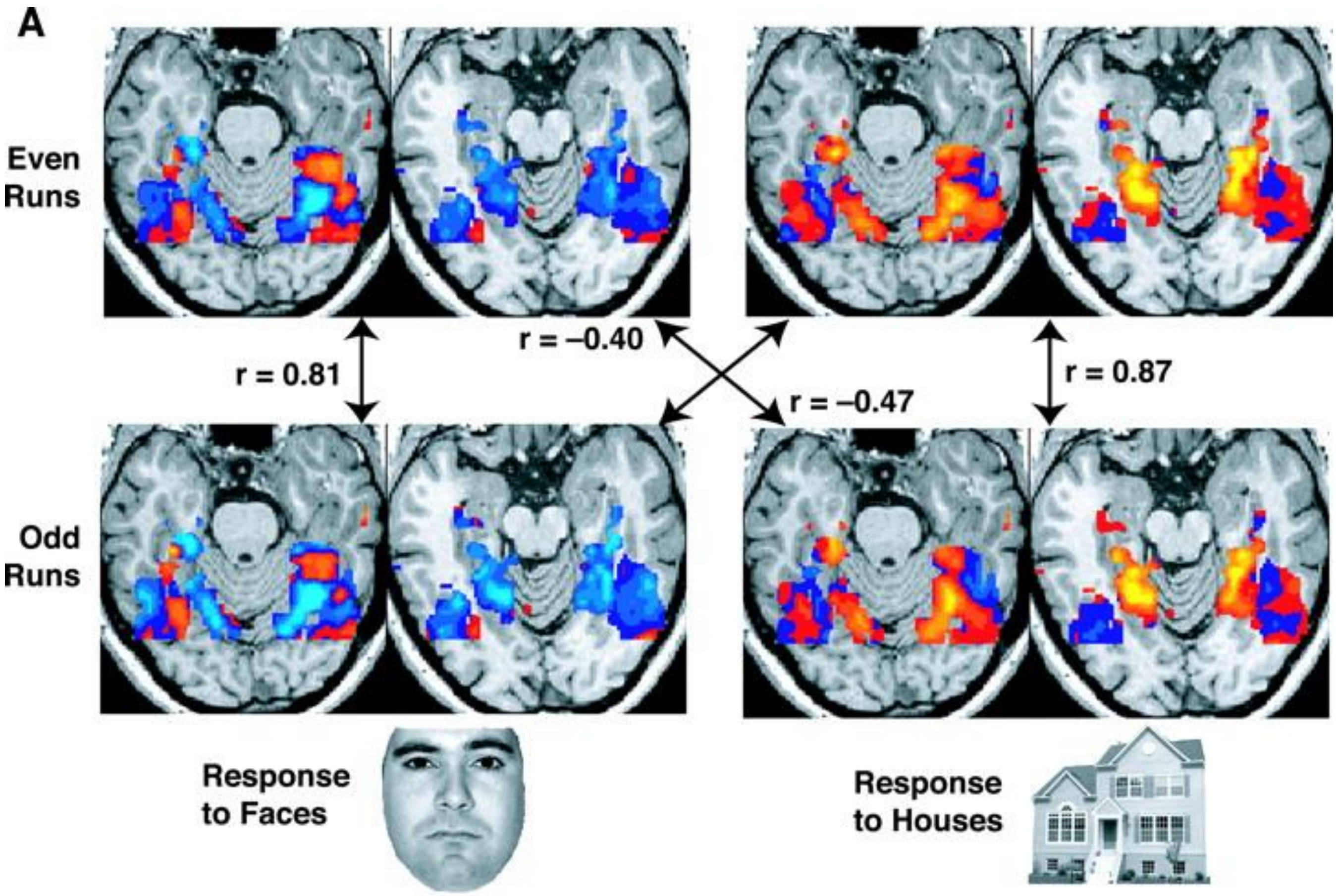

Figure 2.3. *Schematic of a correlation-based MVPA paradigm from Haxby et al., (2001). Neural response patterns were estimated for faces and houses (stimulus conditions), for even and odd stimulus runs. Patterns are restricted to a region of interest, and correlated pairwise within and between- conditions across the data splits.*

2.3.2 Leave One Participant Out (LOPO) MVPA

The majority of MVPA studies investigate patterns of response within participants. This is based on the assumption that patterns of neural response are to some degree idiosyncratic. However, other studies have utilised a Leave-One Participant Out (LOPO) MVPA paradigm (Shinkareva et al., 2008; Poldrack, Halchenko & Hanson, 2009; Kaplan & Meyer, 2012). This LOPO paradigm has the advantage of being able to determine the consistency of responses across individuals. Instead of cross validating neural response patterns across runs within participants, it iteratively compares the neural response to all stimulus conditions between an individual participant and the rest of the group. To perform this method, a group average of the neural response to each condition is created (minus one participant), it is then compared to the neural response profile in the individual whose data has been left out of the group analysis for those same conditions. This analysis is then iteratively repeated once for each participant. The outcome of this yields a number of correlation coefficients for each condition comparison equal to the number of participants. Significance testing can then be carried out in the same way as correlational MVPA (Haxby et al., 2001), by comparing the correlational coefficients for neural responses to within vs between category conditions. If a stimulus category generates a distinct pattern of activity, then the within-condition correlations for the individual participant and rest of the group should be higher than the between-condition correlations.

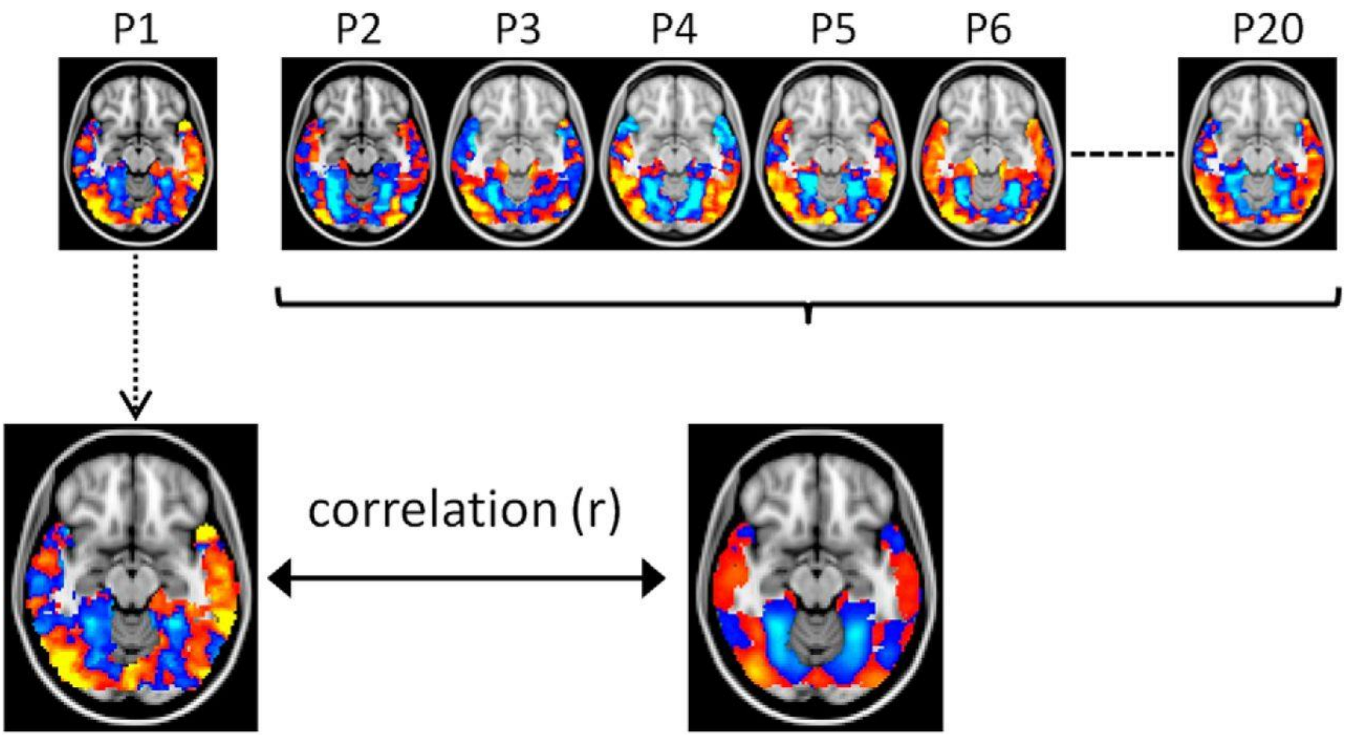

Figure 2.4. *Schematic diagram of a LOPO MVPA paradigm. Group analyses compare individual patterns of response with the group pattern of response derived from all participants except that individual, this process is then repeated across all LOPO iterations for all conditions.*

2.4 Principal Components Analysis (PCA)

In addition to neuroimaging techniques, other statistical methods have been used in face perception and recognition- such as principal components analysis. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically (Jung et al., 2002). This is the case when there is a strong correlation between observed variables. The central idea of using PCA for face recognition is to express the large 1-D vector of pixels constructed from 2-D facial images into the compact principal components of the feature space. This can be called eigenspace projection. Eigenspace is calculated by identifying the eigenvectors of the covariance matrix derived from a set of facial images (vectors).

Using this technique, it has been possible to model a number of perceptual facial properties, including face distinctiveness, the other race effect and emotional expressions (Calder, Burton, Miller, Young, & Akamatsu, 2001; Hancock, Bruce, & Burton, 1998; Hancock, Burton, & Bruce, 1996; O'Toole, Deffenbacher, Valentin, & Abdi, 1994). What makes this technique advantageous is its capacity to represent variability of multidimensional data in far fewer dimensions (known as eigenfaces). Originally shown by Kirby and Sirovich (1990) as well as Turk and Pentland (1991), it was revealed that face images can be reconstructed using as few as 50 eigenfaces compared to many thousands required in a pixel-by-pixel representation (Burton, Bruce, & Hancock, 1999). Another advantage of this technique is that many faces can be entered into the PCA analysis. Thus, natural variation across faces can be approximated and represent our pre-existing experience with faces in real life by using images that are ambient in nature- capturing variability across age, pose, lighting conditions, emotional expressions, image quality, and ethnicity. Similarly, varying numbers of images of different identities can be entered into the analysis in order to simulate different levels of familiarity.

Eigenfaces are created once a set of face images is subjected to PCA. To do this, the face images due to be entered are first resampled to a common pixel resolution ($r \times c$). Each image is then treated as one vector, simply by concatenating the rows of pixels in the original image, resulting in a single column with $r \times c$ elements. For this implementation, it is assumed that all images of the set are stored in a single matrix T, where each column of the matrix is an image. The images are mean centred by subtracting the mean image from each image vector. The eigenvectors and eigenvalues of the covariance matrix are then calculated. Each eigenvector has the same dimensionality (number of components) as the original images, and thus can itself be seen as an image. The eigenvectors of this covariance matrix are referred to as eigenfaces (principle components). They are the directions in which the images differ from the mean image i.e. the directions of the new feature space, and the eigenvalues determine their magnitude, in other words they explain the variance of the data along the new feature axis. These eigenvalues are often sorted so that the first principal components explain the most variance within an image set. Information is lost by projecting the image on a subset of the eigenvectors, but losses are minimized by keeping those eigenfaces with the largest eigenvalues. For instance, working with a 100 × 100 image will produce 10,000 eigenvectors. In practical applications, most faces can typically be identified using a projection of between 50 and 150 eigenfaces, meaning that most of the 10,000 eigenvectors can be discarded.

A common practice when using PCA with faces is to perform shape normalisation aiming to separate face shape and face texture. There are a number of different ways of operationalising the distinction between shape and texture properties to allow them to be manipulated quasi-independently. For example, 3D laser scans can be used to derive the 3D structure of the human face and can be overlaid with a two-dimensionally defined texture map (O'Toole, Price, Vetter, Bartlett and Blanz 1999). However, more commonly, shape and surface texture properties are extracted in a 2D manner using a process of delineation. This process often examines the second order configural properties of face shape, those that are defined by the spatial layout of the features. This process requires a standard grid that is positioned over each face and altered to map out the key landmark fiducial points (e.g. corners of the mouth and eyes). To then separate the shape and texture of the face, the images are morphed into a standard face shape (typically the average shape of all of the faces in the set) generating each image into a shape-free face space, which are then used for the texture PCA. The shape component, therefore, codes the original position of the points in the grid while the texture component codes the pixel intensities in its standardised shape. After this separation, PCA is applied to independently to the images containing just the shape and texture of the faces, leading to images that are then assigned a unique set of shape and texture coefficients that describe the principle components (eigenfaces) that represent the image in this lower-dimensional space.

Chapter 3- The emergence of view-symmetry for faces in humans and deep networks

This Chapter is adapted from: Rogers, D., & Andrews, T. J. (2022). The emergence of view-symmetric neural responses to familiar and unfamiliar faces. *Neuropsychologia*, *172*, 108275.

3.1 Abstract

Successful recognition of familiar faces is thought to depend on the ability to integrate view-dependent representations of a face into a view-invariant representation. It has been proposed that a key intermediate step in achieving view invariance is the representation of symmetrical views. However, key unresolved questions remain, such as whether these representations are specific for naturally occurring changes in viewpoint and whether view-symmetric representations exist for familiar faces. To address these issues, we compared behavioural and neural responses to canonical (yaw) and noncanonical (roll) rotations of the face. Similarity judgements revealed that symmetrical viewpoints were perceived to be more similar than non-symmetrical viewpoints for both canonical and non-canonical rotations. Next, we measured patterns of neural response from early to higher level regions of visual cortex. Early visual areas showed a view-dependent representation for natural or canonical rotations of the face, such that the similarity between patterns of response were related to the difference in rotation. View-symmetric patterns of neural response to canonically rotated faces emerged in higher visual areas, particularly in face-selective regions. The emergence of a view-symmetric representation from a view-dependent representation for canonical rotations of the face was also evident for familiar faces, suggesting that view-symmetry is an important intermediate step in generating view-invariant representations. Next, we measured neural responses to non-canonical rotations of the face. We found that view-symmetric patterns of response were also evident in face-selective regions. However, in contrast to canonical rotations of the face, these view-symmetric responses did not arise from an initial view-dependent representation in early visual areas. This suggests differences in the way that view-symmetrical representations emerge with canonical or non-canonical rotations. The similarity in the neural response to canonical views of familiar and unfamiliar

faces in the core face network suggests that the neural correlates of familiarity emerge at later stages of processing. Finally, we measured responses to viewpoint in a deep convolutional neural network trained on faces (VGG-Face). We found view-specific responses in the convolutional layers. View-symmetric responses emerged in the fully connected layers and were predicted by behavioural responses. Together, these findings provide new insights into the importance of view symmetry in human and artificial neural networks.

3.2 Introduction

Recognising the identity of a familiar face is a simple and relatively effortless process for most human observers. However, the appearance a face can change dramatically as a person moves their head. The visual system must ignore these sources of variation in order to recognise identity, yet at the same time be able to process these changes because of their role in social communication (Bruce & Young, 1986; Andrews & Ewbank, 2004; Baseler et al., 2014). The challenge of familiar face recognition is demonstrated by the difficulty in the recognition of unfamiliar faces when they are seen from different views (Bruce, 1982; Hancock et al., 2000; Longmore et al., 2008). Cognitive models of face perception suggest that a solution to the problem of familiar facial recognition is through view-invariant representations (Bruce & Young, 1986; Young & Burton, 2017). The successful generation of view-invariant representations relies on variable input, and experience with multiple facial viewpoints (Bruce, 2017).

How view-invariant representations are generated from view-specific representations is critical to understand how we recognise faces. A simple model for how a view-invariant representation could emerge involves the convergence of multiple view-dependent representations in a single step. However, a more recent suggestion is that the process of view-invariance occurs by a two-step process that involves the convergence of view-dependent representations into view-symmetrical representations and then the convergence of these view-symmetric responses into view-invariant representations (Freiwald & Tsao, 2010). Behavioural support for the role of view-symmetric representations in face recognition comes from studies that have shown that the perceptual similarity of faces with symmetrical viewpoints (e.g. two profiles) is greater than for non-symmetrical viewpoints (e.g. profile and ¾ view) and also by studies that have shown that recognition judgements are more accurate when the test viewpoint is symmetrical to the learnt viewpoint (Troje & Bulthoff, 1998; Busey & Zaki, 2004; Flack et al., 2019).

Neurophysiological studies provide further support for a model of face recognition that is initially view-specific, with an intermediate view-symmetric representation before view-invariance emerges. For example, different face-selective neurons in the temporal lobe of nonhuman primates have been shown to be selective for single views, symmetric views and invariant to changes in view (Perrett et al., 1991). Studies using fMRI guided neurophysiological recordings in non-human primates show that face regions at early stages of processing have a more view-specific representation, with intermediate face regions showing more view-symmetric responses and later face regions showing more view-invariance (Friewald & Tsao, 2010). Taken together these results imply that there is a functional hierarchy of facial representation within these regions that could underpin recognition. Interestingly, these symmetrical responses were evident to naturally occurring or canonical rotations of the head (yaw - left to right changes in viewpoint), as well as to less common or non-canonical rotations (roll) that occur as a result of within-plane rotations of the image. The demonstration of view symmetrical responses to non-canonical rotations is intriguing, as it suggests that these responses might reflect a more general response to symmetry in the visual brain (Bertamini et al., 2018), rather than something that is directly linked to face recognition.

Neuroimaging studies have also revealed a similar representational hierarchy for viewpoint in face-selective regions. fMRI studies have shown view-selective responses to faces in the OFA (Grill-Spector et al., 1999; Andrews & Ewbank, 2004; Fang et al., 2007; Carlin et al., 2011; Guntupalli et al., 2017; Weibert et al., 2018). However, other studies have also reported view-symmetric representations in regions such as the FFA (Axelrod & Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019). Interestingly, these view-symmetric neural responses are predicted by the perceptual similarity of faces from different viewpoints, suggesting that they might play an important role in face recognition (Flack et al., 2019).

There are two important limitations of previous neuroimaging studies in humans. The first is that these studies have only used naturally occurring viewpoint changes. So, it remains unknown whether view symmetric neural responses are also evident to more unnatural, noncanonical rotations of the face (such as in plane rotation), as has been reported in neurophysiological studies (see Friewald & Tsao, 2010). If view-symmetrical responses to faces were found for these rotations, it could be argued that they reflect a general property of visual cortex, rather than being directly linked to generating view-invariant representations for face recognition. The second limitation is that all previous neuroimaging studies have used unfamiliar faces (Axelrod & Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019). It is not clear therefore whether view-symmetric responses are also evident for familiar faces. Demonstrating view-symmetric responses to familiar faces would provide further evidence for the role of these representations as an intermediate step toward view-invariant representations.

In more recent years, the development of deep convolution neural networks (DCNNs) has made substantial progress on the complex problem of recognising faces across variations of viewpoint, illumination and appearance (O'Toole, Castillo, Parde, Hill, Chellappa, 2018). Despite the development of DCNNs, the nature of the visual representation that emerges from the fully connected layers is not fully understood. For example, what properties of the image are still evident in the fully connected layers. Interestingly, a recent study has shown that, although the fully-connected layers of the DCNN are able to identity face identity across image variation, they also contain information about the image, such as viewpoint (Parde et al., 2017). However, it is yet to be determined if mirror-symmetric representations are also present.

The aim of this study was to explore view-symmetry using behavioural measures, neuroimaging techniques and artificial neural networks. First, we wanted to determine if view-symmetric representations are evident for both canonical and non-canonical rotations of the face. If mirror-symmetric representations are an important process that precedes the generation of a view-invariant representation, then it might be expected that the view-symmetric responses would only be evident for naturally occurring rotations of the face. To address this question, we compared the perceptual similarity to faces that were rotated canonically (left/right rotations of the head, yaw) with faces that were rotated noncanonically within the plane of the image (within plane rotation, roll). As a further test of whether view symmetry is important for recognition, we measured perceptual similarity of familiar faces for which view-invariant representations are thought to exist. Second, to determine how patterns of viewpoint selectivity response emerge in visual cortex, we measured the pattern of response in early visual areas and in face-selective regions. Our hypothesis was that face images will initially be represented by a view-dependent (image based) representation in early visual areas, but that view-symmetric representations will emerge in higher-level face regions. Third, we addressed how viewpoint is represented within a DCNN that has been trained on millions of face images for the purpose of recognition. Here we wanted to ask whether view-symmetric representations are a feature of artificial neural networks and whether these representations are similar to those seen in human observers during fMRI. To do this, we compared the outputs of pairs of images with different viewpoints to ask whether view-symmetric responses are evident in the DCNN and in what layers of the network they emerge. Finally, we asked whether behavioural responses (perceptual similarity ratings for pairs of images) in humans could predict patterns of response in the DCNN addressing the question of whether humans and DCNNs compute view symmetry in similar ways.

3.3 Methods

3.3.1 Participants

Participants were recruited separately for the behavioural and fMRI experiments (Experiment 1: n = 38, female = 26, mean age = 23.6 years, SD = 5.58; Experiment 2: n = 25, female = 14, mean age = 23.5 years, SD = 6.87; Experiment 3: n = 69, female = 43 , mean age = 21.3 years , SD = 3.27). All participants had normal or corrected to normal vision and were drawn from an opportunity sample of students and staff at the University of York. All participants gave their written informed consent. The study was approved by the Psychology Department Ethics Committee and the York Neuroimaging Centre Ethics Committee.

3.3.2 Stimuli

Figures 3.1 and 3.2 show the stimulus sets that were used for the behavioural and neuroimaging experiments, as well as the DCNN analysis. The stimuli were either familiar (well-known celebrities) or unfamiliar faces. There were three conditions: (1) canonical-familiar, (2) canonical-unfamiliar and (3) noncanonical-unfamiliar. The unfamiliar faces were taken from the Radboud Faces Database (Langner et al., 2010). The familiar images were taken from five celebrities popular to a UK student demographic (Angelina Jolie, Brad Pitt, George Clooney, Jennifer Aniston, Taylor Swift). Naturally occurring changes in view for the familiar and unfamiliar faces are shown in Figures 3.1 and 3.2. These images show canonical rotations (yaw) of the head at approximately -90°, -45°, 0°, 45°, 90°. The symmetric views for the unfamiliar faces were created by different cameras being set up at precise angles and all photos being taken simultaneously (Langner et al., 2010). View-symmetric images of familiar faces were created by taking the mirror image of each viewpoint. Otherwise, it would have been impossible to get symmetric views with similar appearance. Non-canonical views were generated by taking the frontal view of each unfamiliar face and rotating it in the frontal plane (roll) by 45° and 90° to the left and right (Figure 3.3). All face images were superimposed on a 1/f amplitude mask and scaled to 500 × 500 pixels, to ensure that all images stimulated the same amount of the visual field despite changes in viewpoint and rotation.

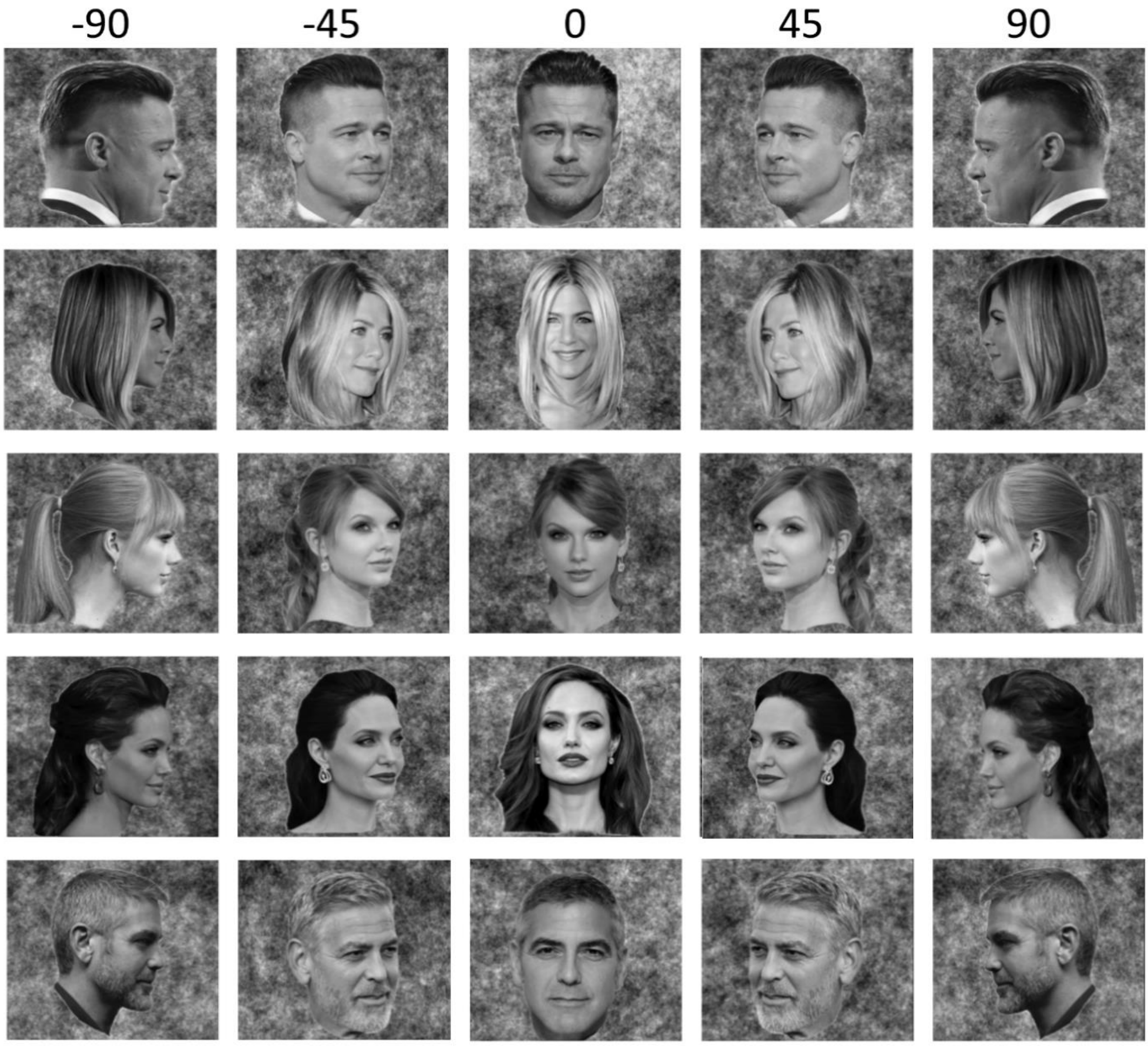

Figure 3.1. *Exemplars from the canonical-familiar condition. Images from 5 different viewpoints are show in columns. Images from 5 familiar identities are shown in rows.*

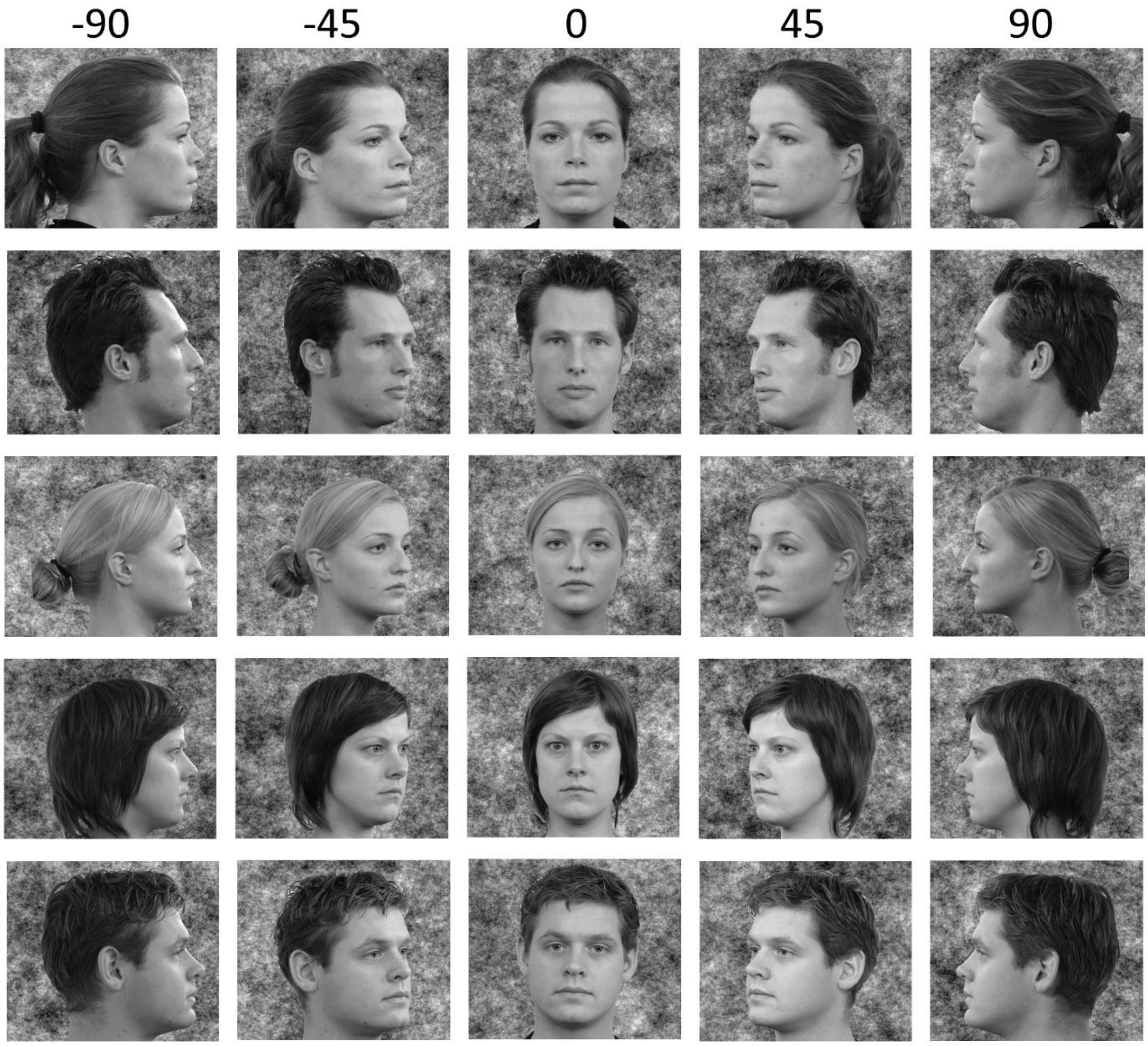

Figure 3.2. Exemplars from the canonical-unfamiliar condition. Images from 5 different viewpoints are show in columns. Images from 5 unfamiliar identities are shown in rows.

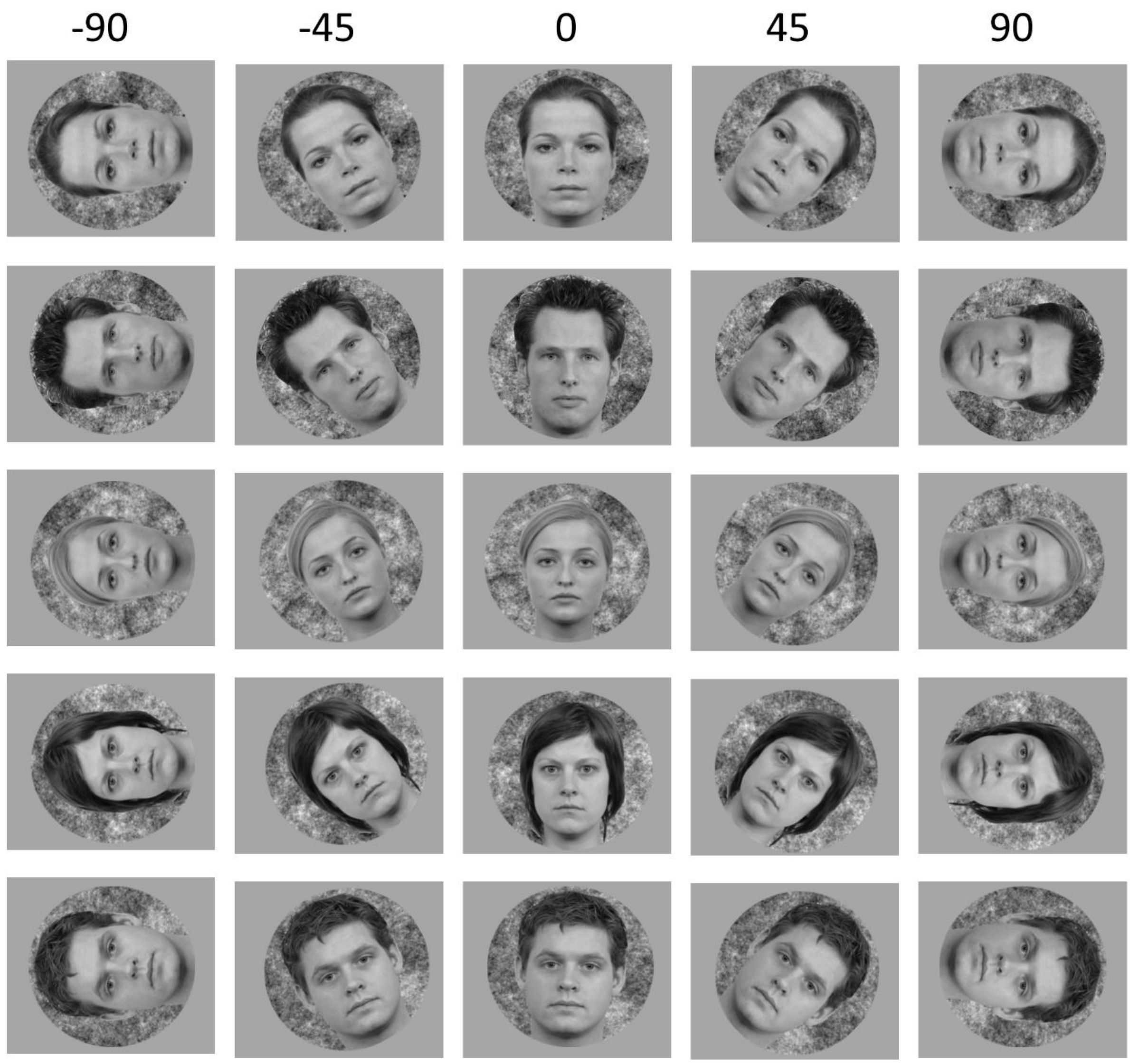

Figure 3.3. *Exemplars from the non-canonical-unfamiliar condition. Images from 5 different viewpoints are show in columns. Images from 5 unfamiliar identities are shown in rows.*

3.4 Experiment 1- Behavioural responses to face view

3.4.1 Methods

To determine whether symmetrical viewpoints were perceived as being more similar than non-symmetrical viewpoints, participants were asked to rate the perceptual similarity of pairs of faces that differed in view (canonical-familiar, canonical-unfamiliar; non-canonical-unfamiliar). Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Each trial began with a white fixation cross superimposed on a grey background for 0.5s. This was followed by a pair of faces (from the same identity) that were presented for 3s. Each view was presented with every other view (10 combinations), there were 5 identities for each of the 3 image sets giving a total of 150 trials. Images subtended approximately 8◦ of visual angle. The order of trials was randomised for each individual participant. Participants were required to respond with a button press indicating how similar they perceived the images to be, on a scale of 1–7 (1 being less similar and 7 being more similar). Participants had an unlimited time to respond.

3.4.2 Results

Participants made perceptual similarity judgements between pairs of faces with different viewpoints. Figure 3.4 shows the average ratings for symmetrical (e.g. − 90◦ & 90◦) and asymmetrical (e.g. − 90◦ & − 45◦) face pairs for each condition (canonical-familiar, canonical-unfamiliar & noncanonical-orientation). An ANOVA with Symmetry (symmetrical, non-symmetrical) and Condition (canonical-familiar; canonical-unfamiliar; noncanonical-unfamiliar) showed a main effect of symmetry ($F(1,37) = 260.52$, $p < .001$) and condition ($F(1.31, 48.48) = 12.99$, $p < .001$) as well as an interaction between symmetry and condition ($F(2,74) = 18.83$, $p < .001$). The effect of symmetry was a result of symmetrical views being more similar than non-symmetrical views for both canonical (familiar $t(37) = 14.10$, $p < .001$); unfamiliar: ($t(37) = 6.78$, $p < .001$) and non-canonical ($t(37) = 8.88$, $p < .001$) rotations of the face. The interaction between symmetry and condition was due to a greater difference between symmetrical and asymmetrical viewpoints in the canonical-familiar condition compared to both the canonical-unfamiliar ($t(37) = 4.67$, $p < .001$, $d = 0.76$) and the non-canonical-unfamiliar ($t(37) = 5.30$, $p < .001$, $d = 0.86$). Importantly, for symmetrical viewpoint combinations there was no difference in perceptual similarity between the canonical-familiar and canonical-unfamiliar conditions ($t(37) = 1.64$, $p = .054$). These findings show a perceptual similarity advantage for symmetrical views is evident for both canonical and non-canonical rotations and is also evident for familiar faces.

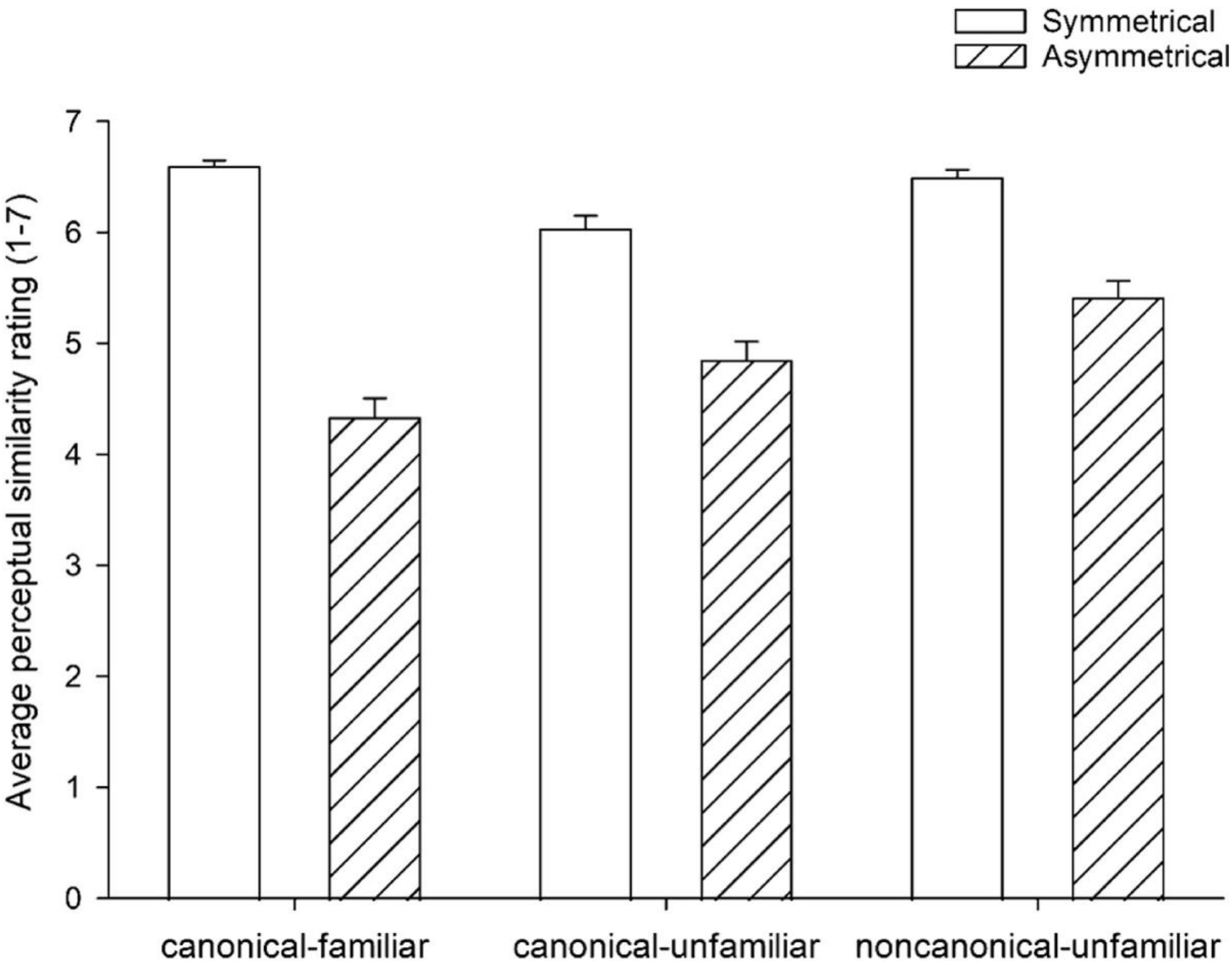

Figure 3.4. *Average perceptual similarity ratings of symmetrical and asymmetrical viewpoints for each condition. For each condition, symmetrical viewpoints were rated as being more similar than asymmetrical viewpoints. Error bars indicate SEM.*

3.5 Experiment 2 - Neural responses to face view

3.5.1 Methods

The main fMRI experimental scans used a block design with 5 different stimulus conditions, each depicting a different rotation (− 90◦, − 45◦, 0◦, 45◦, 90◦). Images from the different conditions (canonical-familiar, canonical-unfamiliar; non-canonical-unfamiliar) were shown in separate scans. In each scan, the 5 images corresponding to each viewpoint (columns in Figures 3.1-3) were shown in 6s blocks. Within each block, each image was presented for 1s followed by a 200ms grey screen. A 9s fixation screen was presented between each block. There were 5 views and each was shown 6 times during the scan, giving a total of 30 blocks. The order of the blocks was pseudorandomised across the scan. Images subtended a retinal angle of approximately 15◦ and were viewed on a screen at the rear of the scanner via a mirror placed immediately above the participant's head. Participants maintained attention during the scans by fixating on a cross in the centre of the images and indicating using a response box when they saw a green cross. Accuracy on this task was very high during each scan (canonical-familiar: 98.3% SD 2.46, canonical-unfamiliar: 97.7% SD 4.35, non-canonical-unfamiliar: 98.3% SD 2.74).

All imaging data was collected using a GE 3 T HD Excite MRI system with an eight-channel phased array head coil tuned to 127.4MHz, at the York Neuroimaging Centre (YNiC), University of York. A T1-weighted structural MRI image (1 × 1.13 × 1.13 mm voxel) was collected and a gradient-echo EPI was used to collect the functional images. A gradient echo EPI sequence with a radio-frequency coil tuned to 127.4MHz was used to acquire 38 contiguous axial slices (TR = 3s, TE = 25ms, flip angle = 90◦, FOV = 260mm, matrix size = 128 × 128, slice thickness = 3mm, voxel size: 2.25 × 2.25 × 3mm) in a bottom-up interleaved acquisition.

Data were analysed with FEAT version 5.0.9 (http://www.fmrib.ox. ac.uk/fsl). The first 9s (3 vol) from each scan were discarded, and MCFLIRT motion correction, spatial smoothing (Gaussian, FWHM 5 mm), and temporal high-pass filtering (cutoff 0.0093Hz) were applied. The BOLD response for each condition was modelled with a boxcar function convolved with a standard haemodynamic response function. To understand how the representation of facial viewpoint might change from early to higher levels of the visual system, we used ROIs based on probabilistic visual-field maps (Wang et al., 2015). These visual-field maps were generated using standardised retinotopic mapping, utilising colour and luminance varying flickering checkerboards, on a large sample of participants (N = 53). These maps were then validated on a separate group of participants to determine the probability of the ROI being in that location. Overall, we investigated 12 ROIs in each hemisphere giving a total of 24 independent regions. The analysis extracted mean percentage signal changes within the given ROI for each cope (condition) for each of the functional scans. We also used the core face-selective regions (FFA, STS, OFA). Face specific regions were defined at the same size (500 voxels), to allow the MVPA analyses to have comparable potential power to detect underlying patterns of response in each region. A group analysis was performed across participants comparing the response to unfamiliar faces compared to baseline. Using masks from a previous study (Flack et al., 2019), we identified the most face-selective voxel for each ROI from the group analysis. ROIs were then created using a flood fill algorithm that progressively selected voxels with the highest face-selectivity until the mask reached 500 voxels in size (Weibert et al., 2018; Flack et al., 2019).

Pattern analyses were performed using the PyMVPA toolbox (http://www.pymvpa.org/; Hanke et al., 2009). Parameter estimates from a univariate analysis of the main experiment were first normalized by subtracting the average response across the five viewpoint conditions (− 90◦, − 45◦, 0◦, 45◦, 90◦). The reliabilities of the neural patterns of response were then determined using a modified form of the correlation-based MVPA method devised by Haxby et al (2001), in which patterns of response from each participant were compared with the patterns resulting from the group analysis with that participant left out. This leave one participant out (LOPO) method allowed us to determine the consistency of the patterns of response across participants by measuring how similar each participant's responses were to those for the rest of the group (Rice et al., 2014; Watson et al., 2014; Coggan et al., 2016; Weibert et al., 2018). The group pattern was derived by entering all but one of the participants' data into a higher-level group analysis (mixed effects; FLAME, http://www.fmrib.ox.ac.uk/fsl). This group pattern of response for each condition was then correlated with the pattern from the participant who was omitted from the group. For each unique pair of conditions, the LOPO method was repeated 25 times, with a different participant being omitted from the rest of the group each time. A Fisher's Z-transformation was then applied to the correlations before statistical analysis. To assess whether there were reliable responses to each view we compared the within-condition and between-condition correlations.

Next, a representational similarity analysis (Kriegeskorte et al., 2008; Flack et al., 2019) was performed to determine how information regarding facial viewpoint was represented across the ROIs using a Viewpoint and a Symmetry model. In the Viewpoint model the value of each cell was proportional to the degree of difference in rotation between views. In the Symmetry model, cells showing symmetrical viewpoints were given a value of 1 (e.g. − 90; 90) and non-symmetrical viewpoints were given a value 0. To prevent differences in the overall magnitude of within-condition and between-condition correlations artificially inflating differences in correlations between matrices, our analysis was only performed on the between-cluster comparisons. All models were normalized using a Z-transform (mean 0, SD 1) which was then inputted into a linear regression analysis, with the outcomes defined as the correlation matrices obtained from the MVPA concatenated across LOPO iterations. For each model, elements within the matrix were extracted and flattened to a vector. These vectors were then repeated and tiled to match the number of participants. For each participant, correlation matrices were extracted and flattened to a vector. These vectors were then concatenated and entered into the model as the outcome variable. This analysis yielded a regression coefficient and an error that reflected variance across participants. All regression analyses included a constant term. From this analysis, it was possible to determine the relative fit to each model in each ROI. To determine how the representations emerged throughout visual cortex, we compared the regression coefficients for each model across different ROIs. Statistical values were corrected for multiple comparisons using Bonferroni-Holm.

3.5.2 Results

We measured the effect of viewpoint on the neural response to faces using LOPO MVPA. To determine the reliability of the patterns of response to different viewpoints, we first compared same-viewpoint similarity with between-viewpoint similarity. Higher same-view compared to different-view correlations shows that the patterns of response were reliable. Reliable patterns of response were evident across most regions of interest (Table 3.1). This demonstrates consistency in the patterns of response across participants to different viewpoints of faces (see also Weibert et al., 2018; Flack et al., 2019).

Next, we asked whether the patterns of response in each region were better explained by a view-dependent model (in which the similarity in the patterns of response to different viewpoints is explained by the difference in rotation) or by a view-symmetric model (in which symmetric viewpoints elicit more similar patterns of response compared to asymmetric viewpoints).

Table 3.1. *Same-view versus different-view comparisons for each condition across all ROIs. Distinct patterns of response were demonstrated by higher within-viewpoint correlations compared with between-viewpoint correlations. ***p < .001, **p < .01, *p < .05.*

		Canonical familiar		Canonical unfamiliar		Noncanonical unfamiliar	
		t	p	t	p	t	p
V1	Left	5.21	***	6.07	***	6.70	***
	Right	6.78	***	6.82	***	4.68	***
V2	Left	3.58	**	7.13	***	5.33	***
	Right	3.71	**	4.03	***	2.56	*
V3	Left	4.09	***	5.63	***	4.92	***
	Right	7.52	***	5.51	***	6.46	***
V3a	Left	2.46	*	1.11	0.298	2.78	*
	Right	3.89	***	3.63	**	5.42	***
V3b	Left	1.63	0.116	1.19	0.247	3.57	**
	Right	2.17	*	2.46	*	2.17	*
V4	Left	3.82	***	4.76	***	5.02	***
	Right	3.44	**	3.87	***	4.65	***
VO1	Left	1.48	0.152	1.99	0.058	4.74	***
	Right	3.21	**	2.69	*	1.33	0.198
VO2	Left	2.44	*	2.48	*	2.30	*
	Right	3.45	**	3.31	**	2.06	*
PH1	Left	0.55	0.589	2.14	*	1.32	0.200
	Right	3.53	**	1.57	0.130	0.22	0.831
PH2	Left	3.80	***	1.36	0.187	3.62	**
	Right	3.69	**	4.05	***	1.66	0.110
LO1	Left	2.35	*	3.21	**	5.82	***
	Right	4.85	***	3.97	***	4.60	***
LO2	Left	5.34	***	3.50	**	5.25	***
	Right	3.39	**	2.06	*	5.05	***
OFA	Left	6.10	***	9.26	***	5.09	***
	Right	7.54	***	8.58	***	7.59	***
FFA	Left	3.61	***	5.88	***	9.73	***
	Right	4.97	***	8.33	***	8.01	***
STS	Left	5.81	***	5.99	***	9.56	***
	Right	4.88	***	6.14	***	5.77	***

First, we measured patterns of response to different viewpoints in the canonical familiar condition. Figure 3.5 (A) shows how the view-dependent model predicts patterns of response across different regions. The data shows that regression coefficients for the view-dependent model were highest in early visual areas, but then decreased in higher visual areas. Figure 3.5 (B) shows how the view-symmetric model predicts patterns of response across different regions. In contrast to the view-dependent model, regression coefficients for the view-symmetric model were lowest in early visual areas, but increased in higher visual areas, particularly the face-selective regions. These findings suggest the emergence of a view-symmetric representations from an initial view-dependent representation. To quantify the transition from a view-dependent to a view-symmetric representation, the regression coefficients were correlated across the different models. There was a significant negative correlation in both the left (r = -0.72, p = .003) and right (r = -0.91, p < .001) hemisphere.

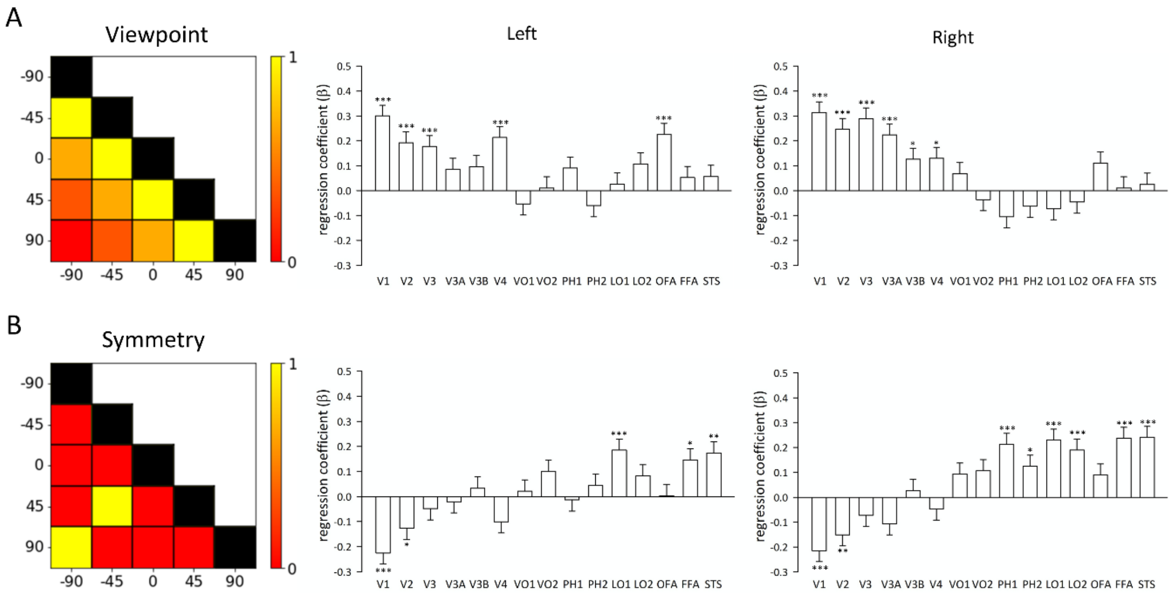

Figure 3.5. *Regression analysis of fMRI data for canonical-familiar condition showing how different models predict patterns of response across ROIs. (A) The Viewpoint model predicted patterns of response in early visual areas. (B) In contrast, the Symmetry model predicted patterns in higher visual areas, including the face-selective regions. ***p < 0.001, **p < 0.01, *p < 0.05.*

Next, we measured patterns of response in the canonical unfamiliar condition. Similar to the pattern for familiar faces, there were high regression coefficients for the view-dependent model in early visual areas, but lower values in higher visual areas (Figure 3.6A). Regression coefficients in the view-symmetric model (Figure 3.6B) were lowest in early visual areas, but increased in higher visual areas, particularly the face-selective regions. Again, these findings show the emergence of a view-symmetric representations from an initial view-dependent representation of faces. To quantify this change from view-dependent to view-symmetric patterns of response, the regression coefficients were correlated across the two models. Similar to the familiar faces, there was a significant negative correlation in both the left (r = -0.63, *p* = .012) and right (r = -0.68, *p* = .005) hemispheres.

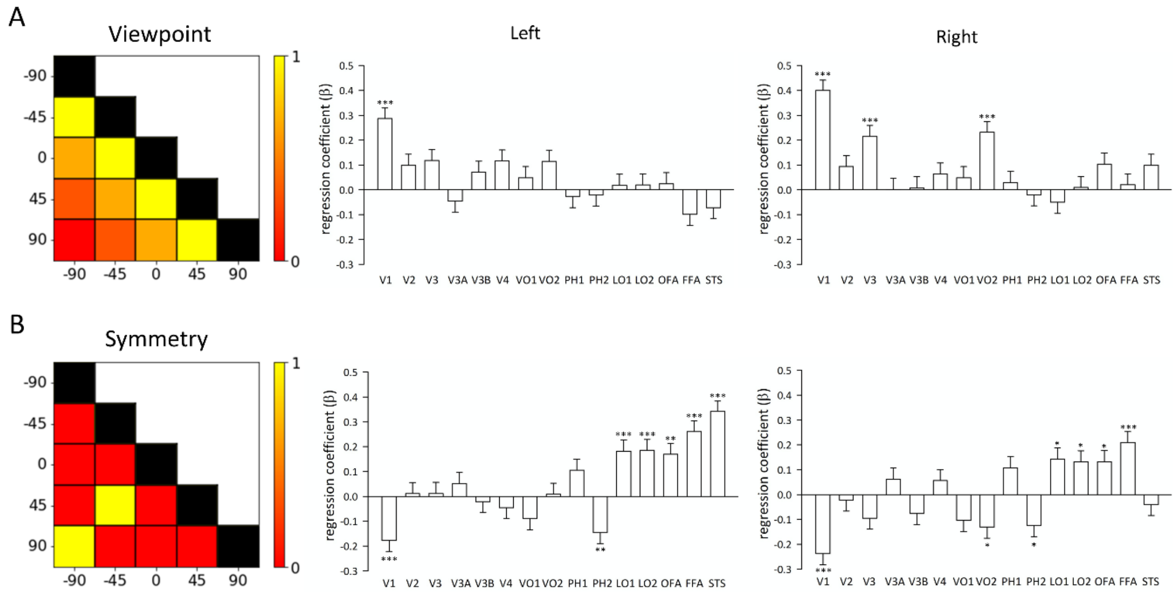

Figure 3.6. *Regression analysis of fMRI data for canonical-unfamiliar condition showing how different models predict patterns of response across ROIs. (A) The Viewpoint model predicted patterns of response in early visual areas. (B) In contrast, the Symmetry model predicted patterns in higher visual areas, including the face-selective regions. ***p < 0.001, **p < 0.01, *p < 0.05.*

Finally, we measured patterns of response to viewpoints in the non-canonical-unfamiliar condition (Figure 3.7). In contrast to canonical rotations of the face, regression coefficients for the view-dependent model were low, with only V1 (left hemisphere) having a significant positive regression coefficient and there was no obvious change in the magnitude of regression coefficients from early to higher visual areas. The regression coefficients for the Symmetry model were, however, significant in many of the higher visual areas. Although these findings demonstrate the existence of view-symmetric representations for non-canonical rotations, this does not appear to emerge from an initial view-dependent representation. This is also shown by the lack of correlation between regression coefficients across the two models in either the left (r = 0.15, *p* = .589) or right (r = -0.35, *p* = .205) hemispheres.

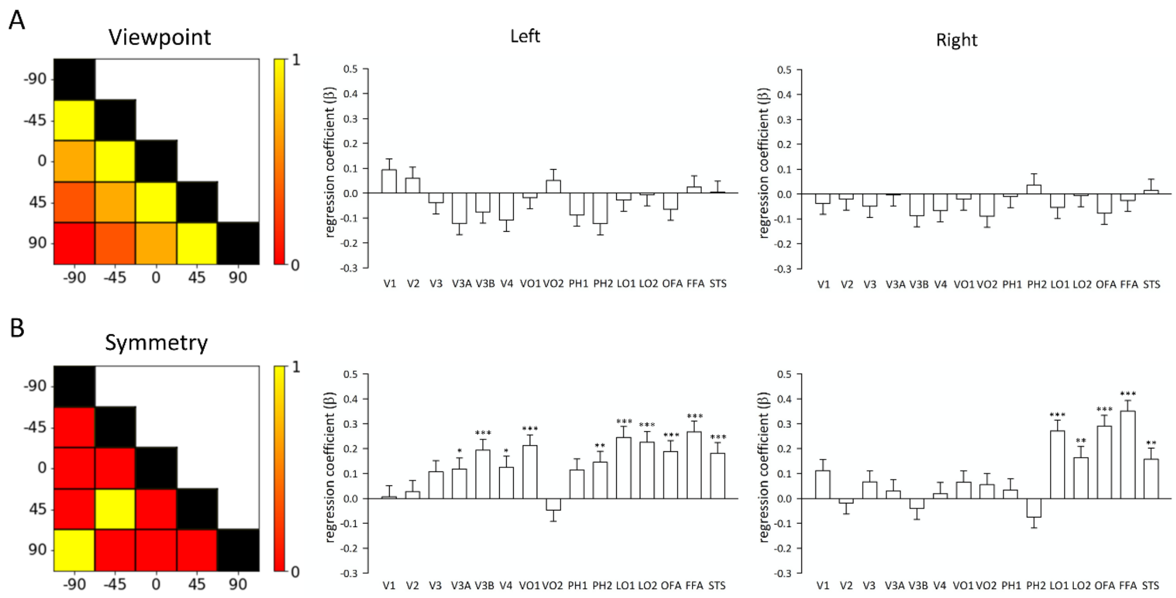

Figure 3.7. *Regression analysis of fMRI data for non-canonical-unfamiliar condition showing how different models predict patterns of response across ROIs. (A) The Viewpoint model failed to predict patterns of response across early and higher visual areas. (B) In contrast, the Symmetry model predicted patterns in higher level visual areas, including the face-selective regions. ***p < 0.001, **p < 0.01, *p < 0.05.*

3.6 Experiment 3 – DCNN responses to face view

3.6.1 Methods

In Experiment 3 we were interested in exploring whether view symmetry would emerge within a deep neural network and whether this would mirror similar patterns found in MRI within humans. To do this we used the VGG-Face DCNN (Parkhi, Vedaldi & Zisserman, 2015) to compare the similarity of face images. This DCNN consists of 13 convolutional layers and 3 fully connected (Fc) layers. The input to the network is images of size 224 x 224 pixels; images are cropped to a square bounding box centred on the face and rescaled to this resolution. Each convolutional layer is followed by one or more non-linear layers, such as rectified linear units or max pooling, which were not used in this analysis. The dimensions of the layers are as follows: Conv1 = 224 x 224 x 64 = 3,211,264; Conv2 = 112 x 112 x 128 = 1,605,632; Conv3 = 56 x 56 x 256 = 802,816; Conv4 = 28 x 28 x 512 = 401,408; Conv5 = 14 x 14 x 512 = 100,352; Fc6 = 4096; Fc7 = 4096; Fc8 = 1000. The DCNN was trained on over 2.6M face images from over 2.6K identities. Face recognition on the Labelled Faces in Wild dataset (Huang et al., 2008) and YouTube Faces (Wolf et al., 2011) for VGG-Face is 99.9% and 97.4%, respectively.

Here we used the five unfamiliar identities (Figure 3.2) each depicted by five different canonical facial viewpoints (− 90◦, − 45◦, 0◦, 45◦, 90◦) as an input for the DCNN. We compared the outputs of each layer of the neural network for pairs of images from different viewpoints (both within and between identity) to ask whether view-symmetric responses are evident in the DCNN and if so, in which layer(s) do they emerge. To do this, we first correlated the output for each combination of viewpoint within each layer of the DCNN, to generate 16 correlation matrices (representing the 16 layers of the DCNN). We then used a multiple regression analysis in which we compared the ability of two models (View-specific & View-symmetric) to predict the patterns shown between two viewpoints in each layer of the DCNN, to see how well each model explained the variance between each combination of viewpoint for each network layer. We did this analysis separately for within identity and between identity viewpoint comparisons, in order to explore how identity and viewpoint were represented within a DCNN.

We next wanted to address whether human ratings of perceptual similarity of viewpoints could be used to predict the outputs of a deep neural network, addressing the question of whether humans and DCNNs compute view symmetry in similar ways. To do this, we replicated and extended the perceptual similarity experiment (Experiment 1) with a different group of participants. The key differences in the design between this experiment and Experiment 1, was that we compared the perceptual similarity of viewpoints both within and between identities, and recorded response times.

Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2021.1). Each trial began with a white fixation cross superimposed on a grey background for 0.5s. This was followed by a pair of faces that were presented sequentially each for 2s. In this experiment we used the same five unfamiliar identities as Experiment 1, each depicted by five different viewpoints, giving rise to 25 unique images. There was a total of 300 unique image pair comparisons (25*24/2). Images subtended approximately 8◦ of visual angle. The order of trials was randomised for each individual participant. Participants were required to respond with a button press indicating how similar they perceived the images to be, on a scale of 1–7 (1 being less similar and 7 being more similar). Participants were instructed to make this judgment as fast and as accurately as possible.

To assess whether human perceptual similarity ratings could be used to predict the outputs of a DCNN, we correlated the perceptual similarity ratings for each combination of viewpoints with the output of the DCNN for that combination of viewpoint (for each layer of the network). We did this separately for both the perceptual similarity ratings as well as the response times. Here we only looked at the within identity viewpoint comparisons.

3.6.2 Results

To explore if view symmetry is also a shared feature of a DCNN, we used a simple correlational analysis to compare the outputs of the DCNN (for each layer) for every unique combination of viewpoints. We did this separately for the within (Figure 3.8B) and between identity (Figure 3.8C) comparisons, importantly we collapsed the data into averages for each viewpoint combination.

In order to quantify if view symmetry emerges within a DCNN and if so in what layer, we used a multiple regression analysis, in which we compared the ability of the two models to predict the patterns shown in different layers of the DCNN (analogous to the fMRI analysis computed in Experiment 2). These models can be seen in Figure 3.8A.

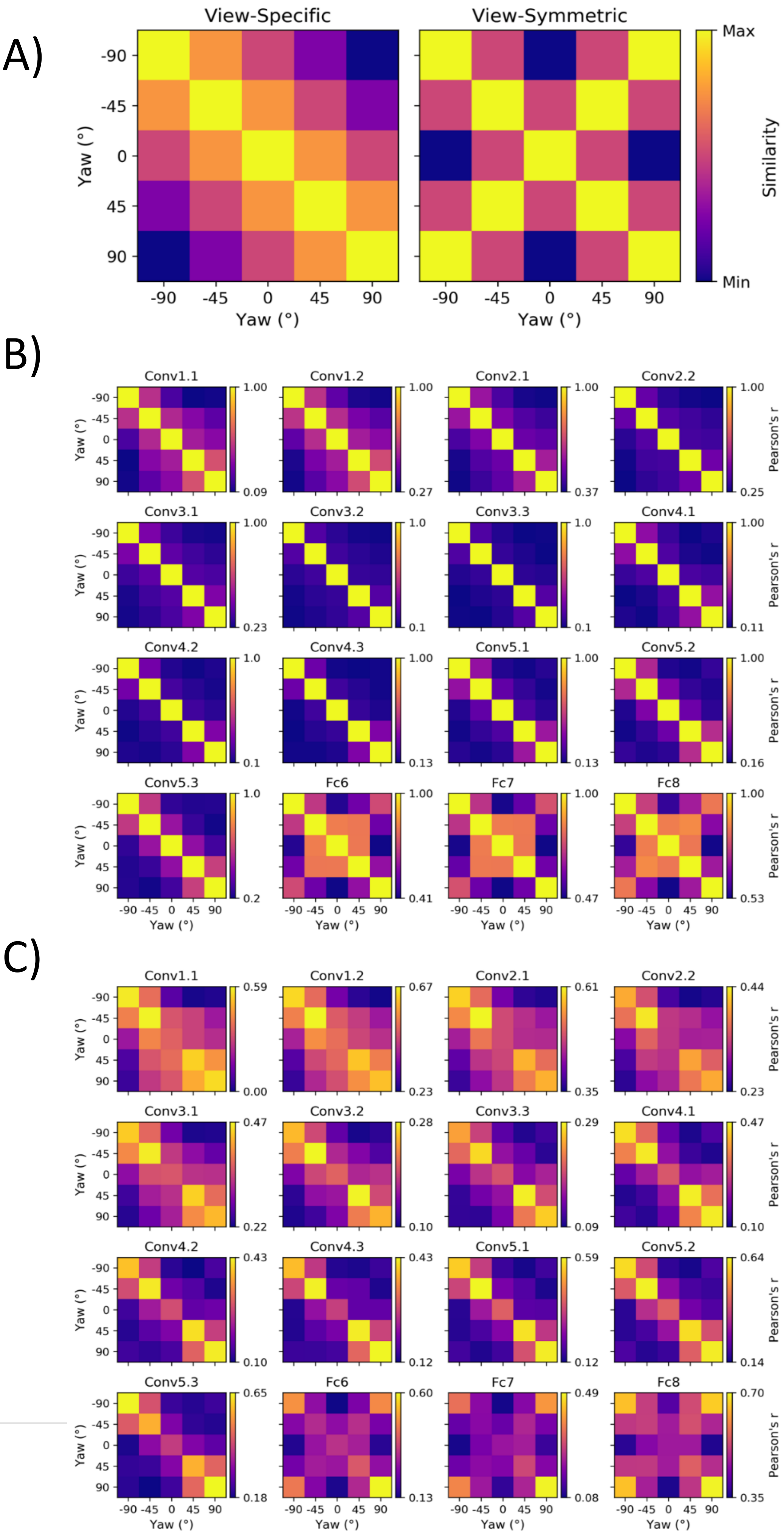
A)
View-Specific
View-Symmetric
Max
Min
Similarity
Yaw (°)
-90
-45
0
45
90
B)
Conv1.1
1.00
0.09
Conv1.2
1.00
0.27
Conv2.1
1.00
0.37
Conv2.2
1.00
0.25
Conv3.1
1.00
0.23
Conv3.2
1.0
0.1
Conv3.3
1.0
0.1
Conv4.1
1.00
0.11
Conv4.2
1.0
0.1
Conv4.3
1.00
0.13
Conv5.1
1.00
0.13
Conv5.2
1.00
0.16
Conv5.3
1.0
0.2
Fc6
1.00
0.41
Fc7
1.00
0.47
Fc8
1.00
0.53
Pearson's r
C)
Conv1.1
0.59
0.00
Conv1.2
0.67
0.23
Conv2.1
0.61
0.35
Conv2.2
0.44
0.23
Conv3.1
0.47
0.22
Conv3.2
0.28
0.10
Conv3.3
0.29
0.09
Conv4.1
0.47
0.10
Conv4.2
0.43
0.10
Conv4.3
0.43
0.12
Conv5.1
0.59
0.12
Conv5.2
0.64
0.14
Conv5.3
0.65
0.18
Fc6
0.60
0.13
Fc7
0.49
0.08
Fc8
0.70
0.35
Pearson's r

Figure 3.8. *A- schematic diagram of the View specific and View symmetric models showing the expected similarity for all viewpoint combinations, in the output of a DCNN layer if that layer represents single viewpoints or symmetry (respectively). B- shows the similarity in outputs for each combination of viewpoint within each layer of the DCNN for the within identity comparisons. C- shows the similarity in outputs for each combination of viewpoint within each layer of the DCNN for the between identity comparisons.*

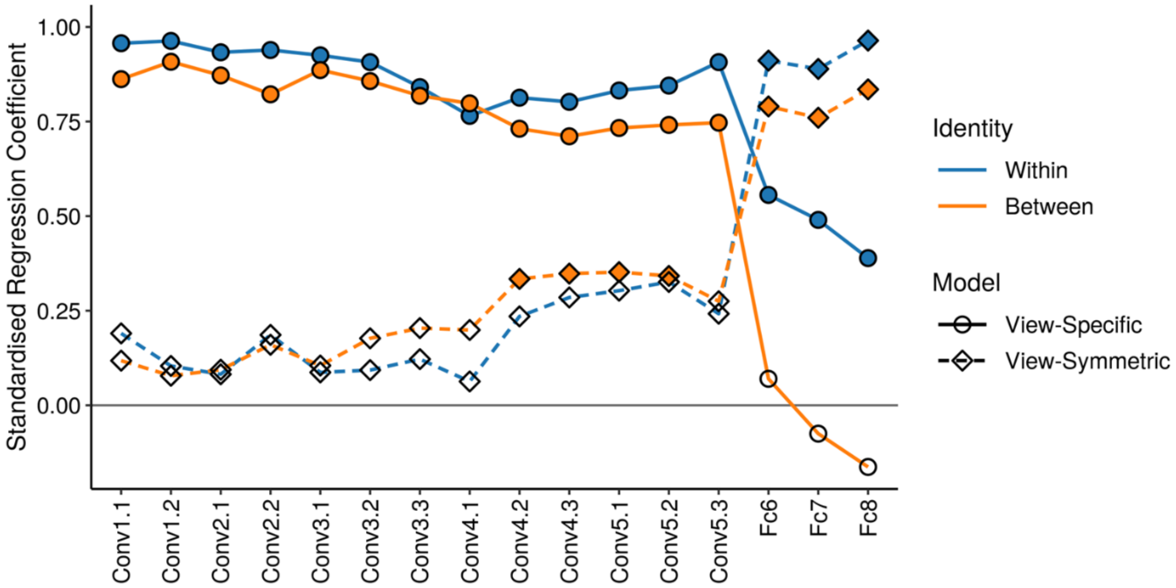

Figure 3.9. *Regression analysis for the DCNN outputs, showing the ability of the View-specific and View-symmetric models at predicting the outputs of each layer of the network, separated into within identity (shown in blue) and between identity (shown in orange) comparisons. Significant results are depicted by the filled symbols.*

The results of the regression analysis can be seen in Figure 3.9. They show that the View-specific model was significantly able to predict the outputs for both the within and between identity viewpoint combinations, within the convolutional layers of the neural network. While the View-specific model was able to predict the patterns of response for the convolutional layers (in particular the earliest layers), there was limited evidence of view-symmetry. Instead, the View-symmetric model was better able to predict the patterns of response in the output of the fully connected layers (for both within and between identity comparisons). This shows that view-symmetry is also a shared feature of deep neural networks and that this emerges in the fully connected layers, where more abstract representations such as identity have been found. Interestingly, the View-specific model was also able to predict the outputs of the fully connected layers of the network, but only for the within identity viewpoint comparisons, which is consistent with the representation of face identity within the fully connected layers.

To address the question of whether humans and DCNNs compute view symmetry in similar ways, we asked whether human ratings of perceptual similarity of viewpoints could be used to predict the outputs of a deep neural network. Figure 3.10 shows the correlation matrix for the perceptual similarity responses (A) and the response time (B) for the different viewpoint combinations, broken down across the different image combinations. We then correlated these matrices the DCNN outputs for each layer of the network

To assess if perceptual similarity ratings and response times could predict the outputs for each layer of the network, we correlated the perceptual measures (rating and response time) with the output of each layer of the neural network. This can be seen in Figure 3.11. Perceptual similarity ratings were moderately able to predict the outcome of the earliest layers of the DCNN, but were very highly correlated with the fully connected layers of the network. Similarly, the response time measure to make the perceptual similarity judgment was also significantly correlated, but only with the fully connected layer outputs. It is important to note that here the negative values for response time reflect the fact that shorter response times indicated better performance. These findings suggest that perceptual similarity ratings given by human observers can be used to predict the outcomes of a deep neural network, and that this occurs mostly for the fully connected layers. This could suggest that humans and artificial deep neural networks may share similar representations of viewpoints, and that the process of facial representation generation is solved in a similar way to humans within a DCNN.

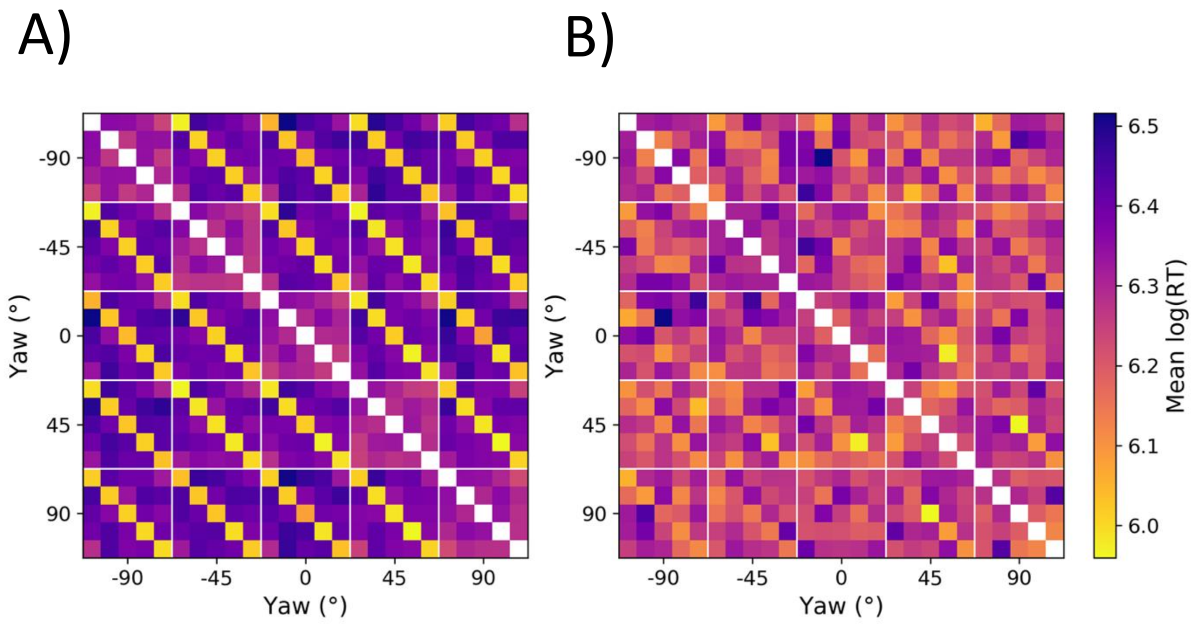

Figure 3.10. *A- Correlation matrix showing the correlations between the perceptual similarity ratings for each combination of viewpoint, broken down across the image combinations. B- Correlation matrix showing the correlations between the response times for each combination of viewpoint, broken down across the image combinations.*

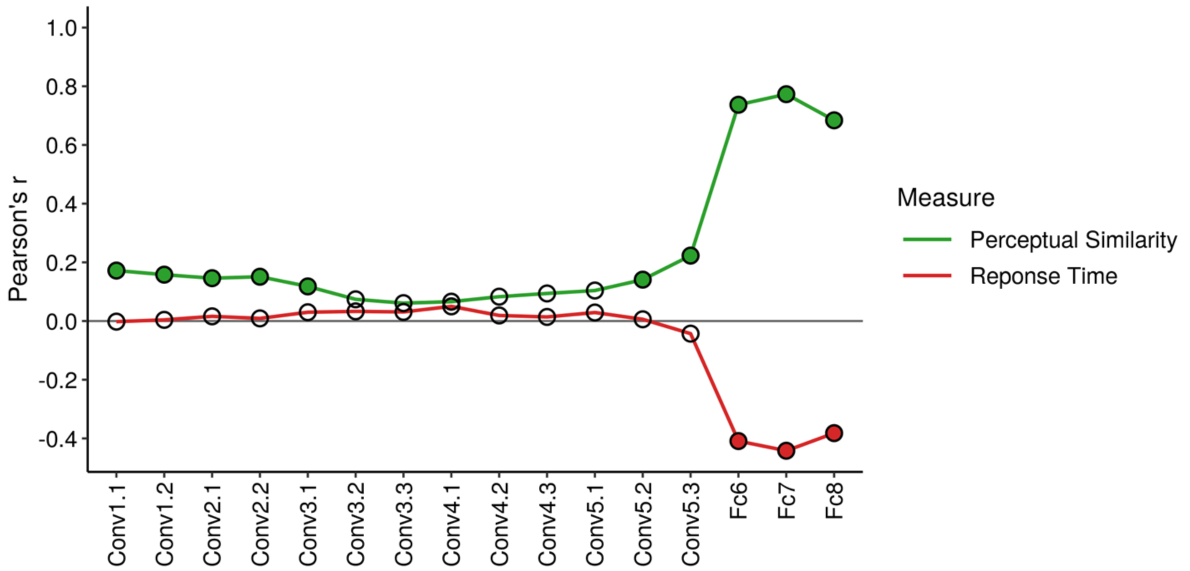

Figure 3.11. *Correlation analysis between the DCNN outputs for each layer with the perceptual similarity ratings (shown in green) and response times (shown in red), showing the ability of the perceptual measures to predict the outputs of each layer of the network. Significant results are depicted by the filled symbols.*

3.7 General Discussion

The aim of this study was to investigate whether view-symmetric representations are an important intermediate step in the generation of view-invariant representations that are used for the recognition of faces in human and artificial neural networks. The main findings from this study are: (1) The emergence of view-symmetric responses is different for canonical and non-canonical rotations; (2) View-symmetric representations are evident for familiar and unfamiliar faces; (3) View-symmetry is also a feature of DCNNs trained for face recognition. Together these findings argue that view-symmetric representations play an important role in the perception and recognition of faces, for both human and artificial neural networks.

First, we investigated the emergence of view-symmetric patterns of response in unfamiliar faces following naturally occurring (canonical) rotations of the head. We found that patterns of neural responses to canonical rotations were view-dependent in early visual areas. That is, the neural response was predicted by the degree of rotation between different viewpoints. These findings are consistent with other neurophysiological (Perrett et al., 1991, 1998; Freiwald & Tsao, 2010; Dubois et al., 2015) and neuroimaging studies (Carlin et al., 2011; Axelrod & Yovel, 2012; Kietzmann et al., 2012; Ramírez et al., 2014; Dubois et al., 2015; Guntupalli et al., 2017; Flack et al., 2019) that have also found selectivity to specific viewpoints of the face. They also fit with behavioural studies that have shown the importance of view-specific representations in the perception and recognition of unfamiliar faces (Bruce, 1982; Hill & Bruce, 1996; Fang & He, 2005; Longmore et al., 2008). However, there was a gradual decrease in the view-specific response from early to higher level visual areas and a corresponding increase in view-symmetric responses, particularly in face-selective regions. The importance of view-symmetrical neural responses is shown by the fact that symmetrical faces are perceived to be more similar than asymmetrical faces (see also, Troje & Bülthoff, 1998; Busey & Zaki, 2004; Flack et al., 2019).

To determine whether view-symmetric responses are specific to naturally occurring rotations of the face, we also measured the behavioural and neural response to non-canonical rotations of the face. We found there was limited evidence for the pattern of neural response in early visual areas being systematically predicted by changes in viewpoint, as was found with canonical rotations. However, we did find view-symmetric neural patterns of response for non-canonical rotations in face regions, which is consistent with the behavioural finding that symmetrical viewpoints were perceived to be more similar than asymmetrical viewpoints. This suggests that the emergence of view-symmetric responses occurs differently for canonical and noncanonical rotations of the face. Our findings are consistent with previous neurophysiological studies that have reported view-symmetric responses in face regions to non-canonical rotations that occur as a result of within-plane rotations of the image (Friewald & Tsao, 2010). These findings may be more consistent with a more general preference for bilateral symmetry in the visual system (Corballis & Beale, 2020; Rhodes et al., 2005; Jacobsen et al., 2006; Bertamini et al., 2018; Keefe et al., 2018; Makin et al., 2012).

The recognition of familiar faces requires the ability to integrate information from different viewpoints into an invariant representation (Bruce & Young, 1986; Young & Burton, 2017; Bruce, 2017). One possible mechanism for generating view invariance is the convergence of view-dependent responses onto a view-invariant representation (Bruce & Young, 1986; Burton et al., 1999). The discovery of view-symmetric neural responses suggests that they may provide an important intermediate computational step before full invariance is achieved (Freiwald & Tsao, 2010). However, a limitation of previous studies is that they have only used unfamiliar faces, whose representations are more closely linked to the image and do not generalise well to new viewpoints when compared to familiar faces (Bruce, 1982; Hancock et al., 2000; Longmore et al., 2008). Our current findings show that view-symmetric neural responses are also evident for familiar faces in core face regions despite the fact that they can be easily recognised across different views. This suggests that the view-invariant representations that are characteristic of familiar faces emerge at later stages of processing (Davies-Thompson et al., 2013; Weibert et al., 2016).

We also found that symmetric views of familiar faces were perceived to be more similar than asymmetric views. This also fits with evidence that symmetrical views may convey an advantage when learning new faces. In a previous study, it was found that when participants were tested with novel face images that were symmetrical to learnt viewpoint, recognition rates were higher than when the learnt and test faces had asymmetrical viewpoints (Flack et al., 2019). Moreover, the pattern of recognition performance was predicted by the pattern of neural response in face-selective regions, such as the FFA. Together, this suggests that view-symmetric representations may play an important intermediate step in the recognition of familiar faces.

In the final experiment we were interested in exploring whether view symmetry would also emerge within a deep neural network. We found that view-symmetry is a feature of a neural network and that this emerges predominantly within the fully connected layers. This view-symmetric representation was found for both within and between identity comparisons. View-specific representations were also evident in the fully-connected layers, but to a lesser degree than view-symmetry. However, view-specific responses were only found for the within identity combinations. The change from the convolutional layers to the fully-connected layers is consistent with change in representation from the early visual areas to the face regions that was found in the neuroimaging data. Human perceptual similarity ratings were also able to predict the similarity of the outputs of a DCNN in the fully connected layers. Taken together these findings suggest that view symmetry is utilised and processed in a similar way for both human observers and DCNNs.

In conclusion, this study investigated the role of view-symmetric responses in face recognition, for humans and within a DCNN. We show that view-symmetrical patterns of response to familiar faces can be found in face-selective regions for both canonical and non-canonical rotations. However, we show distinct differences in the way that view-symmetric responses emerge along the visual hierarchy for canonical and non-canonical rotations of the face. Finally, we show that view-symmetry is also a feature of DCNNs and that view symmetry is computed in a similar way within human and artificial neural networks. These findings provide important evidence in support of the role of view-symmetry as an important intermediate processing stage in the perception and recognition of faces.

Chapter 4- The roles of shape and texture in the recognition of familiar faces.

This Chapter is adapted from: Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2022). The roles of shape and texture in the recognition of familiar faces. Vision Research, 194, 108013.

4.1 Abstract

The surface texture of the face is proposed to be the dominant cue in face recognition. In this study, we investigated the role of shape information in face recognition. We compared the roles of shape and surface texture in the recognition of face identity using familiar and unfamiliar faces. In the first experiment (n = 53), participants had to match the name of a familiar person to one of eight hybrid face images, in which the average shape from one facial identity was combined with the average texture of a different identity. In texture trials, all images had the correct shape, but only one image had the correct texture. In shape trials, all images had the correct texture, but only one image had the correct shape. Although performance was lower for the shape trials (81%) compared to texture trials (99%), both were significantly above chance-level (12.5%). In the second experiment (n = 110), participants had to name hybrid faces using a free recall paradigm. Thus, there were two potentially correct answers for each face image: one based on the texture and one based on the shape. Participants reported the correct name based on the texture information on 61% of trials and the correct name based on the shape information on 12% of trials. Importantly, neither task could be performed by perceptual matching. In the third experiment (n = 19), fMR-adaptation was used to measure the neural sensitivity to changes in the shape or texture. The core face-selective regions showed a similar sensitivity to both shape and texture properties. These findings confirm that texture is the dominant cue utilised for face recognition, but also show that shape plays an important role in the recognition and neural response to familiar faces.

4.2 Introduction

Recognising the identity of a person from their face is fundamental for appropriate social interactions. Processing what visual information is used to recognise faces is central to understanding this behaviour. In face perception, a distinction can be made between the texture (or surface) properties of the face and its shape properties (Bruce & Young, 1998; 2012). Any facial image comprises of a set of edges generated by abrupt changes in reflectance due to the shapes and positions of facial features. These shape properties usually arise from how the 3D geometrical description of the face is projected onto a 2D image (Maurer, Le Grand, & Mondloch, 2002). Facial images also contain a broader pattern of reflectance based on the surface properties of the face. 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 (Bruce & Young, 2012).

Texture plays a critical role in the perception of a face identity (Burton, 2013). For example, familiar face recognition is still possible when surface properties are projected onto a standardised shape (Burton, Jenkins, Hancock, & White, 2005) or when linearly stretching or morphing a face image in a way that dramatically alters the shape of the face (Hole et al., 2002; Sandford & Burton, 2014; Baseler et al., 2016; Itz et al., 2014; Itz et al., 2016). Changes to the texture of the face, on the other hand, caused by contrast negation or spatial blurring have a dramatic effect on recognition, even when the shape of the face is unchanged (Bruce & Langton, 1994; Kemp et al., 1996; Hole et al., 2002). It is also difficult to recognise line drawings of a familiar face that have the correct shape, but limited texture information (Leder, 1999). Moreover, perceptual matching of facial identity has been shown to be more accurate when based on texture compared to shape (Andrews, Baseler, Jenkins, Burton, & Young, 2016).

Although these studies imply that texture information provides the dominant cue for face recognition, manipulations of shape can have a significant effect on the judgements of recognition. For example, non-linear manipulations of shape can have a significant effect on the ability to recognise identity (Hole et al., 2002). Further support for the role of shape in face recognition comes from studies that show shape information can be used to discriminate unfamiliar face images (O'Toole et al., 1999; Jiang et al., 2006; Russell et al., 2007; Russell &

Sinha, 2007; Caharel et al., 2009; Jiang et al., 2011; Lai et al., 2013; Itz et al., 2016). Although judgements based on texture are more accurate than judgements based on shape, it is still possible to make some use of shape information in matching tasks involving familiar faces (Andrews et al., 2016). However, in all of these studies it is possible that these tasks involving shape could be performed by lower-level perceptual matching of features rather than higher-level processes critical to the recognition of identity in natural viewing conditions (Burton et al., 2015). Moreover, training individuals with poor face recognition skills using faces that have been caricatured for their shape or texture can both lead to improvements in face processing (Limbach, Itz, Schweinberger, Jentsch, Romanova & Kaufmann, 2022). Indeed, a challenge for a central role of shape in face recognition is that shape cues (particularly those involving the internal features of the face) can vary quite dramatically across different images of the same person (Burton, 2013; Burton et al., 2015). For example, the spatial distances between features can often vary as much within-person as between-person.

The behavioural sensitivity to the shape and texture of faces should be mirrored by the neural responses of face-selective regions involved in recognition. Neuroimaging studies have revealed a core network of face-selective regions in the occipital and temporal lobes that are involved in the perception and recognition of faces (Haxby et al., 2000; Kanwisher et al., 1997). Within this network, the fusiform face area (FFA) is held to be important for representing invariant facial characteristics that play an important role in the recognition of facial identity (Haxby et al., 2000; Grill-Spector et al., 2004; Rotshtein et al., 2005). Support for the importance of the FFA in processing facial identity is found in neuroimaging studies that have shown adaptation to repeated images of different faces in this region (Andrews & Ewbank, 2004; Grill-Spector et al., 1999). This suggests that the neural response in the FFA represents the identity of the face and that this representation is being adapted during exposure to repeated images.

A more robust link between activity in the FFA and face recognition would be a demonstration that adaptation is still found when the images vary along a dimension that is not important for face recognition (i.e. changes in shape). For example, Jiang and colleagues (Jiang, Dricot, Blanz, Goebel, & Rossion, 2009; see also- Caharel et al., 2009; Itz et al., 2016) found an equal release from adaptation to identity in the FFA with changes in either the shape or texture. This suggests that both properties are represented in this region, which differs from behavioural studies of familiar faces that show a greater sensitivity to changes in texture. These findings might be explained by their use of unfamiliar faces, however a similar release from adaptation to shape and texture with familiar faces has also been found (Andrews et al., 2016). Although this provides further support for a dissociation between the behavioural and neural response to faces, it is possible that a more sensitive adaptation paradigm could show a difference in the neural response to shape and texture.

The aim of this study is therefore to achieve a more detailed understanding of the relative roles of shape and texture in the recognition and neural response to familiar and unfamiliar faces. We used hybrid face images in which the surface texture from one identity is combined with the shape from another identity (Andrews et al., 2016). Our aim in this study was to test recognition directly using tasks that had no component of perceptual matching and instead relied on previously learnt representations for recognition. In all experiments, we compared familiar and unfamiliar faces as previous research has shown differences in how shape and texture information are used when making judgments of familiar and unfamiliar faces (Itz et al., 2014; Itz et al., 2017; Zhou et al., 2021). In the first experiment, a name was shown and participants had to match that name to one of 8 hybrid face images. The images shown varied in either shape or texture. In the second experiment, participants viewed individual hybrid images and were asked to name the person. There were two potentially correct answers for each familiar face: one based on texture and one based on shape. This directly compared the relative role of shape and texture in the representation of familiar faces. In the final experiment, we measured the relative sensitivity to shape and texture in face-selective regions of the human brain, using an fMR-adaptation paradigm that has previously been used to reveal invariant responses to faces in face-selective regions such as the FFA (Davies-Thompson, Newling, & Andrews, 2013). If shape information is important for familiar face recognition, then we would expect: (1) there should be above chance level performance when matching the correct face to a familiar target name (Experiment 1); (2) there will be occasions when participants select the identity whose shape information is present within a hybrid face (Experiment 2); (3) there will be neural sensitivity to changes in shape within the face selective regions of the brain (Experiment 3).

4.3 Methods

4.3.1 Participants

Participants were recruited separately for the behavioural and fMRI experiments (Experiment 1: n = 53, female = 38, mean age = 26.9 years, SD = 9.8; Experiment 2: n = 110, female = 62, mean age = 22.7 years, SD = 6.8; Experiment 3: n = 19, female = 10, mean age = 25.4 years, SD = 1.39). A priori power analyses (0.9, 1-β err prob) were conducted for Experiment 1 (suggested N = 55) and Experiment 2 (suggested N = 110). Sample size for Experiment 3 was based on previous studies using similar paradigms (Andrews et al., 2016; Baseler et al., 2016). Participants were drawn from an opportunity sample of staff and students from the University of York. All participants had normal or corrected to normal vision with no history of neurological illness and gave their written informed consent. The study was approved by the Psychology department Ethics Committee and the York Neuroimaging Centre Ethics Committee.

4.3.2 Stimuli

Figure 4.1 shows the stimuli sets that were used for the behavioural and neuroimaging experiments. The faces used were either familiar (UK celebrities) or unfamiliar (Australian celebrities) in the UK. The familiar images were grayscale average images that were generated by combining 12 different images from each of the 8 celebrities who are generally familiar to UK participants (Alan Sugar, Chris Moyles, Derren Brown, Gary Lineker, Jeremy Paxman, Jeremy Kyle, Louis Walsh). The unfamiliar images were also based on average images generated by combining 12 different images from each of the 8 Australian celebrities who are likely to be unknown to our participants (Brendan Nelson, Don Burke, Grant Hackett, Guy Sebastian, Kyle Sandilands, Mark Holden, Morris Iemma, Shannon Noll).

The averaging procedure was performed using graphics software (Interface) in which key fiducial points on the face were defined in each image (by hand), and then connected to form a grid showing the shape or the second-order configural properties of the image (for details see Burton et al., 2005; Burton, Schweinberger, Jenkins, & Kaufmann, 2015). A common shape was then determined for each identity by averaging the spatial location of corresponding points on the grid across all images taken of the same identity. Average textures for each identity were created through standard face morphing techniques, where each face is modified to conform to a common shape. This is achieved by warping the pixels within each triangle of the source image so that they match the shape of the corresponding triangle in the standard shape. Finally, texture averages are generated by averaging across all images with the same identity. For further information, refer to Beale and Keil (1995). This process enables separate analysis of face shape (represented by grid points before morphing) and face textures (shape-free faces) where the shapes and the placement of features are consistent across all faces in the analysis.

The raw photo images were selected using an internet image search on the celebrities' names. To generate an ambient image set, reflective of natural viewing, the only selection criteria were that the full face was visible in high resolution. Selecting images in this way has been shown to provide robust averages (Jenkins, White, Van Montfort, & Burton, 2011; Burton, Kramer, Ritchie, & Jenkins, 2016; Jenkins, Burton, & White, 2006). One important consideration when selecting images in this way, is that the colour balance between images can be highly variable. Here, colour balance refers to the adjustment of colours in an image to achieve a desired and natural appearance. It is a fundamental aspect of colour correction and image editing to ensure that the colours in an image are accurate and consistent, representing the image as it would appear to the human eye. Thus, due to this variability in colour balance across images, coupled with the need to average images within identities, it was necessary to first convert all images to greyscale. Converting images to greyscale has been shown to have a minimal impact on familiar face recognition (Kemp, Pike, White & Musselman, 1996), and no impact at all when the images are in high resolution, implying that colour does not provide diagnostic information for face recognition (Yip & Sinha, 2001).

The images on the diagonal (top left to bottom right) in each panel of Figure 4.1 show shape and surface properties from the same identity. Because the shape and surface information are generated separately, it is also possible to combine them across different identities to generate hybrid images. Hybrid faces are shown in the off-diagonal images. Images in each column have the same shape, whereas images in each row have the same surface properties.

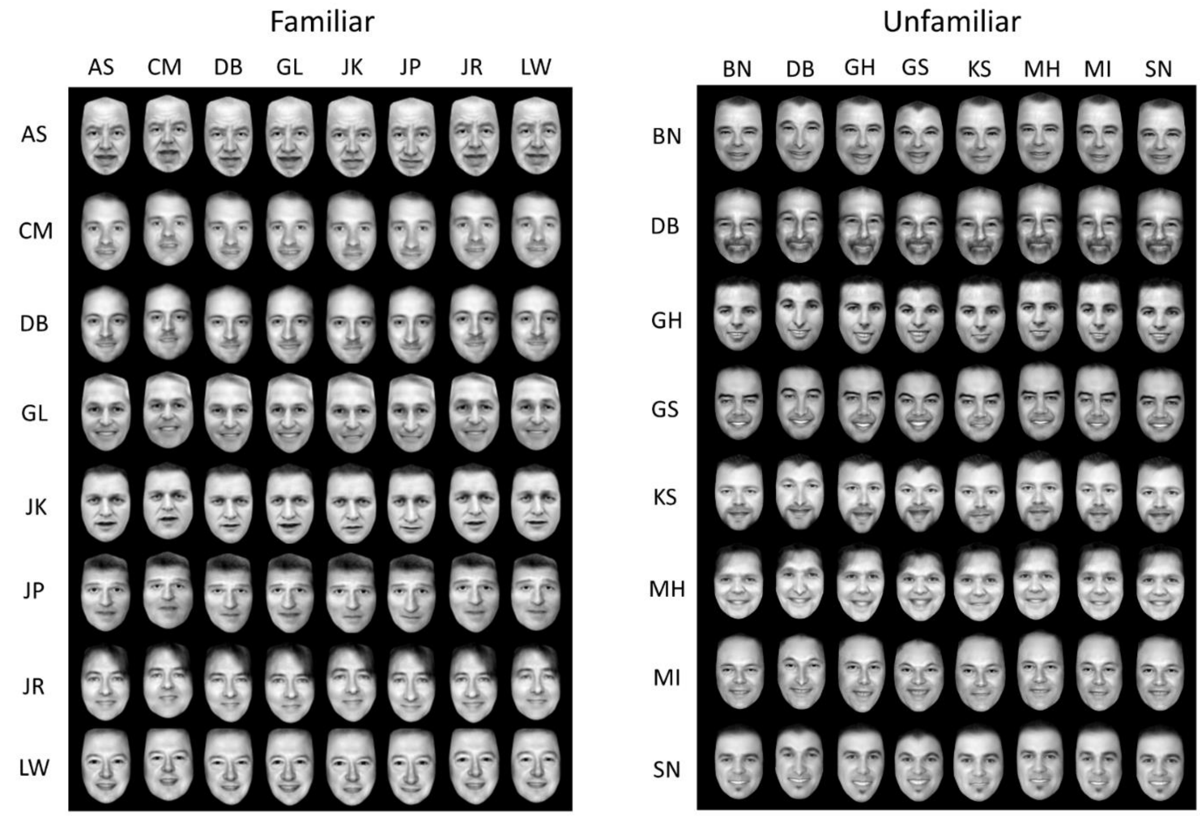

Figure 4.1. *Familiar and unfamiliar hybrid face images. Hybrid images were created by combining the average shape from one identity with the average texture from another identity. The diagonal images (top left to bottom right) contain the average shape and texture properties of the same identity. Rows depict images containing the average texture of one identity and the average shapes of other identities. Columns depict images containing the average shape of one identity and the average textures of other identities.*

4.4 Experiment 1- Name to face matching task

4.4.1 Methods

To compare the relative roles of shape and texture in the recognition of familiar faces, participants had to match a name to faces that varied in either shape or texture. Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Participants were first presented with the name of an identity that was displayed centrally on the screen, this was followed by eight simultaneously presented hybrid face images (Figure 4.2). There were two categories of trials, to explore the role of texture during face recognition, within a texture trial, all the faces had the shape of the target, but only one also had the correct texture. To explore the role of shape information within a shape trial, all faces had the same texture as the target, but only one also had the correct shape. Participants used a button press to indicate which face corresponded to the target. There were 32 trials (8 familiar and 8 unfamiliar identities x shape/texture). This experiment was self-paced and no feedback was given. After the task was completed, participants then completed a familiarity check to test their ability to recognise the familiar faces used in the main experiment. For each identity used in the main experiment, three novel high-resolution colour images were presented to participants and their task was to name (or enter sufficient biographical detail) the identity depicted in each image set. Overall, 88.9% of intended familiar identities were recognised; identities that were not familiar were removed on an individual participant basis. Only 3.3% of the intended unfamiliar identities were recognised; these were also removed on an individual participant basis.

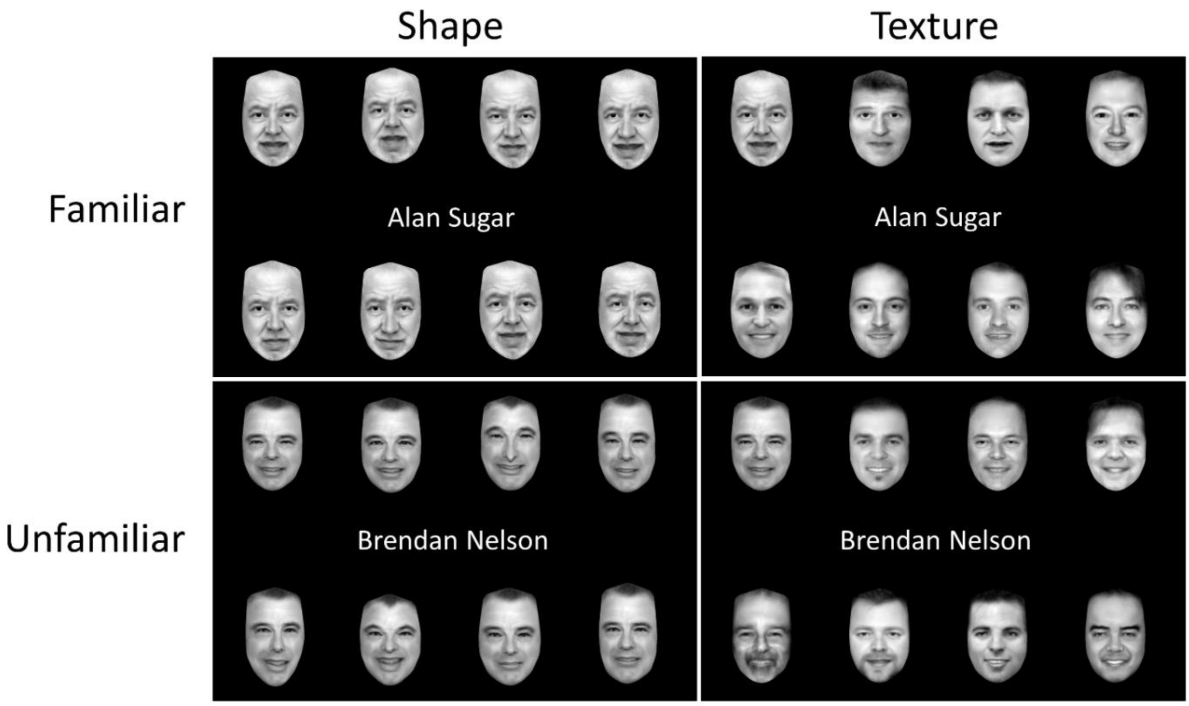

Figure 4.2. *Experiment 1- Name to face matching task: Examples of shape and texture trials for familiar and unfamiliar faces. Participants had to match a name to one of 8 hybrid face images. In shape trials, all faces had the same texture, but only one face had the correct shape. In texture trials, all faces had the same shape, but only one shape had the correct texture.*

4.4.2 Results

Experiment 1 aimed to measure the ability to recognise faces based on either shape or texture. Figure 4.4 shows the recognition accuracy for familiar and unfamiliar faces, separated by trial type (shape or texture cue). To determine whether recognition accuracy differed when using a shape or texture cue, or when faces were familiar or unfamiliar, a 2 × 2 repeated measures ANOVA with Familiarity (familiar, unfamiliar) and Cue (shape, texture) as the main factors was computed. Significant main effects were found for Familiarity ($F(1, 52) = 1081.86$, $p < .001$, $\eta p^2 = 0.954$) and Cue ($F(1, 52) = 28.41$, $p < .001$, $\eta p^2 = 0.353$). There was also a significant interaction between Familiarity and Cue ($F(1, 52) = 17.48$, $p < .001$, $\eta p^2 = 0.253$). This interaction reflects higher performance for texture compared to shape with familiar, but not unfamiliar faces. For familiar faces, there was a significant difference between accuracy between shape (mean ± SEM = 81.0 ± 13.62) and texture (mean ± SEM = 99.4 ± 2.59) trials for familiar faces ($t(52) = 9.66$, $p < .001$, $d = 1.67$). For unfamiliar faces, recognition rates were lower (texture: mean ± SEM = 31.7 ± 17.7; shape: mean ± SEM = 30.2 ± 17.0) and there was no difference between shape and texture trials ($t(52) = 0.46$, $p = .648$, $d = 2.10$). In this experiment chance level was 12.5%, reflecting participants' ability to select the correct image from the eight faces presented in each trial. To determine whether recognition accuracy was greater than chance level, one sample t-tests were conducted for all conditions. For familiar faces, recognition accuracy was greater than chance level on shape trials ($t(52) = 36.61$, $p < .001$, $d = 5.03$) and texture trials ($t(52) = 244.62$, $p < .001$, $d = 33.61$). Recognition accuracy was also greater than chance level for unfamiliar shape trials ($t(52) = 7.59$, $p < .001$, $d = 1.04$) and unfamiliar texture trials ($t(52) = 7.89$, $p < .001$, $d = 1.08$).

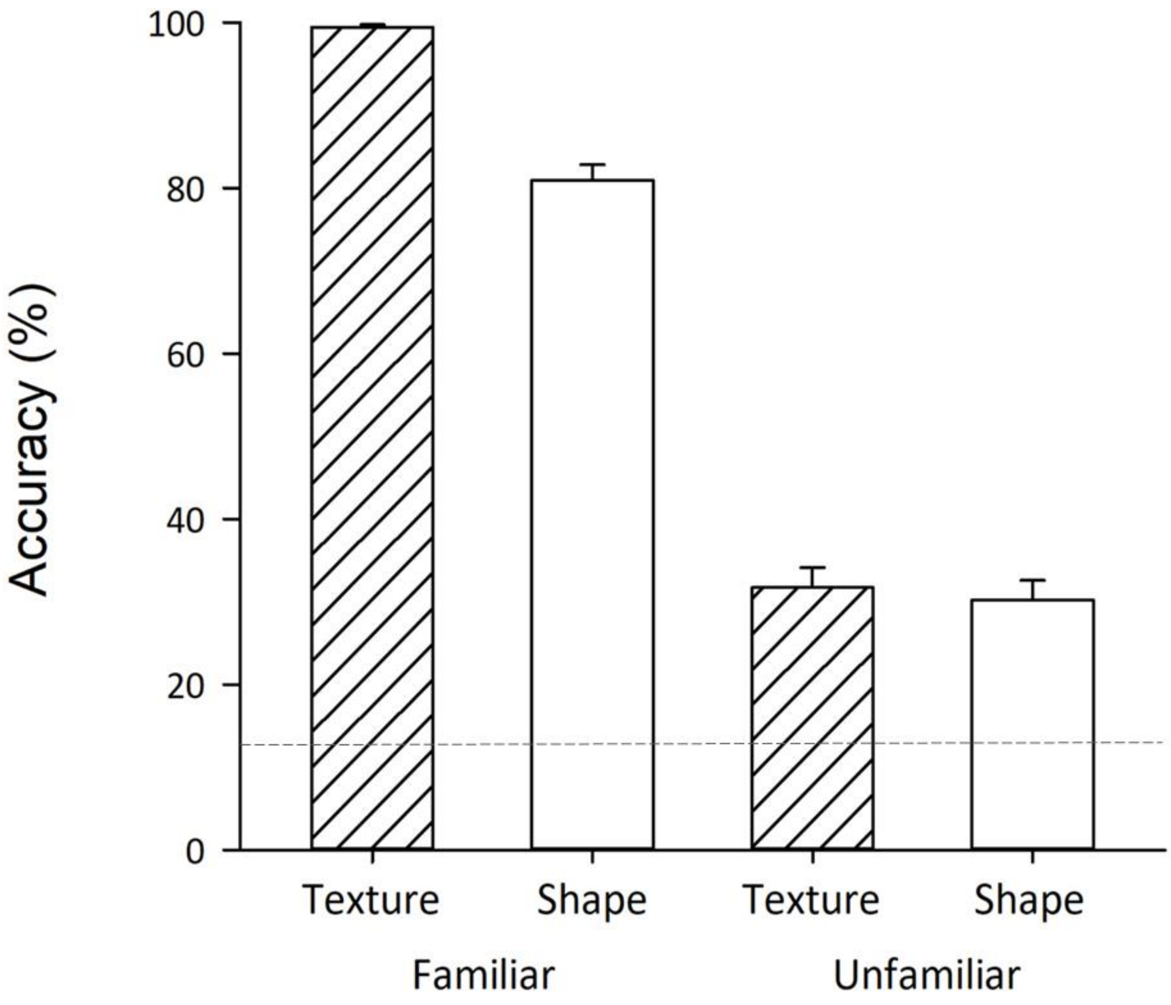

Figure 4.3. *Accuracy on shape and texture trials in Experiment 1. Accuracy for familiar faces was above chance (12.5%-represented by the dotted line) for both shape and texture. However, accuracy for texture trials was significantly higher than for shape trials. Accuracy for unfamiliar trials was substantially lower but still significantly above chance, despite the fact that participants were not familiar with the identities. However, there was no difference between shape and texture in the unfamiliar trials. Error bars represent SEM.*

4.5 Experiment 2- Face recognition task

4.5.1 Methods

In a complementary behavioural experiment, a separate group of participants performed a recognition task on the familiar and unfamiliar hybrid images used in Experiment 1. In this task, participants viewed 16 hybrid faces (8 familiar and 8 unfamiliar). Each image was presented sequentially, and participants were instructed to name the identity depicted in the image with no time constraints. We used multiple groups of participants whereby each group viewed different combinations of hybrid images, such that the shape and texture from all identities was shown equally across the participants. Moreover, the shape or texture of each identity was contained only once in the images shown to each participant. This prevented any effect of priming that might have occurred (for example, if the texture of a face in one hybrid increased the chance of recognising the shape of a face in another hybrid or vice versa). To provide a baseline of performance, one group of participants viewed non-hybrid images in which the shape and texture were from one identity (veridical images in Figure 4.1). Following this, participants then completed the same familiarity test used in Experiment 1. Again, identities that were expected to be familiar or unfamiliar but were not, were removed prior to analysis on an individual participant basis.

4.5.2 Results

In this experiment, participants had to report the identity of hybrid face images that contained the texture from one identity and the shape from another identity, using a free recall paradigm relying on previously stored facial representations. For each trial, there were two potentially correct answers- the identity whose shape was depicted and the identity whose texture was depicted. Figure 4.5 shows the proportion of trials in which participants were able to recognise the face based on the shape or texture of the image. For familiar faces, participants reported the identity based on the texture (mean ± SEM = 61.2 ± 16.4 %) more often than based on the shape (mean ± SEM = 12.3 ± 11.4 %) of the hybrid image ($t(69) = 19.87$, $p < .001$, $d = 2.38$). This shows that texture is a more dominant cue for recognition. Nevertheless, there were trials in which the shape was the dominant cue for recognition. The reported shape ($t(69) = 9.08$, $p < .001$, $d = 1.08$) and texture ($t(69) = 31.31$, $p < .001$, $d = 3.73$) were both significantly greater than 0. Contrastingly and rather unsurprisingly, there were no correct identifications based on shape or texture for the unfamiliar faces.

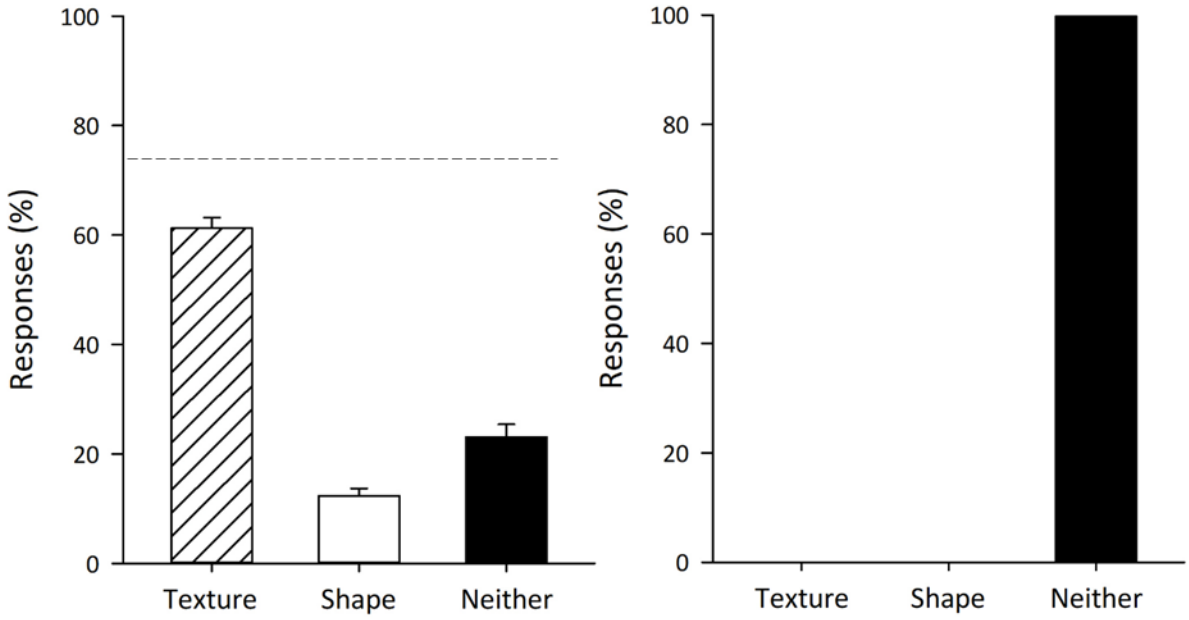

Figure 4.4. *Distribution of responses for Experiment 2. Participants had to recognise the identity of familiar or unfamiliar hybrid faces. There were two potentially correct responses. For familiar faces, participants reported the identity based on the texture more often than the shape of the face. Nevertheless, there were a significant number of hybrid faces that were recognised from their shape. Error bars represent 1 SEM. Dotted line shows performance when the shape and texture were from the same identity (74%).*

A separate group of participants were shown veridical hybrid images in which the shape and the texture were from the same person. The recognition rate for this group can be seen as the maximum expected recognition rate for the hybrid images. After taking out identities that participants reported not knowing during the familiarity checklist, the accuracy rate of the control group was 74% of faces (Figure 4.4 – dotted line). A one-sample t-test showed there were significant differences between this maximal rate and the rate based on shape ($t(69) = 45.69$, $p < .001$, $d = 5.46$) and texture ($t(69) = 6.70$, $p < .001$, $d = 0.80$). This implies that shape and texture properties both carry information regarding identity.

One possible explanation for these findings is that the responses are dominated by faces that have a particularly recognisable texture or shape. To address this issue, we measured the percentage of correct texture or correct shape responses that corresponded to each of the 8 familiar face identities. As can be seen in Table 4.1, the shape and texture hits were evenly distributed across all identities. This implies that our results do not simply reflect the properties of identities with a particularly dominant shape or texture.

Table 4.1. *Percentage of responses for each familiar identity relative to the total Shape hits or Texture hits.*

Identity	Shape hit	Texture hit
Alan Sugar	11.6	14.0
Chris Moyles	7.3	10.2
Derren Brown	13.0	11.1
Gary Lineker	14.5	13.1
Jeremy Kyle	15.9	13.7
Jeremy Paxman	13.0	12.5
Jonathan Ross	11.6	13.1
Louis Walsh	13.0	12.2

4.6 Experiment 3

4.6.1 Methods

To measure the neural sensitivity to shape and texture, we used a block design fMR-adaptation paradigm with 5 different stimulus conditions (see Figure 4.5 for familiar faces): (1) no change (same shape, same surface); (2) shape change (alternating between two shapes, same surface); (3) surface change (alternating between two textures, same shape); (4) shape & surface change-2 (different shape, different texture alternating between two identities) (5) shape & surface change-8 (different shape, different surface-alternating between eight identities). The shape & surface change-2 condition, was included in order to be comparable to the shape change and surface change conditions that alternated between two identities, whilst the shape & surface-8 condition, was used in order to increase the sensitivity of the paradigm by showing the maximum release from adaptation. Similar fMR-adaptation designs have been used in previous experiments to reveal invariant representations of identity in face-selective regions (Davies-Thompson et al., 2013). Data were collected separately using this design for familiar and unfamiliar faces.

Figure 4.5. *FMRI experimental stimuli depicting the familiar faces (British celebrities). Each row portrays an example of images presented during a single 9 s block. A. No change condition; B. Shape change only, alternating between two shapes (AB design); C. Texture change only, alternating between two textures (AB design); D. Shape and texture change, alternating between two identities (AB design); E. Shape and texture change, 8 different identities presented in a block.*

In each stimulus block, 8 images were shown for 975ms followed by a 150ms blank screen. Blocks were 9s in duration and were separated by a 9s fixation screen (a white fixation cross on a mean grey background). Each of the 5 stimulus conditions was repeated 8 times, giving a total of 40 blocks for each scan, which were presented in a counterbalanced order.

Participants performed a red dot detection task during the scan, in which they were required to press a button when a red dot appeared on any of the images. Mean accuracy was 92% across all familiar conditions (mean response time-494ms) and 94% across all unfamiliar conditions (mean response time-493ms). 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 gradient-echo EPI and a T1-weighted structural MRI (1 × 1.13 × 1.13 mm voxel) were acquired for each participant. The gradient-echo EPI sequence used a radio-frequency coil tuned to 127.4 MHz to acquire 38 axial slices (TR 3 sec, TE 33 msec, flip angle 90, FOV 260 mm, matrix size = 128 × 128, slice thickness = 3 mm, voxel size: 2.25 × 2.25 × 3 mm). Data were analysed with FEAT version 4.1 (http://www.fmrib.ox. ac.uk/fsl). The first 9s (3 volumes) from each scan were discarded, and MCFLIRT motion correction, spatial smoothing (Gaussian, FWHM 6 mm), and temporal high-pass filtering (cutoff 0.0093 Hz) were applied.

A localiser scan was used to identify face-selective regions. The localiser scan images included faces, bodies, inanimate objects, places, and scrambled images. The identity of the faces was different to those used in the main experiment. Images from each category were presented in blocks of 10 images in which images were shown for 700ms, followed by a 200ms blank screen. A 9s grey screen with a central fixation cross was presented between each block. Stimulus blocks were repeated 4 times and were presented in a counterbalanced order. A boxcar function convolved with a standard haemodynamic response function was used to model the BOLD response. Face-selective voxels were defined by contrasting the response to faces with each non-face condition, then averaging the resulting statistical maps and thresholding at $p < .001$ (uncorrected). Neighbouring clusters of voxels located within the occipital and temporal lobes were defined as the FFA, OFA and pSTS in each participant.

The experimental scans were analysed by measuring the time series of response to each condition. Across each scan, the response of each voxel was converted to % signal. A single time series for each ROI was then calculated by averaging across all voxels. Each block was then normalized by subtracting the magnitude of response at the start of the block from the response at each time point in the block. The normalized response to the same stimulus blocks was then averaged to produce a mean time series. The average of the % signal change at 9s and 12s post stimulus onset was taken as the peak response for each condition within an ROI for each participant. The peak responses were then analysed using repeated measures

ANOVAs and post hoc t-tests. Specific contrasts were used to compare each experimental condition to the no-change condition. This allowed us to determine whether there was a release from adaptation (or sensitivity) to each manipulation.

To determine whether any differences in the release from adaptation could reflect differences between the image properties of the familiar and unfamiliar faces, we measured the mean change in image intensity across images. This was calculated by taking the average of the absolute differences in grey value at each pixel for successive pairs of images within a block. A 2 × 5 ANOVA with Familiarity (familiar, unfamiliar) and Condition (No change, Shape change, Texture change, Shape and Texture Change (2), Shape and Texture Change (8)) as the main factors was ran. There was a significant main effect for Condition ($F(4,220) = 194.24$, $p < .001$), but there was no main effect of Familiarity ($F(1,55) = 0.35$, $p = .555$) or any interaction between Familiarity * Condition ($F(4,220) = 0.40$, $p = .811$). The largest change in low-level properties was found when both shape and texture changed. However, shape and texture changes for familiar and unfamiliar had a similar effect on this image measure. There was also no difference between the shape change and texture change for familiar faces ($t(55) = -1.85$, $p = .070$) and unfamiliar faces ($t(55) = -0.16$, $p = .977$). These findings ensure that any releases in adaptation for shape and texture changes are not due to low-level image properties such as image intensity

4.6.2 Results

A localiser scan was performed to reveal the location of face-selective regions. The average location of the core face-selective regions: fusiform face area (FFA), occipital face area (OFA) and posterior superior temporal sulcus (pSTS), is shown in Fig. 4.6A and Table 4.2. We next determined how these regions responded to changes in shape or surface properties of faces. A 3-way ANOVA found no interaction effect of hemisphere * condition (familiar: $F(1,14) = 2.09$, $p = .170$; unfamiliar: $F(1,15) = 2.06$, $p = .172$), so the responses from each hemisphere were combined.

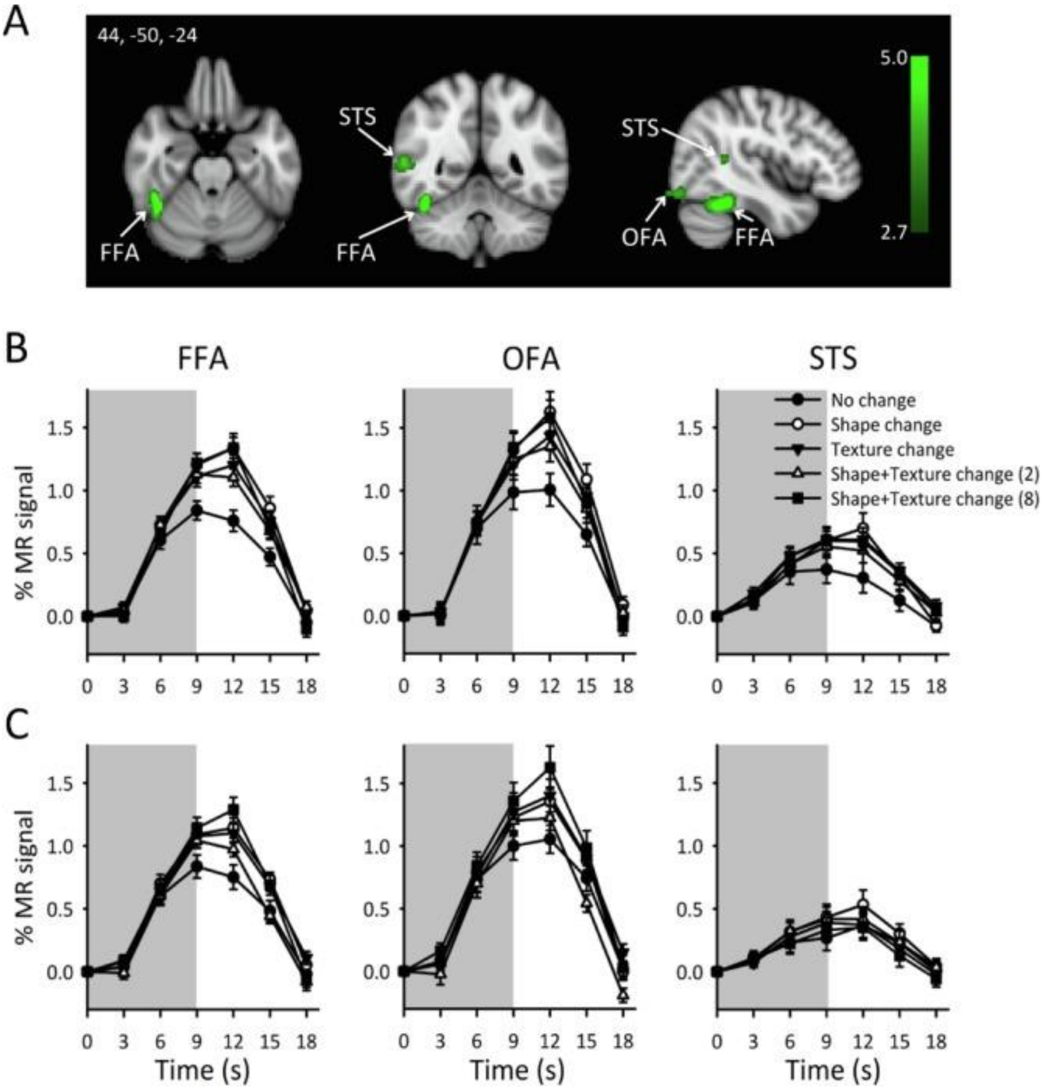

Figure 4.6. *(A) Location of face-selective regions-of-interest (FFA: fusiform face area, OFA: occipital face area, STS: superior temporal sulcus. (B) The average timeseries for face-selective regions of interest in response to familiar faces. There was a significant release from adaptation (compared to no change) for familiar faces in all regions for all conditions. There was a similar release from adaptation to texture and shape. (C) There was a similar release from adaptation with unfamiliar faces in the FFA and OFA, but there was no effect in the STS. Time shows the response relative to the onset of the block. Grey shading shows the stimulus duration. Error bars show SEM.*

Table 4.2. *Mean (SEM) MNI coordinates of regions of interest (centre of gravity). Regions defined by localiser scan (Faces > (Bodies + Objects + Places + Scrambled images).*

Region	**x**	**y**	**z**
FFA			
L	−41.94 (0.82)	−55.63 (1.52)	−21.66 (0.96)
R	42.55 (0.63)	−52.04 (1.34)	−21.33 (1.00)
OFA			
L	−39.34 (1.31)	−83.52 (1.01)	−14.97 (1.57)
R	41.00 (0.95)	−79.20 (1.16)	−13.68 (0.97)
STS			
R	52.38 (1.79)	−49.52 (1.91)	4.97 (1.39)

Figures 4.6B and 4.6C show the time course of response to different conditions in the different regions when viewing familiar and unfamiliar faces respectively. The effect of condition was analysed using the peak responses with a 1-way ANOVA. There was a significant effect of condition for all face regions with familiar faces (FFA: ($F(4,68) = 14.51$, $p < .001$), OFA: ($F(4,68) = 7.98$, $p < .001$) , pSTS: ($F(4,68) = 6.02$, $p < .001$)). However, for unfamiliar faces, there was only a significant effect of condition for the FFA ($F(4, 72) = 8.81$, $p < .001$) and OFA ($F(4,72) = 6.41$, $p < .001$). The pSTS showed no significant effect of condition for unfamiliar faces ($F(4,72) = 1.25$, $p = .297$).

To measure the release from adaptation in each region, the response to each condition was compared to the no change condition. In the FFA, there was a lower response (adaptation) to the no change condition compared to the shape change (familiar: $t(17) = 6.41$, $p < .001$, unfamiliar: $t(18) = 3.50$, $p = .003$), texture change (familiar: $t(17) = 4.92$, $p < .001$, unfamiliar:

$t(18) = 3.49$, $p = .003$), shape and texture change with 2 identities (familiar: $t(17) = 6.12$, $p < .001$, unfamiliar: $t(18) = 2.84$, $p = .011$), shape and texture change using 8 identities (familiar: $t(17) = 6.25$, $p < .001$, unfamiliar: $t(18) = 6.05$, $p < .001$). However, there was no difference in the response when comparing a shape change to a texture change for either familiar or unfamiliar faces (familiar- [$t(17) = 1.43$, $p = .170$], unfamiliar [$t(18) = 0.32$, $p = .754$). This suggests that the FFA is equally sensitive to changes in shape and texture.

The OFA showed a similar pattern of response to the FFA. There was a lower response (adaptation) to the no change condition compared to the shape change (familiar: $t(17) = 4.50$, $p < .001$, unfamiliar: $t(18)$,= 2.58, $p = .019$), texture change (familiar: $t(17) = 3.89$, $p = .001$, unfamiliar: $t(18) = 3.17$, $p = .001$), shape and texture change when using 2 identities (familiar: $t(17) = 3.03$, $p = .001$, unfamiliar: $t(18)$,= 2.02, $p = .058$) and shape and texture change using 8 identities (familiar: $t(17) = 4.15$, $p = .001$, unfamiliar: $t(18) = 4.54$, $p < .001$). Similar to the FFA, there was no difference in the response when comparing a shape change to a texture change for either familiar ($t(17) = 1.45$, $p = .165$), or unfamiliar ($t(18)$,= 0.61, $p = .555$) faces, suggesting the OFA is also equally sensitive to shape and texture changes irrespective of familiarity.

The pSTS was only found to show an effect of condition with familiar faces. Similar to the FFA and OFA, there was a lower response (adaptation) to the no change condition compared to the shape change ($t(17) = 3.94$, $p = .001$), texture change ($t(17) = 3.61$, $p = .002$), shape and surface change with 2 identities ($t(17) = 2.85$, $p = .011$) and shape and texture change using 8 identities ($t(17) = 2.74$, $p = .014$). There was no difference in response when comparing a shape change to a texture change ($t(17) = 0.68$, $p = .505$), suggesting a similar sensitivity to shape and texture.

4.7 General Discussion

In this study, we investigated the roles of shape and texture in the perceptual and neural representation of familiar (as compared to unfamiliar) faces. The main findings are that: (1) shape can contribute to the recognition of familiar faces in tasks that cannot be performed by perceptual matching; (2) texture is, however, the dominant source of information for familiar face recognition; (3) face-selective regions are equally sensitive to changes in shape and texture.

In the first experiment, we asked how shape and texture information in face images contribute to the recognition of person identity. To address this issue, we used hybrid images that were created by combining the average shape information from one identity with the average texture information from a different identity. We then asked to what extent shape or texture information could be used to match a name to a face. In a previous study (Andrews et al., 2016), it was possible to match a previously presented hybrid face that contained the shape and texture from one identity with a subsequent array in which either the shape or the texture varied. Andrews et al, (2016) found that it was possible to do this task for both shape and texture, but performance on texture trials was higher. However, the task used by Andrews et al, (2016) could be performed with perceptual matching between the target and the test array, limiting its relevance to the ways in which we recognise faces in natural viewing conditions. To address this issue in the current study, we used a task in which participants had to match a written name to an array of faces that either varied in texture or shape. This gave no opportunity for perceptual matching, as participants were only able to rely on previously stored mental representations for recognition.

The findings from this experiment clearly show that participants were able to correctly match the target face with the identity name that was presented, when the identities were familiar to the participant. Participants were able to do this significantly above chance level when the distractor images varied either in shape or texture. Whilst accuracy for texture trials was significantly greater than accuracy for shape trials, performance on shape trials was very high (81%). These findings suggest that both shape and texture information can be used independently when making identity judgments without relying on perceptual matching, supporting findings of previous research (Andrews et al., 2016).

An interesting finding in this experiment was that performance on unfamiliar faces was above chance for both shape and texture trials. This was unexpected because participants were not familiar with the identities (as shown in the post-experiment familiarity test) and hence could not have reflected the association between the name and the correct hybrid image. The accuracy on unfamiliar shape trials was similar to the accuracy on unfamiliar texture trials. This suggests that participants were not using a similar mechanism to that used for familiar faces, in which performance on texture trials was significantly higher than for shape trials. Rather, it would appear that participants were able to reject hybrid images (thus, inflating chance-level) for which the combination of shape and texture did not appear naturally face-like. Here the term ‘face-like’ refers to the extent to which an image of a face looks like a face that could be seen in an everyday environment (not the extent to which a particular image looks like the person depicted- Ritchie, Kramer & Burton, 2018). Due to the image manipulations necessary to generate hybrid images; averaging, shape landmarking and shape-free (texture) warping, the appearance of a resulting hybrid image can sometimes look less face-like than others. This typically occurs when two identities have drastically different shape properties to one another, for example if the size or spacing between certain facial features is heavily different between identities, then the resulting hybrid image can appear less face-like. Thus, when asked to select one out of eight images, participants might be more inclined to select a hybrid image that is more face-like, as this combination of shape and texture information is more prototypical in the real world. These findings suggest the importance of including unfamiliar faces as a point of comparison in studies of familiar faces. The difference in the use of shape and texture in familiar and unfamiliar faces that we show here, converges with previous studies that have also found that texture is disproportionately more important than shape for familiar compared to unfamiliar faces (Itz et al., 2014; Itz et al., 2017; Zhou et al., 2021).

To further explore whether participants were able to use shape or texture for the recognition of identity, we presented participants with hybrid faces and asked them to name the person depicted. In a previous study (Andrews et al., 2016), participants performed a similar experiment in which hybrid faces were presented with a list of possible names. Included in those names was the name associated with the shape of the hybrid and another name that was associated with the texture of the hybrid. It was found that participants selected the identity whose texture was present on 90% of trials and the identity whose shape was present on only 5% of trials. However, a possible limitation of this study is that participants were not directly recognising the face, but were rather using a more cognitive based strategy to relate the appearance of the hybrid face with one of the names. To address this issue, the current experiment simply presented each hybrid face and asked participants to name the person. This task had no component of perceptual matching and could not involve any non-visual cognitive strategy, instead, participants relied on their stored mental representations of the identities. Nevertheless, we found that both shape and texture information were used in this pure recognition task. Whilst more hybrid face images were recognised from their texture compared to their shape, there were some hybrid faces in which the shape was more dominant, over the identity whose texture information was displayed. We also measured performance in participants in which the hybrid faces contained the shape and texture from the same familiar identity. We found that performance with these images was best approximated to the sum of performance on texture or shape alone. The images used in this study were all grayscale because colour is known to have at best a limited role in recognition (Bruce & Young, 2012). An interesting question for further investigations might therefore be the extent to which colour can influence neural responses to surface properties.

The importance of shape in the recognition of familiar faces has been challenged by well-established behavioural findings that show (1) large changes in shape can leave recognition unimpaired, (2) large changes in texture have a significant effect on recognition, (3) texture dominates shape in judgements of identity (Burton et al., 2015). Across behavioural Experiments 1 and 2 the findings confirm the fact that texture is a more dominant cue for recognition, however they also show that shape can make a significant contribution to familiar face recognition. Previous studies that have investigated the role of shape have often manipulated the configuration of facial features in unfamiliar faces (Freire et al., 2000; Le Grand et al., 2001; Rossion, 2008). The typical task is to determine whether two faces are the same or different and the extent to which performance is affected by inversion. However, it has not been clear if this has any relevance to judgements of familiar faces in natural viewing in which it is necessary to recognise a face in the absence of any comparison to other faces. Our results provide the first evidence that shape information plays an important role in recognition, albeit less than for texture information.

The aim of Experiment 3 was to investigate the neural sensitivity of face-selective regions to changes in shape and texture. The intention here was to reveal which regions showed a corresponding sensitivity to that shown in the behavioural Experiments 1 and 2. Using a fMR-adaptation paradigm, we compared neural responses to changes in texture, shape, or both texture and shape with the response to a 'no change' baseline that would create maximal adaptation. In a previous study, the release from adaptation to shape and texture was measured finding that there was an equal release to both changes (Andrews et al., 2016; see also Jiang et al., 2006). However, the lack of any difference in sensitivity to shape and texture may have resulted from a design in which 8 different images were presented in a block. In the current study, a more sensitive paradigm was utilised in which 2 images alternated. Previous research has shown that that this paradigm is able to demonstrate invariant representations in face-selective regions (Davies-Thompson et al., 2013).

We found a lower response (adaptation) in the FFA and OFA to repeated images of the same face compared to faces that differed in both shape and texture (see also Andrews & Ewbank, 2004; Grill-Spector et al., 1999; Weibert et al., 2016). However, the critical conditions were those in which either the shape or the texture changed independently. Given our behavioural results, our predictions were that face-selective regions responsible for the recognition of facial identity should show a release to both shape and texture, but that there should be more sensitivity to changes in texture. We did find a release from adaptation in the FFA and OFA to both shape and texture, but we did not find a difference between shape and texture. The similar sensitivity to shape and texture could not be explained by greater low-level image differences between these changes, as these were similar for both changes. Although, our results show a similar release from adaptation to shape and texture in the OFA and FFA, this does not mean that both regions represent information in the same way. Indeed, a recent study using MVPA (Tsantani et al., 2021) it was shown that the OFA and FFA encode distinct types of face identity information.

There is mixed evidence for whether the FFA has an image-invariant representation of face identity. A number of studies have reported image dependent responses in the FFA (Andrews & Ewbank, 2004; Davies Thompson, Gouws, & Andrews, 2009; Grill-Spector et al., 1999; Pourtois et al., 2005; Weibert & Andrews, 2015; Xu et al., 2009), whereas others have shown varying degrees of image invariance (Davies-Thompson et al., 2013; Eger et al., 2005; Ewbank & Andrews, 2008; Loffler et al., 2005; Rotshtein et al., 2004). In a large-scale study of 80 participants, we reported image-invariant adaptation to identity in face-selective regions, such as the FFA, but no difference in the magnitude of adaptation to familiar and unfamiliar faces (Weibert et al., 2016). This fits with our current findings, where we do not find any difference between the pattern of neural response to familiar and unfamiliar faces. Overall, this suggests that the FFA does not process identity to a degree by which full image invariance is achieved. It seems more likely that the FFA is involved in a form of image normalization that contributes to face recognition. This would fit with studies of developmental prosopagnosia in which normal patterns of response in face regions can occur despite impaired face recognition (Avidan & Behrmann, 2014; Furl et al., 2011; although see Jiahui, Yang, & Duchaine, 2018). This should not, however, undermine the role of regions such as the FFA and OFA in face processing. Other studies have shown that the response in the FFA is linked with individual differences in familiar face recognition (Furl et al., 2011; Weibert & Andrews, 2015) and disruption to these regions is known to affect face recognition (Barton, 2008; Rossion et al., 2003; Jonas et al., 2012; Parvizi et al., 2012). Rather, it seems likely that interactions between the core and extended face processing networks are important for familiar face recognition (Collins & Olson, 2014; Weibert et al., 2016).

Models suggest that a dorsal pathway leading to the posterior superior temporal sulcus (pSTS) plays a key role in processing changeable aspects of faces such as emotional expression and gaze direction (Haxby et al., 2000). We found a different pattern of response in the pSTS compared to the OFA and FFA, in which there was a release from adaptation to familiar faces, but not unfamiliar faces. This increased sensitivity to familiar faces converges with previous studies that have shown that the response of the pSTS is more sensitive to familiar compared to unfamiliar faces (Davies-Thompson, Gouws, & Andrews, 2009). Although it is not clear why the pSTS is more sensitive to familiar faces, it has previously been shown that connectivity with the FFA may play a role in tracking meaningful changes in the face (Baseler, Harris, Young, & Andrews, 2014).

In conclusion, our results demonstrate that both shape and texture are used in the recognition of facial identity. These findings provide the first direct evidence for the importance of shape in a paradigm that is similar to face recognition in natural viewing. The equal sensitivity to shape and texture in the neural response of core face-selective regions provides evidence that these regions contribute to the early stages of face recognition.

Chapter 5- A narrow band of image dimensions is critical for familiar face recognition.

This Chapter is adapted from: Andrews, T. J, Rogers, D., Mileva, M., Watson, D., Wang, A., & Burton, A. M. (2023). A narrow band of image dimensions is critical for face recognition. *Vision Research, 212,* 108297.

5.1 Abstract

A key challenge in human face recognition is to differentiate information that is diagnostic for identity from other sources of image variation. Models of face processing suggest that the representation of familiar faces is based on image-invariant representations. However, it remains unclear what image properties underlie this image-invariant representation. Here, we used a behavioural approach in combination with principal components analysis (PCA) to reveal the critical image dimensions for face recognition. First, PCA was used to reveal the image dimensions of a large set of naturally varying faces. To determine which image dimensions were important for recognition, images of familiar faces were manipulated to remove specific combinations of principal components. Participants performed a naming task on the faces. We found that recognition of familiar faces increased when the early image dimensions were removed, decreased when intermediate dimensions were removed, but then returned to baseline recognition when only later dimensions were removed. Next, we asked what information is important when learning new identities. To do this, we employed a face learning paradigm using images that have had specific image dimensions removed. We found that subsequent recognition of newly learned identities, improved when the early image dimensions were removed, decreased when intermediate dimensions were removed and was not affected when the later image dimensions were removed. Together, these findings suggest that early image dimensions reflect ambient changes, such as changes in viewpoint or lighting, that do not contribute to face recognition. However, there is a narrow band of image dimensions that are critical for the recognition of identity and during face learning.

5.2 Introduction

The ability to recognise a person from their face is fundamental to the way we interact with them. Models of face processing propose that faces are first represented in a pictorial code that contains detailed information about the image, but is then transformed into a more abstract structural code that can be used for perception (Bruce & Young, 1986, 2012). This transformation from a pictorial to a structural representation is important because, as we interact with faces in a natural environment, the shape and texture of a face can vary dramatically due to movement of the head and changes in lighting. To be useful, the cognitive processes involved in recognition must be able to ignore these ambient image changes to reveal an invariant, structural representation that can be utilised for recognition (Burton, 2013).

The distinction between familiar and unfamiliar faces demonstrates the transformation from a pictorial to a structural code. While photographs of unfamiliar faces can be remembered and later recognised remarkably well, recognition performance with unfamiliar faces degrades as soon as any changes are made between learnt and test images (Bruce, 1982; Hancock, Bruce, & Burton, 2000; Kemp, Towell, & Pike, 1997). In contrast, the behavioural hallmark of familiar face recognition is that it is remarkably stable across substantial changes in expression, viewing angle, and lighting conditions (Bruce, 1994; Bruce & Young, 2012; Burton, 2013). Models of face recognition propose key structural representations for familiar faces that are known as Face Recognition Units (FRUs), which selectively respond to faces from a particular identity (Bruce & Young, 1986; Burton et al., 1990). Although classical models of face recognition (Bruce & Young, 1986) recognise the importance of FRUs, the nature of the visual properties that are used in this structural code are not specified.

A number of studies have used principal components analysis (PCA) to explore how the image properties of faces can be related to perception (Turk & Pentland, 1991; O'Toole et al., 1993; Hancock, Burton & Bruce, 1996; Burton, Bruce & Hancock, 1999; Calder et al., 2001; Jozwik et al., 2022). When PCA is applied to a set of faces, it delivers a number of dimensions that that can characterise any face image (Scheuchenpflug, 1999; Tredoux et al., 2002; Nestor et al., 2013). Within this framework, 'early' dimensions capture the most variance within a learning set, and tend to be associated with coarse-scale image variation, such as changes in head orientation or whether an image is brighter on one side or the other (Burton, Kramer, Ritchie & Jenkins, 2016). Later components, capturing progressively smaller variance, tend to capture finer-scale information. When PCA is applied to a particular set of faces, it is important to note that, given a sufficiently large sample, the resulting space generalises well. So, components derived from one set of faces, tend to capture the variance of novel sets adequately – particularly if training sets incorporate a range of variation in face images. However, it remains unclear whether and to what extent certain image dimensions are more useful than others for capturing identity information, and how certain components contribute to the process of learning.

In a recent study, Andrews and colleagues (Andrews et al., 2023) used PCA to explore the role of different image dimensions in face recognition. They found that the ability to predict judgements of identity in sorting or matching tasks was improved when the early image dimensions were removed. They also found that the recognition of familiar faces increased when the early image dimensions were removed. In contrast, there was a decrease in performance when intermediate image dimensions were removed. Finally, there was no significant effect when only later dimensions were removed. Together, these findings suggest that early image dimensions may reflect changes in the image that do not contribute to face recognition, but there is an intermediate band of image dimensions that are critical for the structural representation that is important for the recognition.

These findings suggest that the structural representation of faces that is used for the recognition of identity focuses on invariant properties of the face and ignores ambient changes in the image. However, an important related question is how these structural representations emerge. One possible mechanism could be that an invariant structural representation reflects face averages. Support for this possibility is evident in studies showing a behavioural advantage for images of familiar faces based on an average of images when compared to single images and the fact that the size of these effects increase as more images are incorporated into the average (Burton, Jenkins, Hancock & White, 2005; Jenkins & Burton, 2011). It would appear that each image of a face helps to modify the representation, by strengthening prototypical properties, while discarding ambient properties of images. Thus, exposure to multiple images of an identity could create an average face representation that underpins our recognition of familiar faces.

Variability in exposure has been shown to be fundamental to learning new faces (Devue & de Sena, 2023). Several studies have shown that, when observers learn faces, they are better able to recognise previously unseen faces when the learnt faces are more variable (Murphy et al., 2015; Baker et al. 2017; Ritchie & Burton, 2017; Kramer et al., 2018). For example, when participants were taught to associate names with faces of unfamiliar individuals, participants were more accurate and quicker in verifying the names of identities they had learned with high variability compared to those learned with low variability (Ritchie & Burton, 2017). Taken together, these results are consistent with the idea that an average face representation could underpin our recognition of faces. If this average representation is built using face images that are more representative of the natural variation of a person, then there will be an advantage in the recognition of new instances of the person. However, it has been argued that face averages may not underpin the structural representation utilised during familiar face recognition. For example, when exploring the concept face likeness- which captures the degree to which a face image looks like the person that is depicted, face averages have been found to be rated poorer for face likeness than exemplar images (Ritchie, Kramer & Burton, 2018). Moreover, only small correlations are observed between likeness ratings and prototypicality, indicating that these two measures are not suitable proxies for one another (Balas, Sandford & Ritchie, 2023).

While it is clear that exposure and variation within images are important for face learning, a few questions still remain. For example, what image dimensions are important for this process? Under the logic of face averages, all image variation is important to learn, however, it is possible that certain image dimensions contribute to the recognition of identity more than others, and thus face learning is reliant on a process of extracting the variability in a subset of image properties. Moreover, what are the temporal dynamics of face learning? Previous face learning paradigms often employ just one learning session to familiarise participants to new identities prior to testing (Murphy et al., 2015; Ritchie & Burton, 2017; Baker, Laurence & Mondloch, 2017). Therefore, it is possible that at test, participants rely on short-term memory strategies as opposed to stored mental representations of the newly learned identities.

Here, we investigated which image dimensions from a PCA of shape and texture are important for familiar face recognition and face learning. The relative importance of different dimensions on the recognition of identity was determined by measuring the effect on recognition of removing different combinations of dimensions from famous faces (Experiments 1 & 2) and the ability to recognise new instances of learnt faces (Experiment 3). In Experiments 1 and 2, UK participants were presented with images of familiar faces (UK celebrities) that had different combinations of PCs removed, using a free recall face naming paradigm. In Experiment 3, participants were asked to learn unfamiliar identities across two learning sessions using images that had different combinations of PCs removed. Participants were then asked to recognise images of these newly learned identities, using images that had no PCs removed.

5.3 Methods

5.3.1 Participants

Participants were recruited separately for each behavioural experiment. For each experiment we computed an a-priori power analysis (α = .05, power level = 0.8) using G*Power (3.1.9.7, Faul et al., 2007) to determine the minimum sample size required to find an effect (if one was present) for each experiment. We recruited 99 participants (61 female, mean age: 25.4) for Experiment 1 (repeated measures ANOVA-within factors, 4 measurements, η_p^2 = .02 indicating a small expected effect size; Cohen, 2013) and 78 participants (46 female, mean age: 26.4) for Experiment 2 (repeated measures ANOVA-within factors, 5 measurements, η_p^2 = .015 indicating a small expected effect size; Cohen, 2013). We recruited 102 participants (71 female, mean age: 22.4) for Experiment 3 (repeated measures ANOVA-within factors, 5 measurements, η_p^2 = .06 indicating a medium expected effect size; Cohen, 2013). All participants had normal or corrected to normal vision and were drawn from an opportunity sample of students and staff at the University of York. All participants gave their written informed consent. The study was approved by the Psychology department Ethics Committee.

5.3.2 Principal Components Analysis (PCA) and image stimuli

The familiar face images used for Experiment 1 and 2 were A-list celebrities, most of who are well-known Hollywood actors/actresses. The raw images entered into the PCA were collected using a Google Image Search by entering the name of a celebrity and downloading images classified as "large" by the search engine (size of 900 x 900 pixels and above) where the face was broadly front-facing and no part of it being obstructed (e.g. by other parts of the body, clothing or accessories). Apart from these criteria, the images varied naturally across lighting, emotional expressions, hairstyle, facial hair, etc.

To approximate natural variation across faces and represent our pre-existing experience with faces in daily life, PCA was performed on a large image set containing 6100 images (see the 'background set' described in Mileva et al., 2020). The set contained a varying number of images for each identity (between 1-170 images) in order to simulate different levels of familiarity and all images were ambient, capturing variability across age, pose, lighting conditions, emotional expressions, image quality, and ethnicity. Images were rescaled to 380 x 570 pixels. To be consistent across all image sets, we converted all images to greyscale. The shape of each image was determined by aligning 82 fiducial points to each face using the InterFace software package (Kramer, Jenkins & Burton, 2016). The x, y coordinates from each image were then entered into the principal components analysis for shape. The texture of each face was generated by warping each image to a standard shape. The intensity values of each pixel within the standard shape were then entered into a principal components analysis of texture. This procedure generated principal components that captured the ways in which images in the set varied, both in terms of shape and texture. We used the first 100 PCs which explained 99.9% of the shape variance and 91.6% of the texture variance.

The InterFace software package (Kramer et al., 2016) was used to perform different manipulations to each image in order to neutralise the effect of a small number of shape and texture PCs. Shape and texture PCs were neutralised as both of these properties have been shown to be important for making recognition judgments, with texture being the dominant property utilised (Rogers, et al., 2022). This was done by assigning a value of 0 to each shape and texture component within the specified range. Three different manipulations were applied to create images for the first experiment, neutralising the effect of both shape and texture PCs 1-4, 1-8, and 1-12. Experiments 2 and 3 used narrower ranges of PCs (1-3, 4-6, 7-9, and 10-12) to more precisely determine how these PCs related to face identity. All other PCs were left intact.

In Experiment 1, we used 24 familiar face images (8 female). Figure 5.1. shows the 4 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 4 (PCs 1-4 removed), 8 (PCs 1-8 removed), 12 (PCs 1-12 removed). This gave a total of 144 (24 * 4) images. From these images, we created 4 stimulus sets in which there were 6 images from each of the 4 conditions giving a total of 24 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

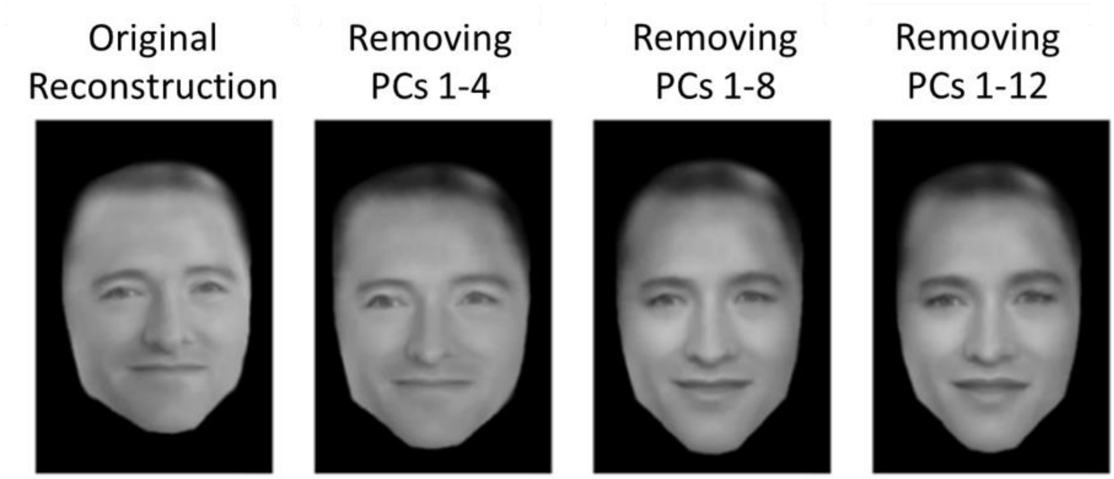

Figure 5.1. *The effect of removing different bands of principal components from an example familiar face image (Hugh Jackman).*

In Experiment 2 (Figure 5.2) we used 20 familiar face images. There were 5 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 1-3 (PCs 1-3 removed), 4-6 (PCs 4-6 removed), 7-9 (PCs 7-9 removed) and 10-12 (PCs 10-12 removed). This gave rise to at total of 100 (20 * 5) images. From these images, we created 5 stimulus sets in which there were 4 images from each of the 5 conditions giving a total of 20 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

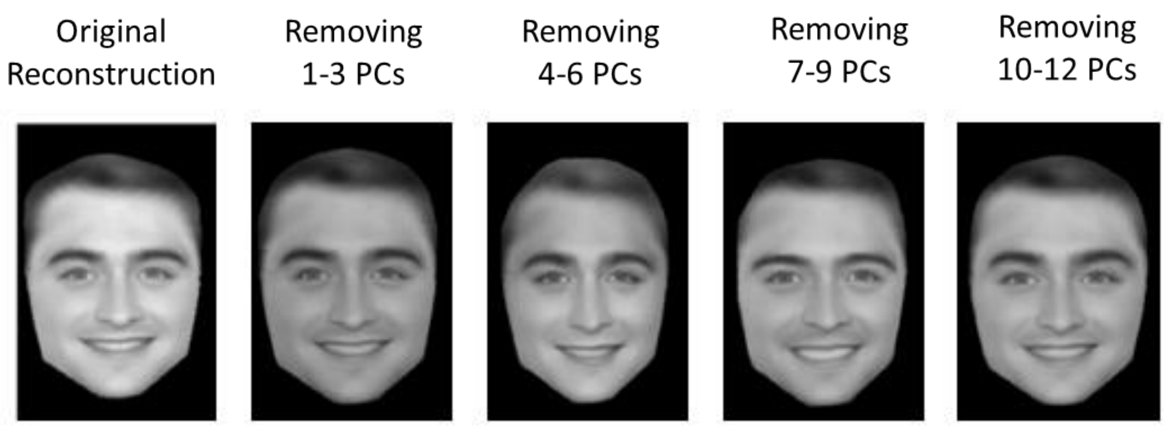

Figure 5.2. *The effect of removing different bands of principal components from an example familiar face image (Daniel Radcliffe).*

In Experiment 3 (Figure 5.3) we used images from five unfamiliar identities. There were 5 image conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 1-3 (PCs 1-3 removed), 4-6 (PCs 4-6 removed), 7-9 (PCs 7-9 removed) and 10-12 (PCs 10-12 removed).

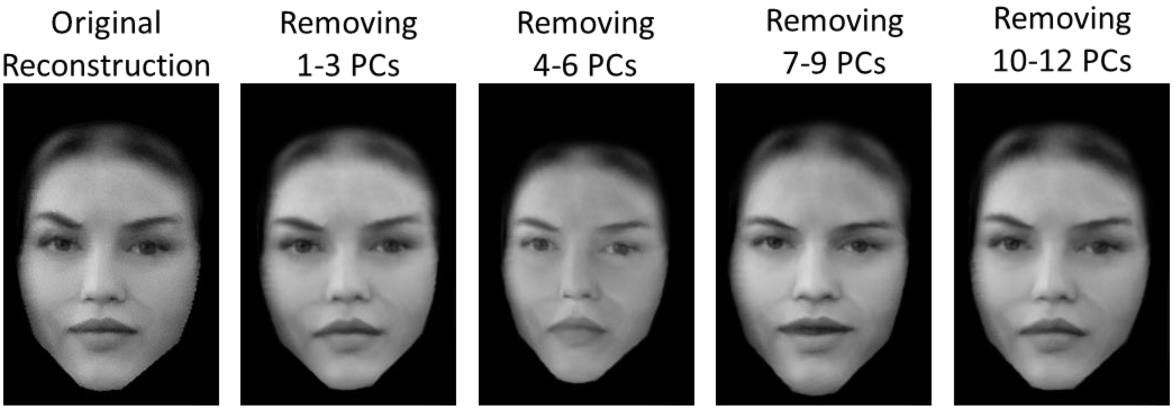

Figure 5.3. *The effect of removing different bands of principal components from an example unfamiliar face image.*

5.4 Familiar Face Recognition Task (Experiments 1 & 2)

5.4.1 Methods

The familiar face recognition task comprised two experiments involving naming familiar faces. Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Each trial began with a white fixation cross superimposed on a grey background for 0.5 seconds. This was followed by a centrally positioned face. Participants pressed one of two buttons to indicate if the face was familiar or unfamiliar. Participants were instructed to respond as quickly and as accurately as possible. If participants indicated that the face was familiar a new screen would appear containing a response box for participants to type the name or biographical information of the person. When this was complete, a new trial began. If participants indicated that the face was unfamiliar a new trial began immediately. The order in which the faces were presented was randomised for each participant. After each experiment, participants completed a familiarity check to test their ability to recognise the familiar faces, in which novel high-resolution images from each identity was presented to participants and their task was to name the identity depicted in each image.

The responses for the familiarity check were cross referenced with the responses given for the main experiment. For each participant, the identities that were not recognised in the familiarity check were automatically removed from the main analysis. Participants entered biographical information about the person (instead of their name) 1.5% of the time for Experiment 1 and 2.8% of the time for Experiment 2. Biographical information was judged to be a match if it was deemed specific enough to the target identity, for example a description of “actor”, “musician” or “politician” would result in a non-match (even if these labels were true) but a description of “actor who played Harry Potter” was deemed specific enough for a match. 86.6% of the faces in Experiment 1 and 93.5% of the faces in Experiment 2 were recognised during the familiarity check. Accuracy and response time were calculated from these trials.

5.4.2 Results

In Experiment 1, we measured the recognition of familiar faces in which different numbers of PCs were removed from the image. There were 4 conditions in which either the first 0, 4, 8 or 12 PCs from both shape and texture were removed from the image. Figure 5.4 shows the accuracy and response time for each condition. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(2.44, 239.1) = 225.7$, $p < .001$, $\eta^2_p = .70$) and response time ($F(2.41, 236.6) = 619.2$, $p < .001$, $\eta^2_p = .86$). Planned comparisons showed that there was a significant difference between the 0 and 4 PCs, which was due to an increased recognition accuracy ($t(98) = 2.72$, $p = .008$, $d = .27$) and a decreased response time ($t(98) = 5.94$, $p < .001$, $d = .60$) for the 4 PC condition. There was also a significant difference between the 0 and 8 PC conditions, and between the 0 and 12 PC conditions for accuracy and response time. However, these differences were due to a decrease in accuracy (0:8, $t(98) = 11.43$, $p < .001$, $d = 1.15$; 0:12, $t(98) = 21.04$, $p < .001$, $d = 2.11$) and an increase in response time (0:8, $t(98) = 26.44$, $p < .001$, $d = 2.66$; 0:12, $t(98) = 28.47$, $p < .001$, $d = 2.86$). These data show that removal of the initial PCs improves accuracy and reduces response time, whereas removal of later PCs reduces accuracy and increases response time.

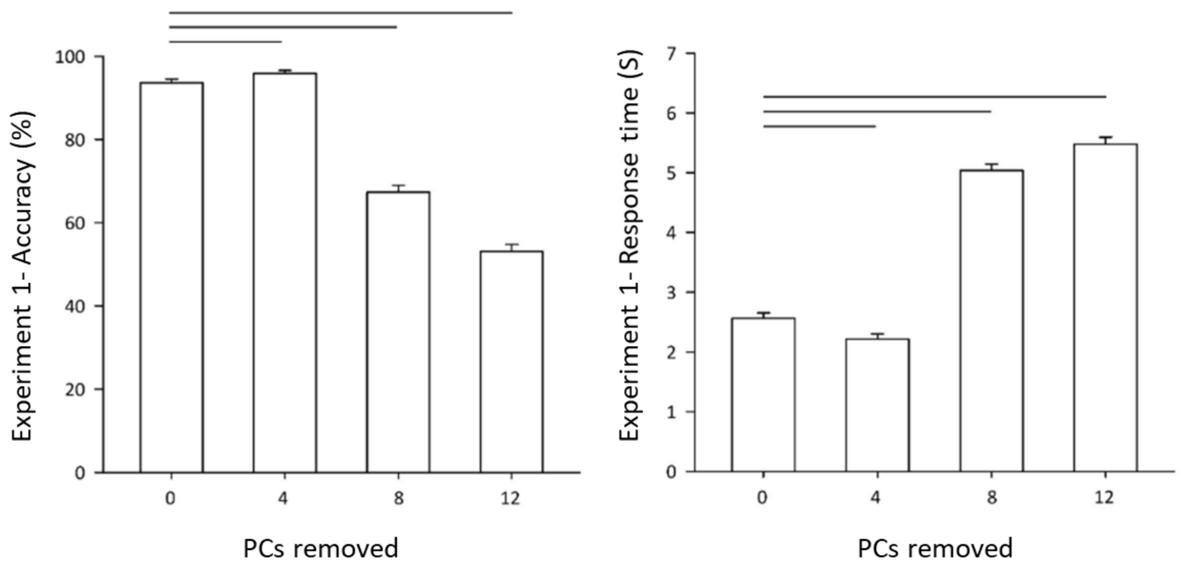

Figure 5.4. *Experiment 1 – The effect of cumulatively removing PCs on familiar face recognition. Removal of 4 PCs resulted in increased accuracy and a reduction in response time to familiar faces. However, removal of 8 or 12 PCs resulted in decreased accuracy and increased response time. Horizontal lines indicate significant differences (p < .05) relative to the 0 PCs condition (original reconstruction). Error bars indicate standard error of the mean.*

In Experiment 2, we investigated the effect of removing bands of PCs on familiar face recognition. Participants viewed images in which 0, 1-3, 4-6, 7-9 or 10-12 PCs of shape and texture were removed from the image. Figure 5.5 shows the accuracy and response time for each condition. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(4, 304) = 13.03$, $p < .001$, $\eta^2_p = .148$) and response time ($F(4, 304) = 5.36$, $p < .001$, $\eta^2_p = .092$). Planned comparisons showed that there was a significant difference between the 0 and 1-3 PC conditions, which was again due to an increased recognition ($t(76) = -1.52$, $p = .034$, $d = .18$) but there was no difference in response time ($t(78) = 0.48$, $p = .630$, $d = .057$) for the 1-3 PC condition. There was also a significant difference between the 0 and 4-6 PCs conditions, and also between the 0 and 7-9 PCs conditions for accuracy and response time. These differences were due to a decrease in accuracy (0:4-6, $t(78) = 4.12$, $p < .001$, $d = .47$); 0:7-9, $t(78) = 4.07$, $p < .001$, $d = .47$) and an increase in response time (0:4-6, $t(78) = -2.20$, $p = .031$, $d = -.26$; 0:7-9, $t(78) = -3.00$, $p = .004$, $d = -.37$)). Finally, there was no significant difference in accuracy or response time between the 0 and 10-12 PCs removed conditions (($t(78) = 1.47$, $p = .147$, $d = .17$); ($t(78) = 0.65$, $p = .519$, $d = .08$)). These data show that removal of the initial PCs improves accuracy and reduces response time, whereas the selective removal of intermediate bands of PCs reduces accuracy and increases response time. Finally, removal of later bands of PCs has a minimal effect on recognition.

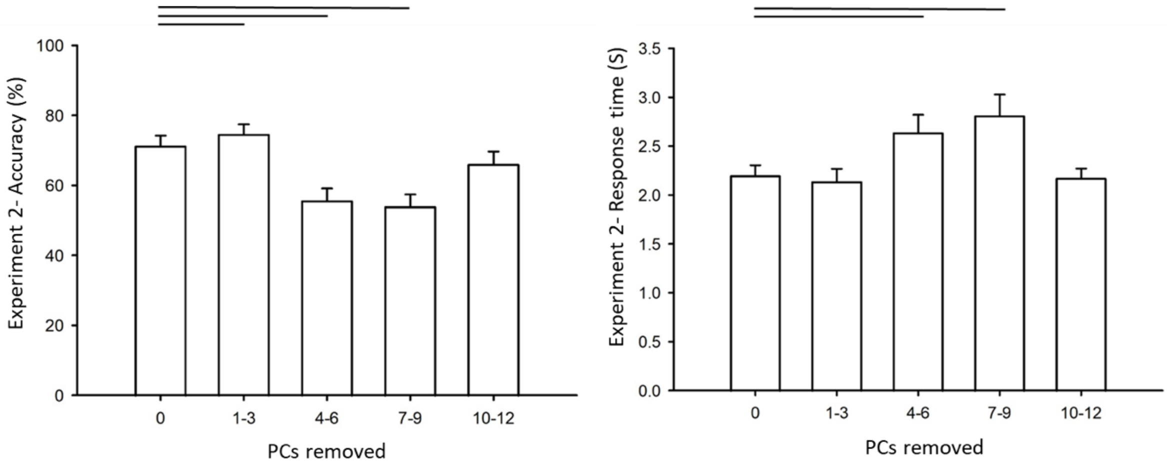

Figure 5.5. *Experiment 2 – The effect of selectively removing bands of PCs on familiar face recognition. Removal of the PCs 1-3 again resulted in increased accuracy but not a reduction in response time to familiar faces. Removal of 4-6 or 7-9 PCs resulted in a significant decrease in accuracy and an increased response time. However, removal of PCs 10-12 had no effect on accuracy or response time.*

5.5 Face Learning Task (Experiment 3)

5.5.1 Methods

The face learning paradigm comprised of three sessions spread out equally across five days (Monday, Wednesday and Friday). The first two sessions consisted of a learning phase and feedback phase. The final session involved a face recognition test of the newly learned identities. Participants completed all sessions online using the Pavlovia platform (PSYCHOJS, Version 2020.2).

During each learning phase, participants were asked to learn five unfamiliar female identities. Participants learned each of the five identities using images from one of the different image conditions (0 PCs removed, 1-3 PCs removed, 4-6 PCs removed, 7-9 PCs removed, 10-12 PCs removed). This was counterbalanced across the participant sample, so that each identity was learned in each of the image conditions. For the learning and feedback phases, participants were shown a total of 14 different images (5 for each learning phase, which were repeated twice and 2 for each feedback phase), where each image was in the image condition for that identity, for that group of participants.

During the learning phases, each trial began with a white fixation cross superimposed on a grey background for 0.5 seconds. This was followed by a centrally positioned face along with a corresponding name for that identity, displayed for 5 seconds. Names were selected from the most popular first names for baby girls in 2020 using birth registration data from the ONS. The selection criteria were the first five names that were two syllables and six letters in length (Harper, Millie, Phoebe, Sophie and Willow). Participants were instructed to learn the facial identity as well as the face and name association. Trials were blocked by identity, but the order of images within a block along with the order of the blocks was randomised for each participant.

After each learning phase participants then entered a feedback phase in order to ascertain the levels of familiarity reached for each identity. Participants were shown 2 new reconstructed images of each identity, as well as the five names of the newly learnt identities. Participants were instructed to press a key to indicate the identity they thought was depicted in the image. Automated feedback was given on each trial, when the participant made an incorrect identification, the correct name for the identity was shown. Each feedback phase was self-paced by the participant. 62.6% of the faces were recognised successfully in the first session and this rose to 73.5% in the second session.

The final session of this experiment was a face recognition test. During this session participants were shown novel images of the newly learned identities as well as foil image distractors. The images used during the final test phase were not manipulated to remove any PCs, for example; if a participant had learned 'Willow' with images that had 4-6 PCs removed, during the test session they had to recognise Willow from images that had no PCs removed. This was important in order to observe how learning facial identities with images that had PCs removed, impacted subsequent recognition of novel images that had no PCs removed but were captured within the PC space. During the test session 5 novel different images were used for each identity (25 in total, 5*5) along with 50 foil distractor images (5 images of 10 unfamiliar identities). Each trial began with a white fixation cross superimposed on a grey background for 0.5 seconds. This was followed by a centrally positioned face. Participants pressed one of two buttons to indicate if the face was familiar or unfamiliar. This task was self-paced and participants were instructed to respond as quickly and as accurately as possible. If participants indicated that the face was familiar, a new screen would appear containing a response box for participants to type the name of the person. When this was complete, a new trial began. If participants indicated that the face was unfamiliar a new trial began immediately. The order in which the faces were presented was randomised for each participant.

5.5.2 Results

In Experiment 3, we asked how learning identities with images that had select bands of PCs removed would affect subsequent recognition of novel images that had no PCs removed. Figure 5.6 shows sensitivity (d prime) and response time (s) for making a correct recognition judgment when learning an identity under each image condition. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(4, 404) = 10.91$, $p < .001$, $\eta^2_p = .10$) and response time ($F(2.88, 175.41) = 7.52$, $p < .001$, $\eta^2_p = .11$). Planned comparisons showed that there was a significant difference between the 0 and 1-3 PC conditions, which was due to an increase in recognition ($t(101) = -2.07$, $p = .002$, $d = .21$) for the 1-3 condition, but no difference in response time ($t(74) = -.58$, $p = .567$, $d = .07$). There was also a significant difference between the 0 and 4-6 & 7-9 conditions. These differences were due to a decrease in accuracy (0:4-6, $t(101) = 3.13$, $p = .002$, $d =.03$); (0:7-9, $t(101) = 2.89$, $p = .005$, $d = .29$) and an increase in response time (0:4-6, $t(74) = -3.13$, $p = .003$, $d = .26$; 0:7-9, $t(73) = -3.71$, $p <.001$, $d = .43$)). Finally, there was no significant difference in accuracy and response time between the 0 and 10-12 conditions ($t(101) = -.14$, $p = .889$, $d = .01$);($t(80) = -2.0$, $p = .053$, $d = .22$)). These data show that removal of the initial PCs during face learning improves accuracy and reduces response time, whereas the selective removal of intermediate bands of PCs during face learning reduces accuracy and increases response time, when tested on images that have had no PCs removed. Finally, removal of later bands of PCs has a minimal effect on recognition.

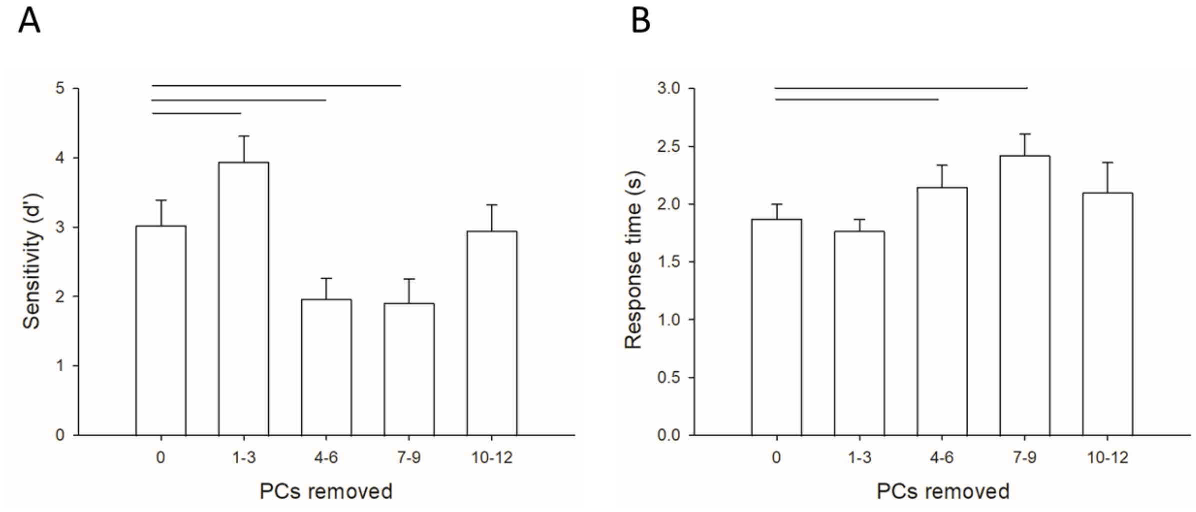

Figure 5.6. *In Experiment 3, removal of the PC band 1-3 from learning images, resulted in an increase in final recognition sensitivity (A) but no difference in response time (B). Removal of the mid bands of PCs (4-6 and 7-9) resulted in a significant decrease in accuracy and an increased response time. However, the removal of a later band of PCs (10-12) had no effect on recognition sensitivity and a limited effect on response time. Horizontal lines indicate significant differences ($p < .05$) relative to the 0 PCs condition (original reconstruction), and error bars represent +1 SEM.*

5.6 General Discussion

The aim of this study was to determine what information is necessary for the recognition of familiar faces, and for the learning of new identities. To address these issues, a principal components analysis was used to reveal the underlying image dimensions of naturally varying face images from different identities. This allowed us to remove certain image dimensions from the face images to compare the importance of these image dimensions on the recognition of familiar identities and when learning of new identities. Our key finding is that the structural representation used for the recognition of identity from faces is dependent on a narrow band of image dimensions.

As faces have a similar structure, the ability to discriminate identity must be based on encoding subtle differences between images. A further challenge for successful face recognition is that, as a result of changes in viewing conditions, each face can generate an almost infinite number of images. So, it is necessary for the recognition system to differentiate between information in the image that provides cues about identity from other information that does not. Models of face processing propose that information about faces is first represented in an image-based or pictorial code, which is then transformed into a structural code that can be used for recognition (Bruce & Young, 1986, 2012; Burton et al., 1990). However, the precise image properties that are used in this structural code have not been fully resolved. The concept of face space provides a framework for explaining how variance across faces (both within and between identities) might be represented in a structural code that is used for recognition (Valentine, 1991; Valentine et al., 2016). Within this framework, different properties of the face are represented along different dimensions. Each face is represented by a location in this multidimensional space, such that faces that are close together are perceived to be more similar and those that are separated by larger distances are perceived to be more different. Nevertheless, it has not been clear how many dimensions are important or what they might represent, nor has it been established how images of the same person could occupy different locations within a face space.

In the first experiment, we found that removing the early image dimensions increased recognition and decreased response time. In contrast, the cumulative removal of more PCs reduced recognition and increased response time. This suggests that, while early image dimensions are not important (in fact removing this information increases recognition), later PCs are important for making identity judgments. However, these findings did not show which PCs are important for recognition. To address this question, Experiment 2 asked whether there are bands of image dimensions that are important for recognition. Again, we found that removing the early PCs improved recognition compared to when no PCs were removed. However, we found that removing intermediate bands of PCs resulted in a significant decrease in the recognition accuracy and increased response time. Interestingly, removing later (10-12) PCs had a minimal effect on recognition. These findings suggest that the initial PCs reflect ambient image information that is not used for recognition. However, there is an intermediate band of PCs that plays a key role in recognition. These intermediate image dimensions could reflect some of the key dimensions within a multidimensional face space model that is important for recognition.

A key and surprising finding from this study was that removal of the initial image dimensions or PCs improved recognition of familiar faces. Previous studies have shown that texture information is important for face recognition (Bruce & Langton, 1994; Russell et al., 2006;

Harris et al., 2014; Andrews et al., 2016; Rogers et al., 2022) and that the recognition of faces also becomes much more difficult when texture is removed from the image (Davies et al., 1978; Leder, 1999; Burton et al., 2005). However, the increased recognition of identity when the initial image dimensions were removed shows that not all texture information contributes to recognition. Presumably, these early image dimensions reflect ambient changes in the texture (e.g. illumination) that are not diagnostic of an identity. The shape or configuration of the face has also been suggested to be important for face recognition (McKone & Yovel, 2009; Tanaka & Gordon, 2011; Piepers & Robbins, 2012; Rogers et al., 2022). However, natural variation in face images caused by rigid changes in viewpoint or non-rigid changes (such as in expression or during speech) can often lead to large changes in the configuration or shape of the face, without changing identity. Our findings show that when the early PCs are removed from familiar face images, recognition of identity increases. This suggests that early principal components for shape reflect changes in viewpoint, which is not diagnostic of identity.

The largest effect on recognition was found when we removed intermediate image dimensions. This suggests that these dimensions are critical when making identity judgments of faces. Burton and colleagues (2016) investigated which aspects of the image were related to different PCs. They found that the early principal components were typically related to rigid head rotations or changes in lighting. Although later PCs were much harder to define, they tended to reflect non-rigid changes in shape or changes in texture that are not related to lighting. They also found that variance in these later PCs was idiosyncratic. For example, the same PC could reflect a different image property in different identities. It is important to note that in their study, the PCA analysis and image reconstructions were performed within individual identities, so is dependent on the image set. However, when PCA is performed using large image sets that incorporates variation across individuals, the PCs can remain relatively stable. In any event, our results show these intermediate dimensions are critical for making judgments of identity.

In the final experiment, we asked which image dimensions are important when learning new faces. We found that removing the early band of PCs from learnt images led to an increase in recognition. In contrast, removal of intermediate image dimensions from the learnt images decreased face recognition. This indicates that these dimensions are critical not only for the recognition of familiar faces, but also during the process of becoming familiar with a face. A key aspect of the learning paradigm employed in this experiment was that the learning phase of the experiments was not immediately prior to the recognition test. Thus, participants had to rely on their newly acquired stored mental representations of the identities. Participants recognition of the newly learned identities was then tested using images that had no PCs removed. This was important in order to assess how certain principal components contributed to the process of face learning. We also used a free recall naming task to provide a more real-world test of recognition. These differences in design provide a more ecologically valid approach to understanding the process by which we become familiar with faces.

Previous face learning research has shown that variability in the exposure of an identity is fundamental to facilitate the generation of a view invariant representation (Jenkins et al., 2011; Murphy et al., 2015; Ritchie & Burton, 2017). This fits with the idea that stable face representations could be formed by averaging across multiple instances of a face (Burton, Jenkins, Hancock & White, 2005; Jenkins & Burton, 2011). This has the effect of removing natural or ambient fluctuations in the image, such as pose, illumination and expression. Our findings extend previous research by showing that removal of the early image dimensions (containing this coarse scale image variation) improved subsequent face learning. This suggests that this information is not important for establishing a stable representation for the recognition of identity. Our data suggest that when we are learning faces in natural viewing, the presence of this ambient variation makes it more difficult to generate a structural representation that can be used for subsequent recognition. Rather, an intermediate band of image dimensions appears to be important for extracting the key face information that is common between encounters (Burton et al., 2016; Young & Burton, 2021). Interestingly, we found that removal of further image dimensions had a minimal effect on our ability to learn new faces. Thus, it could be argued that structural representations relying upon averaging all image properties are not necessary for the process of familiar face recognition and face learning. This notion reflects more current work showing that face averages generated in this way are rated as having lower face likeness than exemplar images (Ritchie, Kramer & Burton, 2018; Balas, Sandford & Ritchie, 2023). Instead, the findings here suggest that extracting and possibly averaging a smaller set of image dimensions that are critical for face recognition may underpin facial structural representations.

A key feature of our study was the use of ambient face images that reflect the image variation that occurs in natural viewing. Although the early image dimensions for both shape and texture explain most of the image variance, they do not appear to contain information that is important for the recognition of identity. In contrast, intermediate image components which represent more subtle changes in the image, appear to be important for face learning and familiar face recognition. These findings are relevant to the debate surrounding the difference between unfamiliar and familiar face perception (Young & Burton, 2018a, 2018b, 2021; Rossion, 2018; Sunday & Gauthier, 2018; Blauch, Behrmann & Plaut, 2021a; 2021b; Yovel & Abudarham, 2021). Our proposal is that an important aspect of the change from a pictorial representation that is used for unfamiliar face perception to a structural representation that is used for familiar face recognition involves the removal of ambient information in the image. The ability to recognise familiar faces would appear to depend on the ability to ignore this irrelevant information and focus on the image properties that are important for recognition. On the other hand, the difficulty in the recognition of unfamiliar faces may reflect the inability to ignore this information.

In conclusion, our results suggest that an intermediate band of image dimensions contains the structural code that is used to not only discriminate identity, but are also fundamental during the process of face learning. Recent studies in face recognition have shown that the discrimination of identity from a PCA is improved by the addition of a classifier (Kramer et al., 2017; Kramer et al., 2018). These results suggest that these classifiers may improve recognition by increasing the weight of these critical band of PCs or image dimensions. These findings provide a new perspective for understanding of the structural code that underpins the recognition of faces.

Chapter 6- The contributions of different image dimensions on the perception of gaze, gender and expression.

6.1 Abstract

Faces provide a wide range of information that help guide our social interactions. In this study, we investigated the visual information in the face that allows us to perceive gaze, gender and expression. First, PCA was used to reveal the image dimensions of a large set of naturally varying faces. To determine which image dimensions were important, images of faces were manipulated to remove specific combinations of principal components. Participants then performed behavioural tasks involving judgements of gaze, gender or expression. We found that the removal of the early PCs had a significant effect on the perception of gaze. On the other hand, the removal of intermediate bands of PCs affected the perception of gender. Finally, the removal of later PCs affected judgements of emotional expression. These findings show that distinct, but overlapping, PCs (image dimensions) in faces are important for the perception of gaze, gender and emotional expression. It remains to be established whether the neural basis of these image dimensions can be found in a generic representation of faces or whether distinct neural representations instantiate each aspect of face perception.

6.2 Introduction

Faces provide a range of information that is critical for everyday social interactions (Bruce & Young, 1986; 2018). The unique appearance of a face allows us to recognise the identity of a person. However, information from the face also allows us to extract what others are thinking and feeling. For example, a warm smile may indicate friendliness and approachability, whereas an angry expression might indicate displeasure. Faces also provide information about the age, gender and race of a person, which can influence the way we interact with them. Given the range of inferences we can make from looking at a face, a fundamental question is what information in the image do we use to make these judgements. In this study, we explored which image properties are important for the perception of gaze, gender and emotional expression.

The ability to perceive the direction of gaze from a face is important for social interaction (Perrett, Hietanen, Oram & Benson, 1992). Humans change their gaze in order to bring different objects onto the fovea where vision is most sensitive. Thus, the ability to perceive the direction of gaze allows us to know what is engaging an individual's attention. Gaze cues can be given by the position of the eyes and head direction (Langton, 2000). Studies have shown that we are very sensitive to the small changes in eye position (Cline, 1967), which is thought to be determined by comparing the luminance of the sclera either side of the pupil (Jenkins & Langton, 2003). The perception of gaze also involves the orientation of the head (Jenkins et al., 2006). For example, the Wollaston illusion shows that head orientation can change the perceived gaze even when the position of the eyes are unchanged (Wollaston, 1824). Indeed, studies using adaptation have shown that we integrate eye and head position to generate a perception of gaze (Hecht et al., 2020).

The perception of the sex or gender of a person is equally important for guiding social interactions. We can categorise the gender of a person from their face quickly and accurately. Male and female faces often differ in both shape and texture. Female faces tend to be shorter and rounder than male faces, which have more angular jawlines (Brown & Perrett 1993). Other studies have found that the relative distance between facial features, such as the eyebrows are also important for gender discrimination and recognition (Campbell et al., 1999). The texture of the face has also been shown to be important for the perception of gender (Brown & Perrett 1993; Yamaguchi et al., 1995). For example, the eyes, eye brows and lips can be used to distinguish male and female faces. The presence or absence of facial hair is another important cue for gender perception (Baudouin, 2006). Thus, it seems that the perception of gender relies on both local and global shape information, as well as texture information from key features.

The perception of facial expressions of emotion is important for understanding the internal state of other people. We are able to make different facial expressions of emotion through changes in the facial musculature (Ekman, 1992; Ekman & Cordaro, 2011; Izard, 1994; Levenson, 2011; Panksepp & Watt, 2011; Calvo & Nummenmaa, 2015). For example, a sad expression is characterised by a pulled down mouth at the corners, eyelid tightening, and a dropping of the outer eye corners. On the other hand, happiness is shown by the raising of the cheeks, lip corners and eyes. Thus, the perception of expression would appear to rely on changes in the shape or configuration of the face. Support for the critical role of shape information in the perception of facial expression is found in studies that show image manipulations affecting texture, but leaving shape information intact, have little impact on perceptual and neural responses to facial expression (Bruce & Young, 1998; Magnussen et al., 1994; White, 2001; Pallett & Meng, 2013; Harris et al., 2014). Similarly, image manipulations that completely remove texture, such as line drawings of faces, also show relatively preserved expression perception (McKelviet, 1973; Etcoff & Magee, 1992). However, other evidence suggests that the texture of the face can also be used to categorise expression (Calder et al., 2001; Sormaz, Young & Andrews, 2016). For example, the categorisation of facial expressions was equally dependent on variation in both the texture and shape properties of the image (Sormaz, Young & Andrews, 2016).

PCA can be used to measure natural variation in the image properties of faces. It generates principal components or image dimensions that capture image variation within the face. In the previous chapter, it was found that an intermediate band of images dimensions from ambient face images contributed to the recognition of identity. However, it remains unclear as to which image dimensions contribute to the perception of gaze, gender and expression. The present study explored how the removal of different principal components (PCs) affected perception. PCs were removed in one of two ways. To first establish what range of PCs are important for the perception of these categories, PCs were removed cumulatively. To then interrogate the finer grained contributions of PCs to the perception of these facial signals, PCs were selectively removed in bands. Our aim was to determine which image dimensions are critical for the perception of these different categories and to determine whether the same or different PCs contribute to the perception of different social judgements.

6.3 Methods

6.3.1 Participants

A single group of participants were recruited for the behavioural experiments. We computed an a-priori power analysis (α = .05, power level = 0.8) using G*Power (3.1.9.7, Faul et al., 2007) to determine the minimum sample size required to find an effect (repeated measures ANOVA-within factors, 8 measurements, η_p^2 = .02 indicating a small expected effect size, Cohen, 2013). We recruited 73 participants (47 female, mean age: 22.4). All participants had normal or corrected to normal vision and were drawn from an opportunity sample of students and staff at the University of York. All participants gave their written informed consent. The study was approved by the Psychology department Ethics Committee.

6.3.2 Principal Components Analysis (PCA)

The images used in Experiments 1, 2 and 3 were all unfamiliar faces that varied in gaze, gender or expression, respectively. Images for each experiment were collected using a Google Image search by entering key words such as “quarter profile face, male face, happy face” and downloading images classified as “large” by the search engine (size of 900 x 900 pixels and above). Images were not selected if the face was obstructed by other parts of the body, clothing or objects. In Experiments 2 and 3, images were only selected when the face was broadly front-facing. Apart from these criteria, the images varied naturally across lighting, hairstyle, facial hair and so on.

To approximate natural variation across faces, PCA was performed on a large image set containing 6100 images (see the ‘background set’ described in Mileva et al., 2020). The set contained a varying number of images for each identity (between 1-170 images). Images were rescaled to 380 x 570 pixels. To be consistent across all image sets, we converted all images to greyscale. The shape of each image was determined by aligning 82 fiducial points to each face using the Interface software package (Kramer, Jenkins & Burton, 2016). The x, y coordinates from each image were then entered into the principal components analysis for shape. The texture of each face was generated by warping each image to a standard shape. The intensity values of each pixel within the standard shape were then entered into a principal components analysis of texture. This procedure generated principal components that captured the ways in which images in the set varied, both in terms of shape and texture. We used the first 100 PCs which explained 99.9% of the shape variance and 91.6% of the texture variance.

The InterFace software package (Kramer et al., 2016) was used to perform different manipulations to each image in order to neutralise the effect of a small number of shape and texture PCs. Shape and texture PCs were neutralised as both of these properties have been shown to be important for making judgments of face perception (Hole & Bourne, 2010). This was done by assigning a value of 0 to each shape and texture component within the specified range. All other PCs were left intact. In each of the three experiments (Gaze, Gender and Expression), eight different manipulations were applied to create images for each experiment, neutralising the effect of both shape and texture PCs 0, 1-3, 1-6, 1-9, 1-12; , 4-6, 7-9, 10-12.

6.3.3 Stimuli

In Experiment 1, we used 96 images, where each image was of a different person. There was an equal split of left and right gaze images and an equal number of male and female images (24*4). Gaze direction ranged from approximately 10-50˚ (yaw). Figure 6.1. shows the 8 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 3 (PCs 1-3 removed), 6 (PCs 1-6 removed), 9 (PCs 1-9 removed), 12 (PCs 1-12 removed), 4-6 (PCs 4-6 removed), 7-9 (PCs 7-9 removed) and 10-12 (PCs 10-12 removed). This gave a total of 768 (96 * 8) images. From these images, we created 8 stimulus sets in which there were 12 images (6 male, 6 female) from each of the 8 conditions giving a total of 96 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

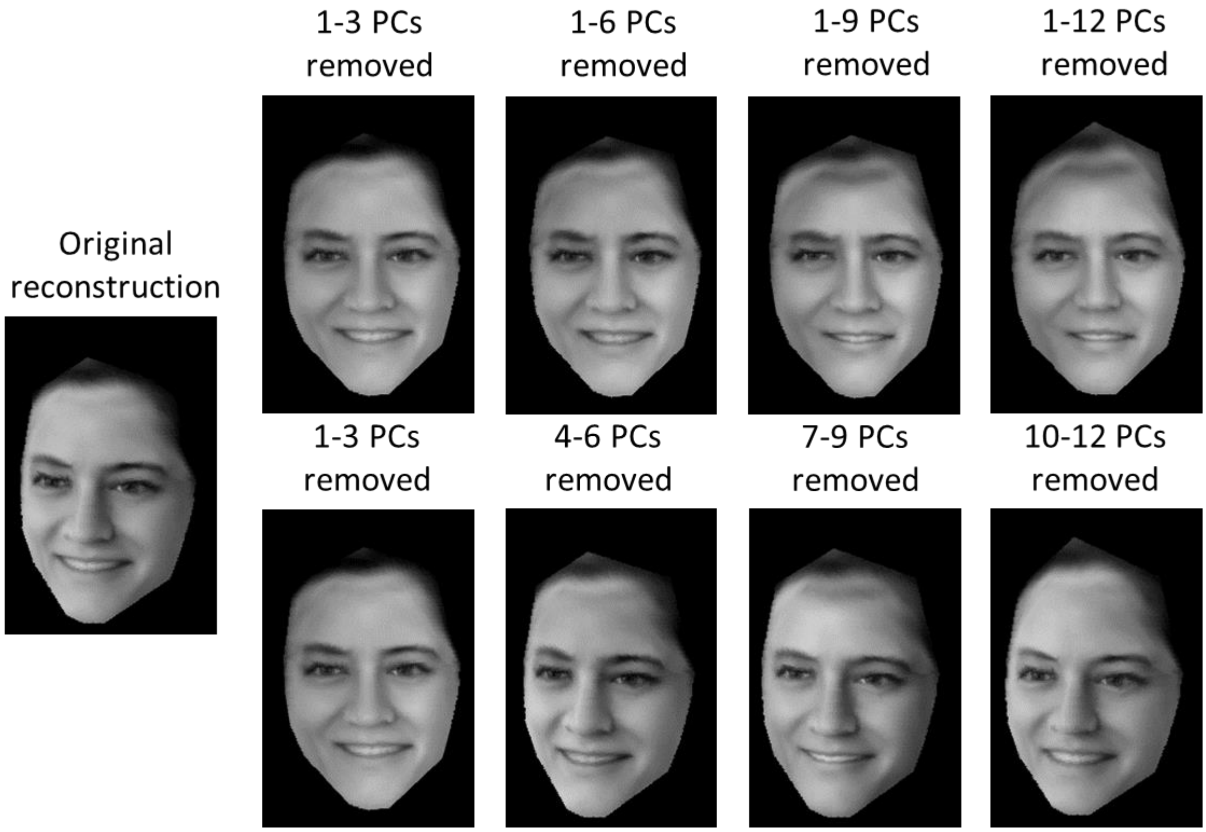

Figure 6.1. *Experiment 1: The effect of removing different principal components on the perception of gaze. The left image shows an original face with a rightward gaze. The top row of images shows the effect of removing PCs cumulatively, and the bottom row shows the effect of selectively removing PC bands.*

In Experiment 2, we used 96 images, where each image was of a different person. There was an equal number of male and female faces. Figure 6.2. shows the 8 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 3 (PCs 1-3 removed), 6 (PCs 1-6 removed), 9 (PCs 1-9 removed), 12 (PCs 1-12 removed), 4-6 (PCs 4-6 removed), 7-9 (PCs 7-9 removed) and 10-12 (PCs 10-12 removed). This gave a total of 768 (96 * 8) images. From these images, we created 8 stimulus sets in which there were 12 images (6 male, 6 female) from each of the 8 conditions giving a total of 96 images. In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

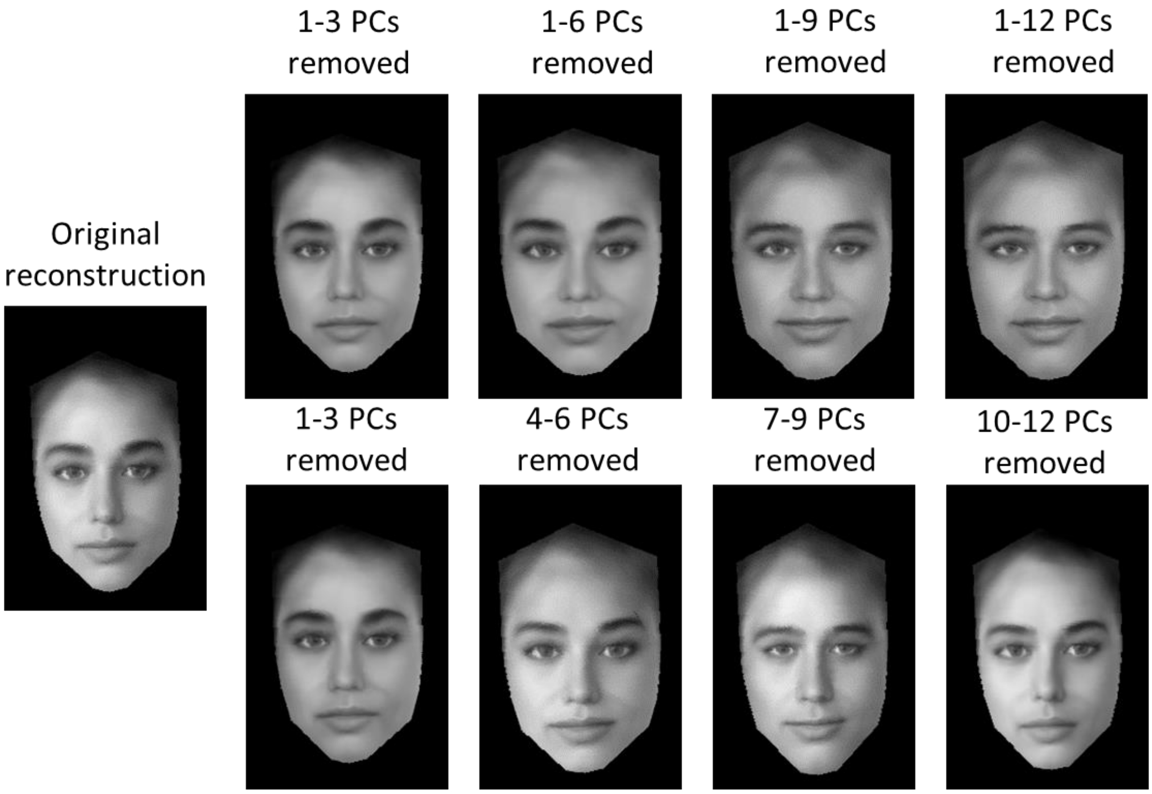

Figure 6.2. *Experiment 2: The effect of removing different principal components on the perception of gender. The left image shows an original female face. The top row of images shows the effect of removing PCs cumulatively, and the bottom row shows the effect of selectively removing PC bands.*

In Experiment 3, we used 96 images for both expressions (happy, sad). In a pilot study, we selected images that were rated high for happiness and sadness. 100 happy, 100 sad and 100 neutral images were selected from a Google Image search. There was an equal split between male and female faces. 30 participants (19 female, mean age: 21.3) rated faces that were presented in two blocks. Each block contained the 100 images of each expression and a further 50 neutral images (150 images total per block). Participants completed this experiment online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Each trial began with a white fixation cross superimposed on a grey background for 0.5 seconds. This was followed by a centrally positioned face. Participants were instructed to rate each face on the expression for that block (happy or sad) using a scale of 1-7 (where 1 equals not happy/sad at all, and 7 equals extremely happy/sad). Participants made this judgement using the number keys (1-7). Participants were instructed to respond as quickly and as accurately as possible. The order in which the faces were presented was randomised within each block for every participant. For each expression the average rating for each image was then calculated across the participants. These images were then separated by gender for each expression and the highest rated 48 images were compiled into separate image sets (giving rise to 4 sets of images- 48 male happy, 48 female happy, 48 male sad, 48 female sad). The average rating for each set was as follows Happy-Male = 5.89 (SD = 0.89); Happy-Female = 6.08 (SD = 0.78); Sad-Male = 5.67 (SD = 0.93), Sad-Female = 5.62 (SD = 0.97).

Figure 6.3. shows the 8 conditions that were created by the selective removal of shape and texture principal components from the images: 0 (no PCs removed), 3 (PCs 1-3 removed), 6 (PCs 1-6 removed), 9 (PCs 1-9 removed), 12 (PCs 1-12 removed), 4-6 (PCs 4-6 removed), 7-9 (PCs 7-9 removed) and 10-12 (PCs 10-12 removed). This gave a total of 768 (96 * 8) images for both expressions. From these images, we created 8 stimulus sets in which there were 12 images (6 male, 6 female) from each of the 8 conditions giving a total of 96 images (for both happy and sad expressions). In each stimulus set, there was only one image from each identity. Participants were allocated randomly to each image set.

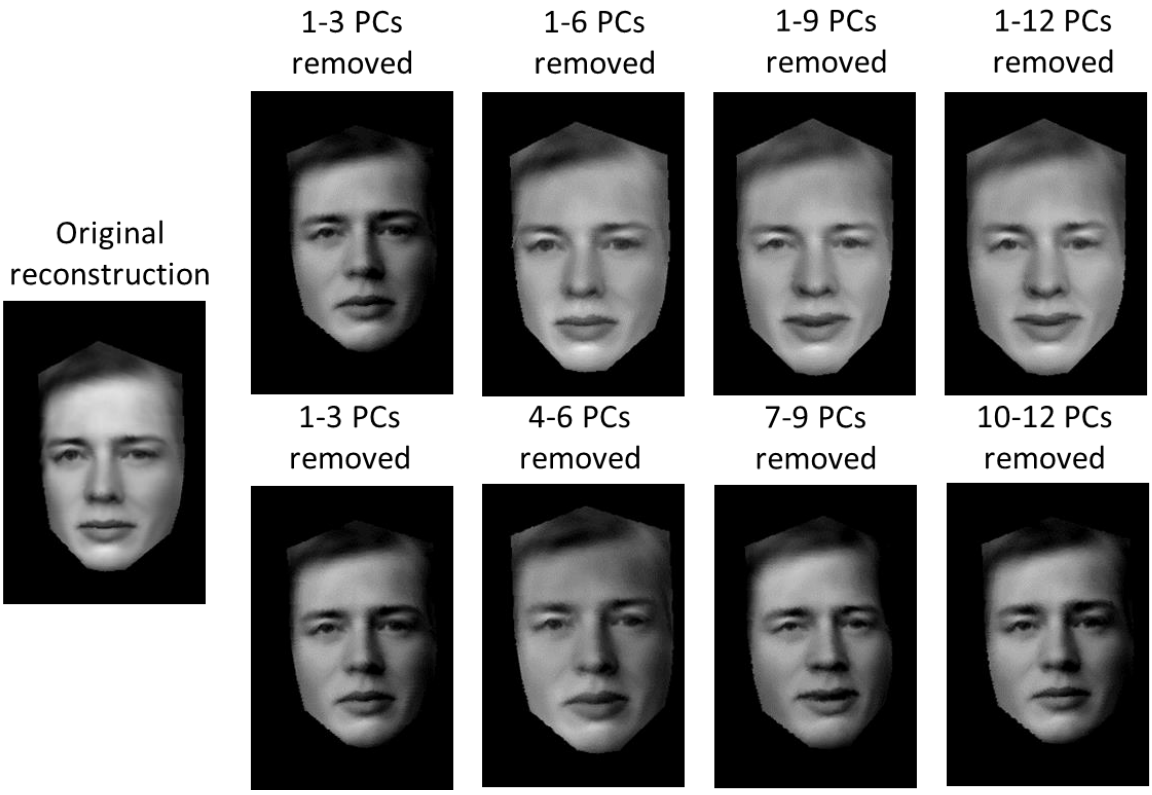

Figure 6.3. *Experiment 3: The effect of removing different principal components on the perception of expression. The left image shows an original face expressing sadness. The top row of images shows the*

effect of removing PCs cumulatively, and the bottom row shows the effect of selectively removing PC bands.

6.3.4 Procedure

Participants completed the study online using the Pavlovia platform (PSYCHOJS, Version 2020.2). Participants completed all experiments in a single session containing 384 trials (96*4- gaze, gender, happy, sad). The session contained three blocks- gaze, gender and expression. The order of the blocks and the order of trials within each block was randomised for each participant. Each trial began with a white fixation cross superimposed on a grey background for 0.5 seconds. This was followed by a centrally positioned face. For each trial, participants made a two-alternative forced choice decision by pressing one of two buttons (left and right arrow keys) to indicate if the face was looking to the left/right (gaze), if the face was male/female (gender), or if that face was happy/sad (expression). On screen labels were present for each trial, reminding participants what each arrow key represented. Participants were instructed to respond as quickly and as accurately as possible and response times were recorded to make this decision.

6.4.1 Experiment 1 – Gaze

In Experiment 1, we measured the perceived direction of gaze from faces in which different numbers of PCs were removed from the image. Participants were asked to judge whether the face was looking to the left or to the right using a 2AFC paradigm. There were eight image conditions in which we removed principal components cumulatively or selectively. Figure 6.4(A,B) shows the average accuracy in performance for each image condition (with chance level being 50%), along with the average response time to make a correct judgment (Figure 6.4 C, D).

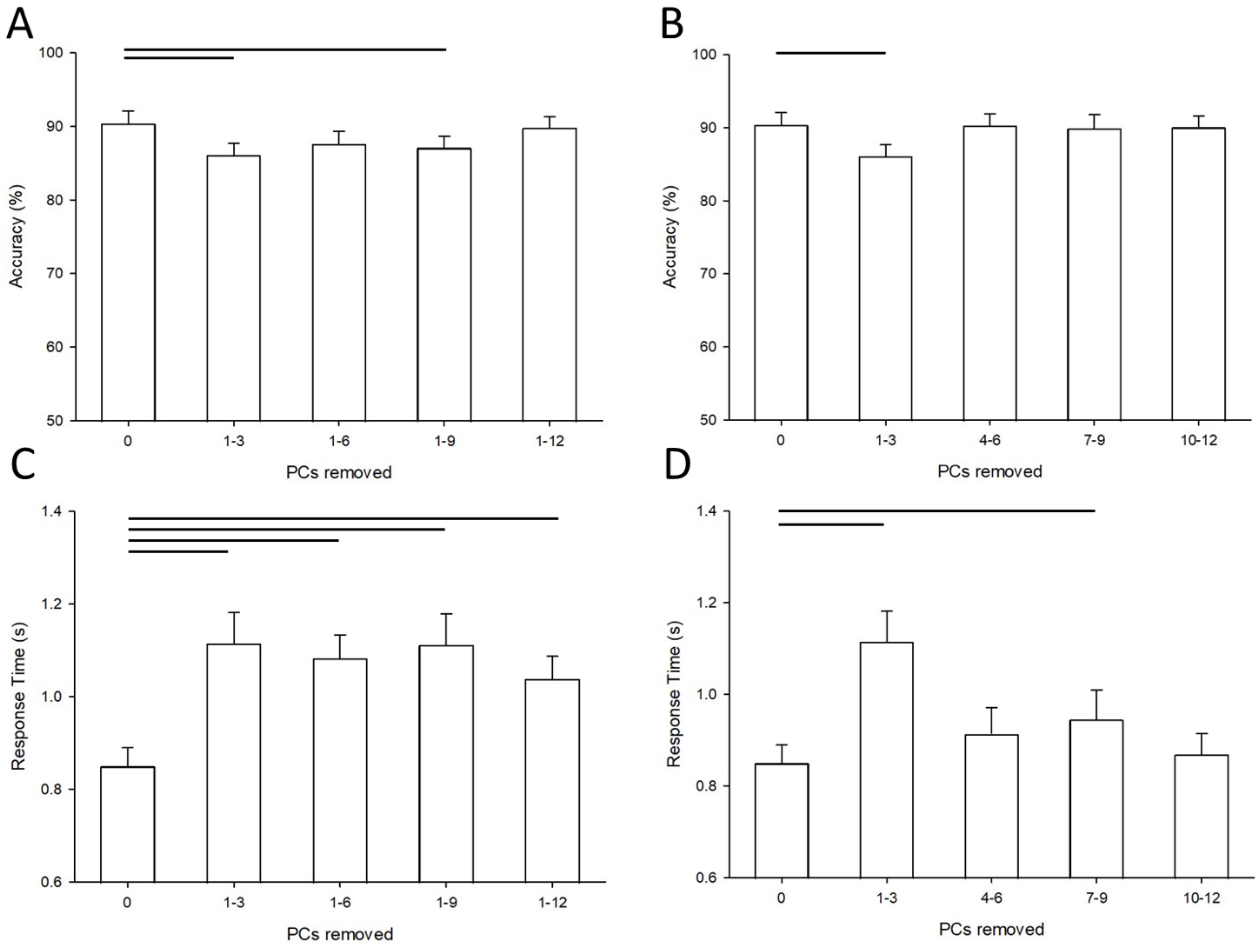

Figure 6.4. *Experiment 1: The effect of removing principal components of the perception of gaze. (A) Performance accuracy when removing PCs cumulatively. (B) Response times to make a correct gaze decision from images that have had PCs removed cumulatively. (C) Performance accuracy when removing bands of PCs. (D) Response times to make a correct gaze decision from images that have had PCs removed in bands. Horizontal lines indicate significant differences (p < .05) relative to the 0 PCs condition (original reconstruction). Error bars indicate standard error of the mean.*

When removing PCs cumulatively, a repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(3.39, 243.76) = 2.82$, $p = .033$, $\eta^2_p = .04$) and response time ($F(3.16, 227.45) = 7.91$, $p < .001$, $\eta^2_p = .10$). Planned comparisons showed that there was a decrease in the perception of gaze when PCs were removed cumulatively for all conditions (when compared to the control condition). However, this was only significant for the removal of PCs 1-3 ($t(72) = 2.61$, $p = .011$) and PCs 1-9 ($t(72) = 2.14$, $p = .036$), with the removal of PCs 1-6 ($t(72) = 1.86$, $p = .117$) and 1-12 ($t(72) = 0.44$, $p = .663$) not having a significant decrease in the perception of gaze. However, the time taken to make a correct gaze judgement was significantly increased for all conditions compared to when no PCs were removed (1:3 (t(72) = -4.01, $p < .001$); 1:6 (t(72) = -4.92, $p < .001$); 1-9 (t(72) = -4.56, $p < .001$); 1-12(t(72) = -4.47, $p < .001$)). These data show that the removal of the initial PCs decreases accuracy and increases response time.

The contribution of PCs to the perception of gaze becomes clearer when bands of PCs were selectively removed. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy (F(2.99, 214.97) = 3.39, $p = .019$, $\eta^2_p = .05$) and response time (F(2.78, 202.02) = 6.78, $p < .001$, $\eta^2_p = .09$). Planned comparisons revealed that this difference in accuracy was due to the removal of the first three PCs only (t(72) = 2.61, $p = .011$). None of the other conditions showed a significant effect (4-6 (t(72) = 0.11, $p = .915$); 7-9 (t(72) = 0.34, $p = .732$); 10-12(t(72) = 0.30, $p = .763$)). The response time to make a correct judgment was significantly increased for the 1-3 (t(72) = -4.01, $p < .001$) and 7-9 (t(72) = -2.12, $p = .037$) conditions. However, there was no significant difference for 4-6 (t(72) = -1.44, $p = .154$) or 10-12 (t(72) = -0.52, $p = .608$) conditions. Taken together, these data show that the first three principal components are most important when making judgments of gaze.

6.4.2 Experiment 2 – Gender

In Experiment 2, we measured the perception of gender from faces in which different combinations of PCs were removed from the image. Participants' perception of gender was measured by indicating whether the face was male or female using a 2AFC paradigm. There were eight image conditions in which we removed principal components cumulatively or selectively. Figure 6.5 shows the accuracy and response time for each condition (with chance level being 50%).

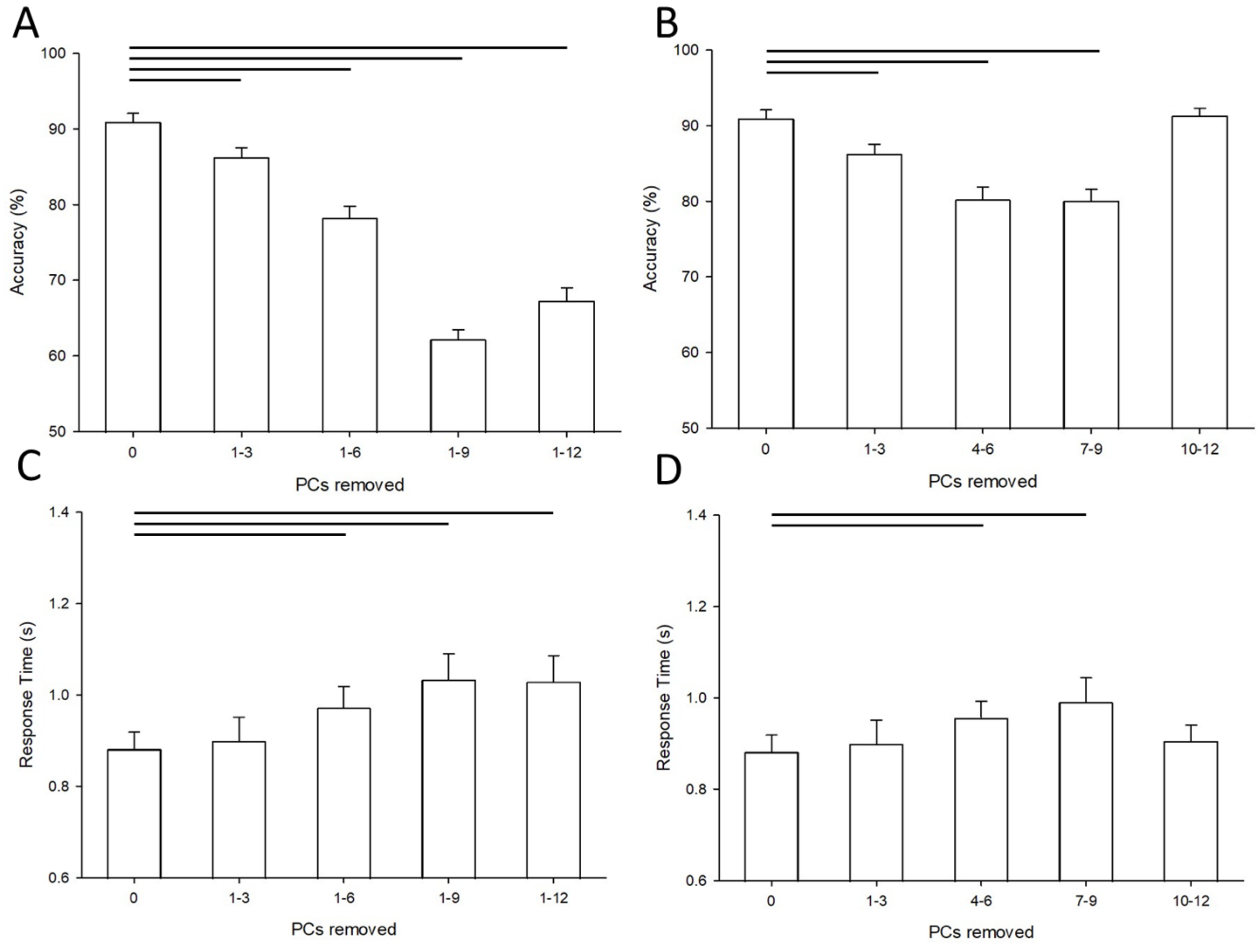

Figure 6.5. *Experiment 2: The effect of removing principal components of the perception of gender. (A) Performance accuracy when removing PCs cumulatively. (B) Response times to make a correct gender decision from images that have had PCs removed cumulatively. (C) Performance accuracy when removing bands of PCs. (D) Response times to make a correct gender decision from images that have had PCs removed in bands. Horizontal lines indicate significant differences (p < .05) relative to the 0 PCs condition (original reconstruction). Error bars indicate standard error of the mean.*

When removing PCs cumulatively, a repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(3.56, 253.33) = 95.45$, $p < .001$, $\eta^2_p = .57$) and response time ($F(3.37, 239.24) = 5.29$, $p < .001$, $\eta^2_p = .07$). Planned comparisons showed that this is explained by a significant decrease in accuracy for all conditions, compared to when no PCs were removed (1:3 ($t(72) = 3.37$, $p < .001$); 1:6 ($t(72) = 2.30$, $p < .001$); 1:9 ($t(72) = 16.74$, $p < .001$); 1:12 ($t(72) = 12.30$, $p < .001$)). A similar pattern of data was found for the measure of response time, whereby the time taken to make a correct gender judgment increased significantly for all PC conditions except 1-3 (1:3 ($t(72) = -0.41$, $p = .680$); 1:6 ($t(72) = -2.16$, $p = .034$); 1:9 ($t(72) = -3.62$, $p = .002$); 1:12 ($t(72) = -3.52$, $p = .002$)). Taken together, these data show that as more

PCs are removed from face images, the perception of gender becomes harder and once the first nine PCs are removed performance drops to just above chance level. This is mirrored by the response time data, suggesting that as more PCs are removed not only does performance decline, the response time to make a correct gender decision also increases.

The contribution of PCs to the perception of gender becomes clearer when bands of PCs were removed selectively. A repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(3.07, 218.07) = 20.12$, $p < .001$, $\eta^2_p = .22$) and response time ($F(3.36, 238.50) = 2.17$, $p < .036$, $\eta^2_p = .04$). Planned comparisons revealed that this difference in accuracy was due to lower accuracy for the removal of PCs 1-3 ($t(72) = 3.37$, $p < .001$), 4-6 ($t(72) = 5.68$, $p < .001$) and 7-9 ($t(72) = 7.59$, $p < .001$). However, there was no significant effect of removing PCs 10-12. The time taken to make a correct gender judgment was significant for the removal of PCs 4-6 ($t(72) = -2.44$, $p = .017$) and 7-9 ($t(72) = -2.49$, $p = .014$). There was no significant difference in response time when removing PCs 1-3 ($t(72) = -0.41$, $p = .680$) or 10-12 ($t(72) = -0.69$, $p = .493$). These data show that the intermediate bands of image dimensions (PCs 4-9) are most important when making judgments of gender, both for accuracy and for the time taken to make a correct judgment.

6.4.3 Experiment 3 – Expression

In Experiment 3, we measured the perception of expression from faces in which different numbers of PCs were removed from happy and sad expression faces. Participants' perception of expression was measured by indicating whether a face had a happy or sad expression using a 2AFC paradigm. Figure 6.6 shows the average accuracy and response time for each image condition (with chance level being 50%).

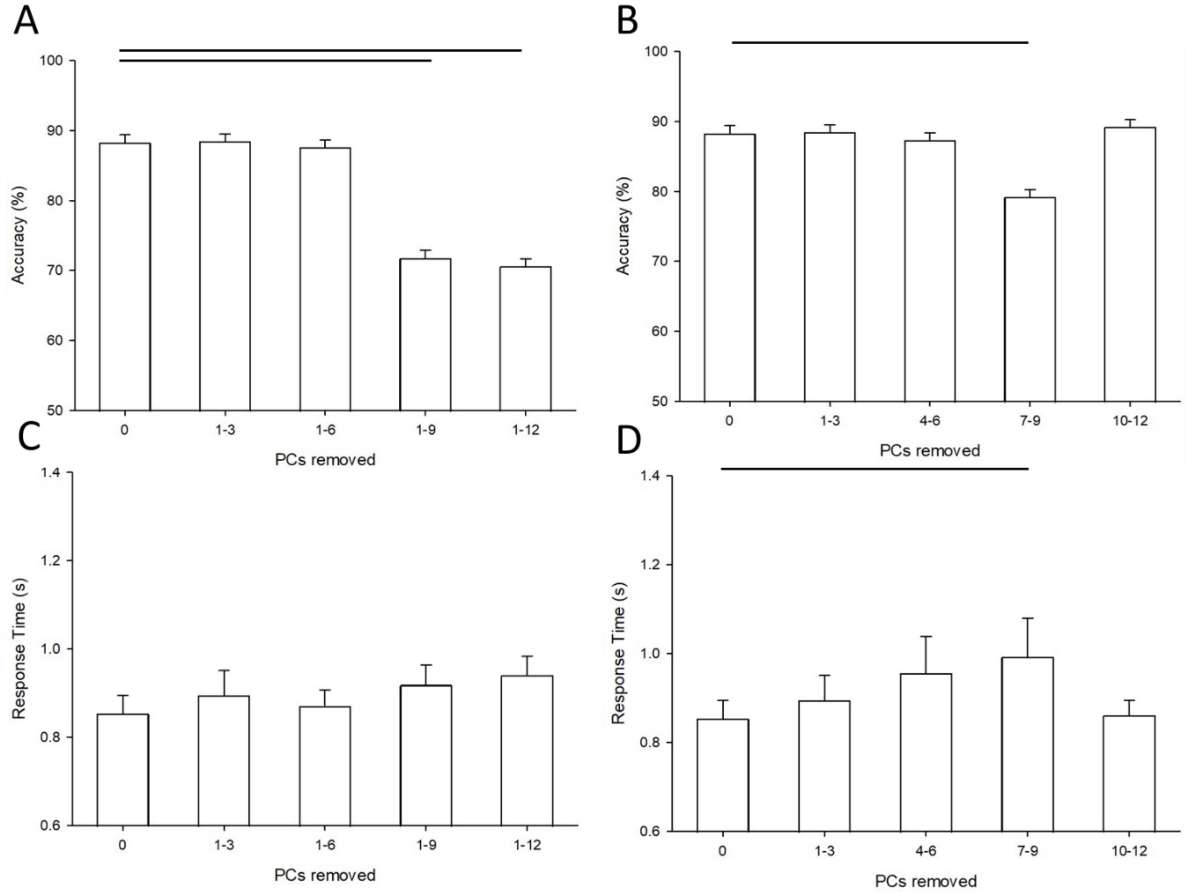

Figure 6.6. *Experiment 3: The effect of removing principal components of the perception of emotional expression. (A) Performance accuracy when removing PCs cumulatively. (B) Response times to make a correct expression decision from images that have had PCs removed cumulatively. (C) Performance accuracy when removing bands of PCs. (D) Response times to make a correct expression decision from images that have had PCs removed in bands. Horizontal lines indicate significant differences (p < .05) relative to the 0 PCs condition (original reconstruction). Error bars indicate standard error of the mean.*

When removing PCs cumulatively, a repeated-measures ANOVA revealed a significant effect of condition for both accuracy ($F(4, 288) = 110$, $p < .001$, $\eta^2_p = .61$) and response time ($F(3.13, 225.07) = 3.51$, $p = .035$, $\eta^2_p = .05$). Planned comparisons showed that the removal of the first 3 ($t(72) = -0.17$, $p = .605$) or 6 ($t(72) = 0.60$, $p = .549$) PCs did not significantly decrease accuracy compared to when no PCs were removed. However, there was a significant decrease in accuracy when 9 ($t(72) = 13.86$, $p < .001$) or 12 ($t(72) = 15.68$, $p < .001$) PCs were removed. But, there were no significant differences in response time relative compared to when no PCs were removed (1:3 ($t(72) = -0.77$, $p = .446$); 1:6 ($t(72) = -0.59$, $p = .556$); 1:9 ($t(72) = -1.57$, $p =$

.121); 1:12 (t(72) = -1.80, p = .076)). Taken together these results suggest that PCs 1-6 do not contribute to the perception of happy and sad expressions. However, later PCs are important.

For the selective removal of PCs, a repeated-measures ANOVA revealed a significant effect of condition for both accuracy (F(3.55, 255.37) = 7.45, p = .008, η^2_p = .09) and response time (F(2.27, 158.99) = 1.90, p = .050, η^2_p = .04). Planned comparisons revealed that similar to the cumulative findings, removing PCs 1-3 (t(72) = -0.17, p = .605), 4-6 (t(72) = 0.74, p = .459) or 10-12 (t(72) = -0.76, p = .452) did not affect accuracy. However, there was an effect of removing PCs 7-9 (t(72) = 7.25, p < .001). The response time data mirrored this pattern of findings, with the removal of PCs 7-9 resulting in a significant increases in response time (7-9 (t(72) = -2.20, p = .031). There were significant differences for all other conditions (1-3 (t(72) = -0.77, p = .446); 4-6 (t(72) = -1.92, p = .059); 10-12 (t(72) = -0.26, p = .799)). Therefore, these findings suggest that a band of intermediate image dimensions (PCs 7-9) is important for the perception of emotional expressions, and that early and later image dimensions to not contribute when making these decisions.

6.5 Discussion

The aim of this study was to determine what information is necessary for the perception of gaze, gender and expression. Specifically, we asked whether the same or different image dimensions are used to process these different aspects of face perception. To address these questions, a principal components analysis was used to reveal the underlying image dimensions of naturally varying face images from different familiar and unfamiliar identities. We then removed PCs or image dimensions from the faces and measured their effect on behaviour. Our key finding is that the perception of gaze, gender and expression is dependent on distinct, but overlapping, image dimensions.

In the first experiment, we explored how different combinations of PCs relate to the perception of gaze. We found that removing the early PCs (1-3) significantly decreased the perception of facial gaze, both in accuracy and the time taken to make a correct judgment. Whereas, gaze judgments remained relatively stable when later PCs (4-12) were removed. Previous research has shown that early image dimensions for both shape and texture contribute to visualisations of gaze perception (Burton et al., 2016). The first three dimensions of shape, typically describe a series of rigid head rotations in three-dimensional space, accounting for the largest proportion of shape variance within ambient images. This is congruent with the fact that head direction provides information about the direction of gaze. The position of the eyes, which is also considered a property of shape information, can also influence the perceived direction of the face (Langton, 2000; Burton et al., 2016). Previous studies have suggested that intermediate PCs for shape reflect the position of the eyes (Burton et al., 2016). However, in the current study, we found that removing more intermediate bands of components (for both and shape and texture) did not impact participants judgements of gaze. Whilst shape properties have been suggested to underpin viewpoint, the perception of gaze can also be understood from texture information. Indeed, when the initial texture PCs are manipulated by increasing or decreasing their variance (+/- 2 SDs) it is observed that the visualisation shows an apparent movement of directional lighting from one side of the image to another, indicative of coding for a left-right rotation (Burton et al., 2016). Our results are consistent with these observations, showing that the perception of gaze becomes harder if these components are removed. Future research should aim to explore what information is contributing to the disruption of gaze perception, for example, is this effect due to changes in head direction (viewpoint), the position of the eyes (e.g. pupil), the changes in illumination patterns or a combination of components.

We next explored the contribution of different combinations of image dimensions to the perception of gender. We first found that the perception of gender decreased as a function of removing more PCs in a cumulative fashion and is just above chance-level when the first nine components are removed. This suggests that to categorise the gender of a face, multiple image dimensions are used to make this judgment. When image dimensions were removed in bands, this effect seems to be driven by the intermediate image dimensions (4-9), with a return to baseline performance when the final band of dimensions are removed. The effect of removing PCs on the perception of gender was thus different to the perception of gaze; whilst both were significantly decreased when the initial components were removed, gender perception relies more heavily on the intermediate dimensions. This is consistent with previous research by Burton et al (2016), that reported notable principal component differences that emerge for male and female faces. For example, in female faces, intermediate texture components corresponded with the skin becoming: more orange, lips a more red hue and a darkening of the eyelids; whereas, men exhibited a component reflecting facial hair, with darkening/lightening of the upper lip and beard area. However, the contribution of these image components to the perception and categorisation of gender, again, remained unresolved. Our research extends these findings by showing that these intermediate dimensions play a causal role in the perception of gender and when these image properties are absent in a face, categorising gender is at chance level. One possible explanation for these findings is that these critical components for making gender discriminations might code for the shape and surface properties that have been found to convey the gender of a face. For instance, PCs 4-9 either as individual components or in tandem, might reflect the global shape properties of face length, jawline angularity (Brown & Perrett, 1993) and the local shape and texture properties of facial hair, eyebrow distance/shape and their corresponding shading patterns (Baudouin, 2006). Whilst a simple one-to-one mapping of PCs to image properties was not possible for the current experimental design, future research should endeavour to explore how single PCs and combinations of PCs interplay with the known diagnostic facial information important for gender discrimination. Moreover, it is also important to explore the potential idiosyncrasy of PCs within and between identities.

Finally, we investigated the impact of removing different combinations of image properties on the perception of facial expressions of emotion. Here we found that the first six components did not contribute to the recognition of expression, but once the final image dimensions were removed, performance decreased. When removing the PCs in bands, we found that this effect was driven by a narrow band of dimensions (7-9) and performance remained at control level when a later band of PCs was removed. These results suggest that there is a narrow band of intermediate image dimensions that appear to be important for making classification judgments of expression. These intermediate dimensions were also important in the perception of gender (but not gaze), suggesting that the same image properties can be utilised for the perception of multiple facial signals. These findings differ from a previous study that showed there was a range of components that were useful for computerised expression categorisation, with early to mid PCs for shape and mid to late PCs for texture observed to be the most useful (Calder et al., 2001). In the current study, we do find evidence that early PCs are important in the perception of expression, but revealed that a much narrower band of dimensions underpinned performance. It is important to note that the PCA conducted by Calder et al (2001) was conducted using the emotional faces from the Ekman and Friesen image set (110 images). This standardised image set differs from the large ambient image set used in the current study. Thus, direct comparisons between the findings of this study and Calder et al (2001) are difficult. Nonetheless, it seems that there is a narrow range of dimensions important for the perception of expression in both humans and models of expression categorisation. Future research should continue to explore the contributions of image dimensions to the perception of expression, expanding this to the full range of emotional expressions, to better understand whether the same dimensions have a shared or independent contribution to multiple emotional expressions.

A key question we asked in this study was, are the same image properties critical for the perception of different facial signals (namely- gaze, gender and expression). The image dimensions in the present experiments were generated from a PCA of a very large set of images compiled from multiple photos of different identities that were ambient in nature. Combinations of these image dimensions were then removed from images that varied in gaze, gender or expression. Here we found that there was a degree of overlap in the image dimensions critical for the perception of each of these facial signals. For example, in the case of gaze perception, only the removal of initial PCs (1-3) decreased performance. For gender judgements, early PCs affected judgements, but the intermediate and later PCs (4-9) appeared to be more critical. Judgements of expression recognition appeared to be most dependent on later PCs (7-9). The overlap of PCs is not particularly surprising, as the same facial properties can be diagnostic for different aspects of perception. For example, the appearance of the eyebrows can be diagnostic for judgements of both gender and expression. However, for each of these facial signals, there were distinct image dimensions. Thus, it seems that faces can be categorised along different perceptual dimensions using a subset of image properties that can be used to perceive gaze, gender, expression and recognition (as we saw in Chapter 5).

Our study opens the interesting possibility that the information space derived from statistical techniques, like PCA, contains discrete bands capturing information about different aspects of the face. Focussing on the statistical information space, opens up a potentially useful route for future research in face perception more generally. It is interesting to note that this possibility arises from a comparatively simple linear analysis such as PCA. Of course, many more complex, non-linear, decompositions of statistical face-space are possible, but nevertheless, the approach described here offers a (perhaps surprising) degree of interpretability. Future research could explore what image properties are important for the perception of race, age and trait judgments (e.g. trustworthiness and dominance).

In conclusion, our results suggest that there are a small set of image dimensions that provide unique and overlapping contributions to the perception of gaze, gender and expression. Our findings provide novel insights into the image dimensions that are important for the perception of different facial signals. This provides a new perspective for understanding how image properties underpin different aspects of face perception.

Chapter 7- General Thesis Discussion

The aim of this thesis was to explore the role of visual information in the perception and recognition of faces. The experiments use a combination of behavioural, computational and neuroimaging approaches to explore how faces are represented in the brain. This combination of methods utilised throughout the experimental chapters were employed to explore two key questions. Firstly, I asked what visual information is needed to generate a view-invariant representation. Here, I investigated the hierarchy of facial representations prior to achieving view-invariance (Chapter 3), as well as what visual properties of the face are important for generating and accessing these representations (Chapters 4 and 5). Next, I asked how visual information is used for different aspects of face perception. To address this question, I explored whether the same or different combinations of image dimensions underlie the perception of identity, gaze, gender and emotional expression (Chapters 5 and 6). Together these questions allow us to extend previous theories and research concerning how the image properties of faces contribute to the perception and neural representations of faces.

7.1 What visual information is needed to generate a view-invariant representation?

Before considering what visual information is critical to forming view-invariant representations, it is first important to understand the theoretical perspectives on how view-invariant representations are generated from view-specific inputs. In classical models of face processing, it is suggested that view-specific inputs converge onto a view-invariant representation (Bruce & Young 1986, 2012; Rolls, 2012). However, a more recent hypothesis posits a two-step process for achieving view-invariance. Here it is suggested that there are two stages of convergence, whereby view-specific representations first converge into view-symmetrical representations which then further convolve into a view-invariant representation of identity (Freiwald & Tsao, 2010). Evidence for an intermediate view-symmetric representation of faces is evident across a range of fields and studies (Guntupalli et al., 2017; Flack et al., 2019). Behavioural evidence is derived from studies indicating that faces with symmetrical viewpoints, such as two opposite profiles, are perceived as more similar than those with non-symmetrical viewpoints (e.g. left profile and left ¾ view). Additionally, recognition judgments following face learning, demonstrate greater accuracy when the test viewpoint aligns symmetrically with the learned viewpoint (Favelle & Palmisano, 2018; Flack et al., 2019). Neurophysiological investigations further support a functional hierarchy of facial viewpoint processing, where early processing stages involve view-specific tuned neurons, intermediate face regions exhibit more view-symmetric responses, and later face regions display increased view-invariance (Perrett et al., 1991; Freiwald & Tsao, 2010). Finally, neuroimaging studies in both humans and primates have supported this representational hierarchy, revealing a transition from view-specific to view-symmetric representations (Axelrod & Yovel, 2012; Kietzmann et al., 2012; Guntupalli et al., 2017; Flack et al., 2019).

A limitation of previous studies investigating view symmetry, is that they have predominantly used unfamiliar faces. A range of evidence shows that there are distinct differences between the perception of familiar and unfamiliar faces (Bruce, 1982; Hancock et al., 2000; Longmore et al., 2008). For example, when judging whether two faces belong to the same identity, participants are more accurate for familiar faces, with unfamiliar face matching using two unfamiliar identities showing the poorest results (Bruce et al., 1999; Davies-Thompson, Gouws & Andrews, 2009). This familiar face advantage is also exacerbated when viewpoint, orientation and expression are manipulated (Bruce, Valentine & Baddeley, 1987; Bruce et al., 2001). Models of face processing suggest that these differences arise due to familiar faces having a mental representation that is view-invariant, whereas unfamiliar faces rely on a more limited pictorial code (Bruce & Young, 2012). Therefore, due to these fundamental distinctions between familiar and unfamiliar face processing, it was unknown whether familiar faces would show evidence of a view-symmetric representation as they have already acquired view invariance.

Another limitation of previous studies is the focus on rotations that result from common or canonical rotations of the head (yaw). So, it was not yet clear if view symmetry is specific to this natural rotation of the head or is also evident for less common or non-canonical rotations (roll) of the head that are less frequently experienced in everyday life. If the same pattern of facial representations is seen for both canonical and non-canonical rotations, it could imply that view-symmetric representations are a product of face-selective areas showing general responses to mirror symmetry as opposed to an intermediate facial representation, underpinning identity recognition.

In Chapter 3, I explored these questions finding that view-symmetric neural responses are evident for both unfamiliar and familiar faces in core face regions. This supports the notion of a two-stage process of view-invariance with view-symmetric representations being generated for both familiar and unfamiliar faces. This suggests that the view-invariant representations that are characteristic of familiar faces emerge at later stages of processing (Davies-Thompson et al., 2013; Weibert et al., 2016). I also found that there was a difference in the way that faces generated from different rotations of the head were represented in the brain. For non-canonical rotations (roll) of the face, there was limited evidence for the pattern of neural response in early visual areas being systematically predicted by changes in viewpoint, as was found with canonical rotations (yaw). However, we did find view-symmetric neural patterns of response for non-canonical rotations in face regions, which is consistent with the behavioural finding that symmetrical non-canonical viewpoints were perceived to be more similar than asymmetrical viewpoints. This suggests that the emergence of view-symmetric responses occurs differently for canonical and noncanonical rotations of the face. Thus, view-symmetric representations for yaw are not simply a product of face-selective areas processing general mirror symmetry, but form a functional hierarchy that emerges from view-specific inputs being processed in early-visual regions, prior to generating full view-invariance.

Similarly, we found a strikingly parallel pattern of findings for how viewpoint was represented in a DCNN, with human perceptual similarity ratings being able to accurately predict the outputs of the fully connected layers. Within the early convolutional layers, there was evidence of a view-specific representation, but view-symmetry emerged in the fully connected layers. In recent studies it has been shown that in the fully connected layers, there are clear representations of identity, and other facial signals, however within these representations a high degree of image information is also kept such as viewpoint/head direction (Parde et al., 2017). Thus, this suggests that a view-symmetry representation is utilised in a meaningful way within the fully connected layers to aid to some degree identity recognition/classification. Taken together the findings presented in Chapter 3, are consistent with the notion that the process of generating a view-invariant representation contains two discrete and connected stages.

Upon establishing the stages of facial representation prior to view-invariance, a full account of face recognition needs to consider what visual information is critical for the generation of a view-invariant representation that underlies identity recognition. The IAC model posits that a subset of image dimensions describes the appearance of a face that underlies the structural representation of identity, using a perceptual front-end model based on principal components analysis (PCA). After testing various models of this PCA front end, Burton and colleagues (1999) concluded that the shape-free model (containing only texture information), best captured the underlying structural representation that could drive the process of face recognition, as the additional shape information was seen as less reliable and contributed little to improving recognition when compared to a shape-free model. Taken together, these models imply that not all variability is equally important for generating a view-invariant representation for example, image variation uncommon to multiple images and variation in shape properties. However, often the behavioural tasks employed in studies of face recognition allows participants to use cognitive strategies instead of stored mental representations of faces (thus not engaging view-invariant representations) such as perceptual matching and recognition memory. Therefore, the importance of shape information in making identity judgements might have been overlooked.

In Chapter 4, I addressed these issues by employing tasks that relied on participants' stored mental representations of familiar faces as well as exploring the sensitivity of face-selective areas to both shape and texture properties. Here we revealed that whilst texture was the dominant visual property utilised when making identity judgments, shape information can also be used for recognition and there are occasions when shape information can override texture information when making identity judgments. Moreover, we found consistent findings with previous research that show that face selective regions show an equal sensitivity to both shape and texture (Andrews et al., 2016; Jiang et al., 2006). However, in contrast to previous research we showed there was an equal sensitivity to both properties both within and between identities and irrespective of the familiarity to the identity, solidifying the importance of both properties. Taken together, these results suggest that both shape and texture information contribute to the generation of a view-invariant representation.

We next wanted to build upon these findings by exploring what specific image properties are fundamental to a view-invariant representation that underpins the recognition of identity.

Based on the findings of Chapter 4, we explored the contributions of both shape and texture image dimensions generated from a PCA conducted on a large ambient image set containing over 6100 images. Over the series of experiments presented in Chapter 5, it was found that there are a small set of intermediate image dimensions that contributed to the perception of identity. Moreover, it was revealed that coarse scale image variation such as pose and ambient illumination contained within the earliest PCs did not contribute to familiar face recognition. Indeed, removing this information resulted in familiar faces being more recognisable. The same pattern of findings was also found for learning new faces. When the early image dimensions were removed from learning images, subsequent recognition increased, and when the intermediate dimensions were removed subsequent recognition decreased. Taken together, these findings extend previous research by suggesting that, view-invariant representations might rely on just a small subset of image dimensions. Thus, the process of generating a view-invariant representation can be suggested to rely upon extracting the degree of variability an identity can show within these components only. For example, as the early image dimensions were shown not to have contributed to familiar face recognition or face learning, understanding how a face can vary in coarse scale image variation is not needed for generating a view-invariant representation. This implies that not all variability is important for generating a view-invariant representation.

Generating view-invariant representations when learning new identities is suggested to be driven by increased exposure to an identity. Specifically, variability in exposure to an identity has been shown to be important for learning new identities (Ritchie & Burton, 2017). This is consistent with findings that suggest that images of faces taken from one identity can be as variable as images taken from different identities (Jenkins, White, Van Montfort & Burton, 2011). Thus, an important computational step in generating view-invariant representations relies on learning the degree to which an identity can naturally vary. The averaging hypothesis proposes that a canonical image whereby image specific information is filtered out and information present in multiple images is kept, is utilised as a facial representation, in essence an FRU. Therefore, implying that not all variability in face images of a single identity is equally important, with image specific variation (e.g. environmental lighting) being deemed as noise.

The notion that not all variability is equally important for learning new identities is not a new concept. Burton (2013) emphasizes the importance of considering how different forms of variability can affect the recognition process. In the work by Ritchie and Burton (2017), they made a distinction between two types of variability: systematic and unsystematic variability. Systematic variability involves alterations in factors like camera angles and the pose of the target within a consistent environment. In contrast, unsystematic variability encompasses changes not only in target appearance (such as hairstyle, makeup, and clothing) and the surroundings, but also in camera angles and target pose. Traditionally, research has primarily concentrated on systematic variability. In such studies, researchers make efforts to keep most characteristics of facial stimuli constant while selectively manipulating only the variable under investigation in regards to face learning and face recognition (Longmore et al., 2008; Liu et al., 2009). These studies have found little advantage in recognition in systematically varying pose or lighting information. Contrastingly, it has been argued that it is unsystematic variability that plays a crucial role in broadening the circumstances under which a person can be recognised. When observers are exposed to multiple images of an individual that capture the full spectrum of natural variations in changeable facial attributes like expression, facial hair, and age, as well as non-facial features such as hairstyle and hair colour, they exhibit quicker responses in verifying the name associated with a recently-learned face and enhanced accuracy in determining whether two photos depict the same person or not (Ritchie & Burton, 2017).

The experiments reported in this thesis complement previous findings, showing that removing systematic variation contained in the earliest image dimensions of learning images, can improve subsequent recognition. Furthermore, our findings extend previous research by revealing the image dimensions that account for the critical variance in face recognition and face learning. Thus, it is plausible to suggest that this band of intermediate image dimensions reflects the unsystematic variation that viewers must extract the range of when learning new identities. Nevertheless, questions still remain regarding the idiosyncrasy of variability between identities. For example, to what extent do identities show similar ranges of variability within these key image dimensions? Here we used just one PCA to generate a set of image dimensions to explain the variance within a large ambient image set. Provided that the PCA can have access to a sufficient number and variation of images during its input, it would be possible to conduct individual PCAs on images of just one identity. This would allow one to compare how different image components contribute to the recognition of individual identities. However, the challenge here would be that a sufficiently large sample of images would be needed in order to compare the roles of bands of components, for example the same band of components generated from an individual PCA of one identity might not reflect the same image dimensions generated from another.

Within this thesis a number of findings have illuminated what visual information is needed to generate a view-invariant representation. Firstly, we have solidified the process in which view-invariant representations are generated. This has shown that a two-stage progression is needed to achieve view-invariance, with view-symmetry representations being an intermediate step. Secondly, we showed that at the basic level of visual information, whilst texture information is the dominant property utilised for familiar face recognition (hence important for view-invariant representations), shape properties also provide discrete and important contributions to familiar face recognition with face-selective areas showing equal sensitivity to both properties. Finally, we explored the image dimensions that are critical for generating and utilised in view-invariant representations that underpin identity learning and judgments. Here, we revealed that not all variability is equally important. Systematic variation was shown to hinder familiar face recognition and face learning, implying that view-invariant representations do not rely on information such as pose, illumination and so on. However, we did reveal that a small band of intermediate image dimensions were critical for familiar face recognition and when generating view-invariant representations during face learning.

7.2 Is the same visual information used for processing multiple facial signals?

Multidimensional face space models aim to provide a framework to explain how faces are represented in memory (Valentine, 1991). Here, each face is represented by a single location within the multidimensional space, where each dimension maps onto either a specific parameter or global property of a face that varies from one face to another. Image based properties (such as the distance between the eyes) or more abstract properties of a face (such as trustworthiness), have all been considered possible dimensions of face-space (Valentine, Lewis & Hills, 2016). The number or nature of the image dimensions within a face space framework remains unclear, however with techniques such as PCA it has been suggested that face images can be reconstructed and subsequently recognised with a high degree of accuracy using less than 50 image components (Burton et al., 2016). Whilst there have many variations of the face-space model each being able to account for different behavioural findings, a commonality they share is that they suggest that there is one set of common axes (or image dimensions) along which faces are coded. This suggests that different identities' faces are represented using a unified set of dimensions.

Burton, Kramer, Ritchie and Jenkins (2016) proposed an alternative multi-dimensional account of how faces are represented, in which each face is represented by its own person-specific coding space. This considers the fact that any face can generate a range of images. In their study they conducted separate PCA analyses on individual identities (from 30 images of each identity), making several observations. They found that certain facial elements were coded by different principal components between identities, implying that separate face spaces would better characterise identity idiosyncrasies. For example, a smile on two individuals might transform the appearance of a face in different ways, reflecting the degree of idiosyncrasy between individuals. Similarly, they use the example of two individuals who vary in nose length, which when combined with a 10-degree head turn, has idiosyncratic effects for both 2D shape and texture whereby this 3D movement in the world translates the tip of the longer nose much further than the tip of the short nose. Thus, concluding that individual faces have their own idiosyncratic variability, describing that all faces vary in appearance, but that they vary in different ways, represented by different axes in face space. Therefore, in Chapter 5, it was of interest to explore whether a single set of image dimensions derived from a large sample of images from different identities with multiple images of individual identities, could underpin the recognition of multiple identities.

The findings of Chapter 5 showed that the same image dimensions when removed from familiar face images impacted the recognition of multiple identities, suggesting that to a certain degree there are a shared set of image properties that contribute to the recognition of identity. Moreover, the same image dimensions were also seen to contribute to the learning of multiple identities. Therefore, this seems to imply that one set of image dimensions can underpin variability both within and between identities, when a sufficient image set is used to generate the image dimensions. Thus, this would suggest that a common set of axes can represent multiple facial identities. It is important to note that we did not explore the generation of image dimensions within specific identities to then compare the difference in face learning between this and image dimensions generated from a larger image set.

Parallel to the representation of identity, in Chapter 6 we used the same set of image dimensions generated using a PCA from the same ambient image set as we did for Chapter 5. This allowed us to explore whether the same image dimensions contribute to the perception of identity as well as other facial signals- gaze, gender and expression. The overarching finding here was that different but overlapping bands of image dimensions underpinned the perception of different face signals.

Early image dimensions (PCs 1-3) were the only image dimensions shown to underpin the perception of gaze. This fits the notion that these dimensions represent aspects of 3D movement for example, head direction, the position of the eye, and the changes in illumination patterns. The same image dimensions were also shown to be important for identity recognition and face learning. However, instead of underpinning recognition, these components showed a negative loading, such that removing them from familiar face images (and learning images) improved recognition. This shows that the same image components can be useful for one aspect of face perception (gaze) and detrimental for another aspect of face perception (identity).

An intermediate band of image dimensions (4-6) was shown to be important for gender and identity recognition. The Bruce and Young (1986, 2012) model propose that when face classification processes which do not involve any information regarding the identity of faces are instead made through a parallel pathway that is not mediated by the FRUs and depend on aspects that are different from the ones that are used for identification (Ganel & Goshen-Gottstein, 2002). Here, we show that the same visual information can underpin both identity and aspects of face perception and is consistent with findings that show that participants cannot selectively attend to either gender or identity without being influenced by the other, even when instructed to ignore the irrelevant dimension (Ganel & Goshen- Gottstein, 2002).

A later band of intermediate image components (7-9) were found to contribute to the perception of identity, gender and emotional expression. This is interesting as classical models of face processing, often differentiate between changeable and unchangeable aspects of the face (namely identity and expression), however, here we show that the same image properties are critical for both of these judgements. It is of course possible that although the same image dimensions are important for these properties, the initial encoding of these properties is then recruited in parallel by different systems to process the facial signals.

Overall, the findings of Chapters 5 and 6, elucidate novel findings within this debate of facial signals and representations. Firstly, they show that a rather small subset of image dimensions is critical for multiple aspects of face processing. Secondly, they reveal that individual facial signals rely on an even smaller subset of image properties, in other words there are plenty of image properties that have no impact on the perception of identity or expression and so on. Next, they show that the same image properties can underlie multiple aspects of face processing, even those that have been suggested to run in parallel. Thus, one possible mechanism to explain these findings would suggest that faces are represented and initially encoded using just one set of axes (image properties) irrespective of familiarity to the face, then different systems recruit the loading of the dimensions that they are tuned to, to support the processing of specific facial signals. Future research should endeavour to further explore the potential idiosyncrasy of PCs within and between identities and how image properties contribute to the representation of multiple facial signals. For example, in our studies a simple one-to-one mapping of image dimensions to facial characteristics was not possible with the current experimental design. By future research exploring how individual PCs or various combinations of PCs contribute to the many signals available in faces a more in-depth construction can be made in regards to how visual information is represented and subsequently accessed during processing.

7.3 Methodological Considerations

Within this thesis a combination of methods and paradigms were used to investigate face recognition. Previous research has utilised several different behavioural measures for face recognition, each providing a metric for the accuracy of how well one can recognise a familiar face. Examples of these measures include famous face tests where participants' metric of recognition is whether a face is famous or not (Fast, Fujiwara, & Markowitsch, 2005), face matching tasks in which a key dimension of one of the faces to be matched has been manipulated (Abudaraham, Shkiller & Yovel, 2019) and face memory tasks where unfamiliar faces are presented to participants and after a certain period of time have to identify which faces they had previously seen (Duchaine & Nakayama, 2006). However, as highlighted in the literature, when these methods are employed to investigate familiar face recognition specifically, they often do not engage the stored mental representations of familiar faces in a naturalistic way (Herzmann et al., 2008). For example, famous face tasks do not allow us to distinguish whether participants are getting a correct answer by activating the relevant FRU (invariant representation), or simply whether they are familiar with the person to an extent to know that they are famous from that individual image (but would not be able to recall their name/biographical information). Similarly, a common theme when using face matching tasks to test familiar face recognition, is to manipulate one aspect of the face that is considered potentially important for recognition. The problem here is that these tasks can be completed by using elements of perceptual matching that do not engage the stored mental representations of the identities. Additionally, another drawback of these methods that use familiar faces is that across a participant sample, there will be differences in the exposure and frequency to the identities. Thus, creating a level of construct-irrelevant variance that will be captured at test, individual for each participant. One way of exploring this would be to use a test at the end of the main experiment to assess familiarity to the identities used. Alternatively, to overcome these issues would be to test familiar face recognition using faces that have previously been learned by participants, therefore reducing the variability in prior exposure across the sample. However, often this method has a relatively short period of time between the learning and subsequent test phases, typically spanning just one session. This is problematic when drawing conclusions about how we recognise faces in the natural world. For example, learning faces in this way is unlikely to generate a substantial representation of an identity, even when multiple different images of each identity are used. Having just one learning session prior to test, also means that the mental representations formed are likely to be held in short-term memory which is unreflective of what we know regarding how faces are represented.

These methodological issues were addressed within this thesis. In Chapter 4 we explored the roles of shape and texture in familiar face recognition using two different behavioural tasks. We first used a matching task, in which participants had to match a name to one of eight familiar hybrid faces. By using a name as the probe, this allows for participants to activate their face recognition systems and generate a mental representation of the target identity.

They then have to match this stored mental representation to each of the images to decide if the structural code generated by each image is a match or not. Thus, this task uses a face-matching paradigm but critically cannot be completed using perceptual matching. Likewise, the second paradigm was a simple free-recall design, in which upon seeing a hybrid face image, participants were instructed to enter the name of the identity (or sufficient biographical detail). This task whilst simple in nature, is deceptively harder to complete than a matching task relying only on stored mental representations, moreover, this was made harder by using hybrid faces (depicting the shape of one identity and the texture information of another). Despite this more challenging task, participants had a high degree of accuracy, thus showing that the paradigm is feasible and allows for the exploration of familiar face recognition using celebrity images that cannot be completed using perceptual matching whilst also controlling for variability in identity familiarity (using the post experiment familiarity test).

This free recall paradigm was also used in Chapter 5. Following the main experiments, participants were then given a post-experiment familiarity test, in which multiple novel colour images were used. This was done to affirm which identities participants were already familiar with to a high degree. Upon supplementary analysis we found that an individual item analysis of recognition rates was similar for all identities tested. In other words, our results were not driven by the high or low recognition of any individual identity. A potential extension to this paradigm would be to include a more comprehensive measure of how well each participant knew each identity in the post-experiment familiarity test. This would allow for a more detailed understanding of the variability in the participant sample to each identity. For example, presenting different images of the identities in a randomised sequential order would be able to provide an average rating of familiarity for each identity. Equally, subjective measures similar to confidence ratings could be acquired in which participants self-report their familiarity to each identity using a Likert-scale. The research surrounding how well confidence ratings reflect one's performance is somewhat mixed. However, research has shown that participants are more confident when making correct decisions as opposed to decisions that are incorrect (Devue, Wride & Grimshaw, 2018). Additionally, moderate-high correlations between self-reported face recognition abilities and performance on the CFMT have been revealed (Bowles et al., 2009), which hold up well when controlling for age, gender and cross-culturally (Livingston & Shah 2018; Ventura, Livingston & Shah, 2018). Therefore, the inclusion of a confidence style rating would be beneficial and add another layer of analysis to affirm how well celebrity faces are recognised.

Furthermore, in Chapter 5, we developed a novel paradigm for investigating face learning. We did this to combat the limitations of previous face learning methodological designs that often only employ one learning session that is immediately followed by a test session. It was therefore arguable that at test, participants rely on short-term memory strategies as opposed to stored mental representations of the newly learned identities, acquired after a period of consolidation. Whilst it is the case that one can recognise a face they have briefly seen in everyday life at a later time point, it is hard to generalise these findings more broadly to assess how familiarisation is achieved. In our paradigm we combatted these issues, by employing a longitudinal design. Here participants completed three sessions (2 learning-feedback sessions, 1 test session) with a 48-hour (approximately) delay between each session. Previous research has found that when learning new faces, sleep passively and transiently protects face recognition memory from interference (Sheth, Nguyen & Janvelyan, 2009). Additionally, an increase in connectivity between the memory centres and fusiform gyrus which indicate face memory formation and consolidation have been revealed to occur at later time points than during the encoding session (Geiger, O'Gorman Tuura & Klaver, 2016). Therefore, the inclusion the of these gaps between the sessions allows participants to solidify the learning of these facial identities and suggests that adequate time is needed to generate face representations that underpin identity judgments. Without these delays between learning phases and the final test phase, participants are required to use face representations held in short-term memory that would be more pictorially based, uncharacteristic of an FRU. This issue is mostly relevant to the time in between the final learning phase (or the only learning phase in previous research) and the test phase. As we showed participants multiple different images of each of the newly-learned identities in each session, it was critical for us to not have a learning or recap session on the same day of the test phase. We did this to eliminate any potential priming effects prior to test, so that participants had to rely on their stored mental representations of the newly learned identities. Future research, should consider the use of this more longitudinal design when investigating the process of face learning, and well as exploring what the optimum number of images and learning sessions is needed to achieve the behavioural benefits indicative of possessing an FRU for newly learned identities. Additionally, exploring the more long-term memory decay of the newly learned identities would allow for a richer understanding of how the number of learning sessions and intermediate delays contribute to face learning. For example, what would the recognition rates of the newly learned identities be, if the test phase was a week after the final learning session?

Related to this, we also explored the immediate effects of the learning sessions using our feedback phases. Here we used novel images of the identities and asked participants to select the corresponding name of the identity. This was important so that we could assess how well participants were learning the identities and if any additional learning would be needed. Overall, the findings from these feedback sessions were positive, in terms of showing a clear progression of learning accuracy. This was fundamental as the learning task we used did not involve an active learning element, participants simply learned the identities and corresponding names passively. Thus, being able to have this live feedback, enabled us to observe any potential non-engagement in the learning process, as well as giving participants the chance to practice recognising the newly learned identities. Future research could extend the use of the feedback sessions, and have these immediately prior to the second learning phase. This would allow us to explore the effects of consolidation from the first learning session in more detail.

7.4 Conclusions

The aim of this thesis, was to further explore the role of visual information in the perception and recognition of faces. The experiments in this thesis use a combination of behavioural, computational and neuroimaging approaches to ask how facial representations are formed and what facial information is critical for the recognition and perception of faces. Results in Chapter 3 supported the notion that generating view-invariant representations involves a two-stage process, with a representation of view-symmetry being an intermediate step. These findings were supported behaviourally and within the neuroimaging data, with view-invariance showing different neural origins for canonical vs noncanonical views of the face. Similar patterns also emerged from the DCNN analysis suggesting that this neural network represented view-symmetry in a similar way. Chapter 4, built upon these findings by exploring what basic visual information in the face (shape or texture) is important when making identity judgments. Here, it was shown that whilst texture properties are the dominant cue for familiar face recognition, shape properties provide unique and important contributions when making identity judgments. Moreover, the face selective areas showed an equal sensitivity to both properties. Probing this further, Chapter 5, found that there was a small intermediate band of shape and texture image dimensions that that were important for familiar face recognition and for face learning. However, early and later image dimensions were not important for these processes. Finally, we investigated what image dimensions were important for the perception of gaze, gender and expression. Early dimensions were critical for gaze judgments, with unique intermediate bands of image dimensions underlying gender and expression decisions.

Overall, we have shown that the visual information needed to generate a view-invariant representation relies on a two-stage process, that is driven by both shape and texture properties (with texture being the dominant property). Importantly the information that is critical for making identity judgments within this view-invariant representation relies on a small intermediate band of image dimensions. Similarly, we found that an extended but comparatively small subsection of the first image dimensions also were fundamental to the perception of gaze, gender and expression. Whilst it is the case that different facial signals rely on unique and discrete bands of image dimensions there is some overlap, meaning that the same image property can be useful for deciphering different signals. Thus, taken together, the work presented within this thesis extends our knowledge of the contributions that low-level image properties have on the perception and neural representations of faces.

Bibliography

Abudarham, N., Shkiller, L., & Yovel, G. (2019). Critical features for face recognition. *Cognition*, *182*, 73-83.

Akiyama, T., Kato, M., Muramatsu, T., Saito, F., Umeda, S., & Kashima, H. (2006). Gaze but not arrows: a dissociative impairment after right superior temporal gyrus damage. *Neuropsychologia*, *44*(10), 1804-1810.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral cortex*, *9*(5), 415-430.

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., Baseler, H., Jenkins, R., Burton, A. M., & Young, A. W. (2016). Contributions of feature shapes and surface cues to the recognition and neural representation of facial identity. *Cortex*, *83*, pp.280-291.

Arcurio, L. R., Gold, J. M., & James, T. W. (2012). The response of face-selective cortex with single face parts and part combinations. *Neuropsychologia*, *50*(10), 2454-2459.

Avidan, G., & Behrmann, M. (2014). Impairment of the face processing network in congenital prosopagnosia. *Front Biosci (Elite Ed)*, *6*, 236-257.

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

Baker, K. A., Laurence, S., & Mondloch, C. J. (2017). How does a newly encountered face become familiar? The effect of within-person variability on adults' and children's perception of identity. *Cognition*, *161*, 19-30.

Balas, B., Sandford, A., & Ritchie, K. (2023). Not the norm: Face likeness is not the same as similarity to familiar face prototypes. *i-Perception*, *14*(3), 20416695231171355.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629.

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), pp.197-225.

Baseler, H. A., Harris, R. J., Young, A. W., & Andrews, T. J. (2014). Neural responses to expression and gaze in the posterior superior temporal sulcus interact with facial identity. *Cerebral cortex*, *24*(3), 737-744.

Baseler, H. A., Young, A. W., Jenkins, R., Burton, A. M., & Andrews, T. J. (2016). Face-selective regions show invariance to linear, but not to non-linear, changes in facial images. *Neuropsychologia*, *93*, pp.76-84.

Baseler, H., Andrews, T., Burton, M., Jenkins, R., & Young, A. (2011). Sensitivity of human cortical face selective regions to face shape and texture. *Journal of Vision*, *11*(11), 649-649.

Baudouin, J. Y., & Humphreys, G. W. (2006). Configural information in gender categorisation. *Perception*, *35*(4), 531-540.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain research*, *342*(1), 91-102.

Beale, J. M., & Keil, F. C. (1995). Categorical effects in the perception of faces. *Cognition*, *57*(3), 217-239.

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.

Berman, M. G., Park, J., Gonzalez, R., Polk, T. A., Gehrke, A., Knaffla, S., & Jonides, J. (2010). Evaluating functional localizers: The case of the FFA. *NeuroImage*, *50*(1), 56-71.

Bertamini, M., Silvanto, J., Norcia, A. M., Makin, A. D., & Wagemans, J. (2018). The neural basis of visual symmetry and its role in mid-and high-level visual processing. *Annals of the New York Academy of Sciences*, *1426*(1), 111-126.

Blauch, N. M., Behrmann, M., & Plaut, D. C. (2021). Computational insights into human perceptual expertise for familiar and unfamiliar face recognition. *Cognition*, *208*, 104341.

Bokde, A. L. W., Dong, W., Born, C., Leinsinger, G., Meindl, T., Teipel, S. J., ... & Hampel, H. (2005). Task difficulty in a simultaneous face matching task modulates activity in face fusiform area. *Cognitive Brain Research*, *25*(3), 701-710.

Bötzel, K., Schulze, S., & Stodieck, S. R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental brain research*, *104*(1), 135-143.

Bowles, D. C., McKone, E., Dawel, A., Duchaine, B., Palermo, R., Schmalzl, L., ... & Yovel, G. (2009). Diagnosing prosopagnosia: Effects of ageing, sex, and participant–stimulus ethnic match on the Cambridge Face Memory Test and Cambridge Face Perception Test. *Cognitive neuropsychology*, *26*(5), 423-455.

Brown, E., & Perrett, D. I. (1993). What gives a face its gender?. *Perception*, *22*(7), 829-840.

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

Bruce, V. (2017). *Recognising faces* (Vol. 3). Routledge.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. *Visual cognition*, *1*(2-3), 141-180.

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

Bruce, V., & Young, A. W., (2012). '*Face perception*'. Psychology Press.

Bruce, V., & Young, A., (1998). '*In the eye of the beholder: The science of face perception*'. Oxford university press.

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.

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.

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.

Burra, N., Baker, S., & George, N. (2017). Processing of gaze direction within the N170/M170 time window: A combined EEG/MEG study. *Neuropsychologia*, *100*, 207-219.

Burton, A. M., (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), 1467-1485.

Burton, A. M., & Vokey, J. R. (1998). The face-space typicality paradox: Understanding the face-space metaphor. *The Quarterly Journal of Experimental Psychology Section A*, *51*(3), 475-483.

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., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of psychology*, *81*(3), 361-380.

Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, *102*(4), 943-958.

Burton, A. M., Jenkins, R., Hancock, P. J., & White, D., (2005). Robust representations for face recognition: The power of averages. *Cognitive psychology*, *51*(3), pp.256-284.

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

Burton, A. M., Schweinberger, S. R., Jenkins, R., & Kaufmann, J. M., (2015). Arguments against a configural processing account of familiar face recognition. *Perspectives on Psychological Science*, *10*(4), pp.482-496.

Burton, A., (2013). Why has research in face recognition progressed so slowly? The importance of variability. *The Quarterly Journal of Experimental Psychology*, *66*(8), pp.1467-1485.

Busey, T. A., & Zaki, S. R. (2004). The contribution of symmetry and motion to the recognition of faces at novel orientations. *Memory & cognition*, *32*(6), 916-931.

Byatt, G., & Rhodes, G. (1998). Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision research*, *38*(15-16), 2455-2468.

Caharel, S., Jiang, F., Blanz, V., & Rossion, B. (2009). Recognizing an individual face: 3D shape contributes earlier than 2D surface reflectance information. *Neuroimage*, *47*(4), pp.1809-1818.

Calder, A. J., Burton, A. M., Miller, P., Young, A. W., & Akamatsu, S. (2001). A principal component analysis of facial expressions. *Vision research*, *41*(9), 1179-1208.

Campbell, R., Benson, P. J., Wallace, S. B., Doesbergh, S., & Coleman, M. (1999). More about brows: How poses that change brow position affect perceptions of gender. *Perception*, *28*(4), 489-504.

Carlin, J. D., Calder, A. J., Kriegeskorte, N., Nili, H., & Rowe, J. B. (2011). A head view-invariant representation of gaze direction in anterior superior temporal sulcus. *Current Biology*, *21*(21), 1817-1821.

Cline, M. G. (1967). The perception of where a person is looking. *The American journal of psychology*, *80*(1), 41-50.

Coggan, D. D., Liu, W., Baker, D. H., & Andrews, T. J. (2016). Category-selective patterns of neural response in the ventral visual pathway in the absence of categorical information. *Neuroimage*, *135*, 107-114.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

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

Corballis, M. C., & Beale, I. L. (2020). *The psychology of left and right*. Routledge.

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), pp.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*(2), pp.370-377.

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.

Devue, C., & de Sena, S. (2023). The impact of stability in appearance on the development of facial representations. *Cognition*, *239*, 105569.

Devue, C., Wride, A., & Grimshaw, G. M. (2019). New insights on real-world human face recognition. *Journal of Experimental Psychology: General*, *148*(6), 994.

Dubois, J., de Berker, A. O., & Tsao, D. Y. (2015). Single-unit recordings in the macaque face patch system reveal limitations of fMRI MVPA. *Journal of Neuroscience*, *35*(6), 2791-2802.

Duchaine, B., & Nakayama, K. (2006). The Cambridge Face Memory Test: Results for neurologically intact individuals and an investigation of its validity using inverted face stimuli and prosopagnosic participants. *Neuropsychologia*, *44*(4), 576-585.

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

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P., & Cordaro, D. (2011). What is meant by calling emotions basic. *Emotion review*, *3*(4), 364-370.

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), pp.1857-1870.

Fang, F., & He, S. (2005). Viewer-centered object representation in the human visual system revealed by viewpoint aftereffects. *Neuron*, *45*(5), 793-800.

Fang, F., Murray, S. O., & He, S. (2007). Duration-dependent fMRI adaptation and distributed viewer-centered face representation in human visual cortex. *Cerebral Cortex*, *17*(6), 1402-1411.

Fast, K., Fujiwara, E., & Markowitsch, H.-J. (2005). Der Famous Faces Test. Göttingen, Germany: Hogrefe

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191.

Favelle, S., & Palmisano, S. (2018). View specific generalisation effects in face recognition: Front and yaw comparison views are better than pitch. *Plos one*, *13*(12), e0209927.

Flack, T. R., Harris, R. J., Young, A. W., & Andrews, T. J. (2019). Symmetrical viewpoint representations in face-selective regions convey an advantage in the perception and recognition of faces. *Journal of Neuroscience*, *39*(19), 3741-3751.

Fox, C. J., Moon, S. Y., Iaria, G., & Barton, J. J. (2009). The correlates of subjective perception of identity and expression in the face network: an fMRI adaptation study. *Neuroimage*, *44*(2), 569-580.

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*(2), pp.159-170.

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

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), pp.1723-1740.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, *19*(4), 207-208.

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: further evidence for the single-route hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 854.

Geiger, M. J., O'Gorman Tuura, R., & Klaver, P. (2016). Inter-hemispheric connectivity in the fusiform gyrus supports memory consolidation for faces. *European Journal of Neuroscience*, *43*(9), 1137-1145.

Goldstein, A. G., & Chance, J. E. (1980). Memory for faces and schema theory. *The journal of psychology*, *105*(1), 47-59.

Gong, S., McKenna, S. J., & Psarrou, A. (2000). Dynamic vision: From images to face recognition. London, UK: Imperial College Press.

Grill-Spector, K., & Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta psychologica*, *107*(1-3), 293-321.

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., 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. (2017). Disentangling the representation of identity from head view along the human face processing pathway. *Cerebral Cortex*, *27*(1), 46-53.

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

Hancock, P. J., Burton, A. M., & Bruce, V. (1996). Face processing: Human perception and principal components analysis. *Memory & cognition*, *24*, 26-40.

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.

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.

Harris, R. J., Young, A. W., & Andrews, T. J. (2014). Brain regions involved in processing facial identity and expression are differentially selective for surface and edge information. NeuroImage, 97, 217–223.

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.

Hecht, H., Siebrand, S., & Thönes, S. (2020). Quantifying the Wollaston illusion. *Perception*, *49*(5), 588-599.

Herzmann, G., Danthiir, V., Schacht, A., Sommer, W., & Wilhelm, O. (2008). Toward a comprehensive test battery for face cognition: Assessment of the tasks. *Behavior Research Methods*, *40*, 840-857.

Hill, H., & Bruce, V. (1996). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 986.

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

Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*.

Huettel, S. A., McKeown, M. J., Song, A. W., Hart, S., Spencer, D. D., Allison, T., & McCarthy, G. (2004). Linking hemodynamic and electrophysiological measures of brain activity: evidence from functional MRI and intracranial field potentials. *Cerebral cortex*, *14*(2), 165-173.

Humphreys, G. W., Donnelly, N., & Riddoch, M. J. (1993). Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia*, *31*(2), 173-181.

Itz, M. L., Schweinberger, S. R., & Kaufmann, J. M. (2016). Effects of caricaturing in shape or color on familiarity decisions for familiar and unfamiliar faces. *PLoS One*, *11*(2), e0149796.

Itz, M. L., Schweinberger, S. R., Schulz, C., & Kaufmann, J. M. (2014). Neural correlates of facilitations in face learning by selective caricaturing of facial shape or reflectance. *NeuroImage*, *102*, 736-747.

Izard, C. E. (1994). Innate and universal facial expressions: evidence from developmental and cross-cultural research.

Jacobsen, T., Schubotz, R. I., Höfel, L., & Cramon, D. Y. V. (2006). Brain correlates of aesthetic judgment of beauty. *Neuroimage*, *29*(1), 276-285.

Jenkins, J., & Langton, S. R. (2003). Configural processing in the perception of eye-gaze direction. *Perception*, *32*(10), 1181-1188.

Jenkins, R., & Burton, A. M. (2008). 100% accuracy in automatic face recognition. *Science*, *319*(5862), 435-435.

Jenkins, R., & Burton, A. M. (2011). Stable face representations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1571), 1671-1683.

Jenkins, R., Beaver, J. D., & Calder, A. J. (2006). I thought you were looking at me: Direction-specific aftereffects in gaze perception. *Psychological Science*, *17*(6), 506-513.

Jenkins, R., Burton, A. M., & White, D., (2006). Face recognition from unconstrained images: Progress with prototypes. In *7th international conference on automatic face and gesture recognition (FGR06)* (pp. 25-30). IEEE.

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

Jiahui, G., Yang, H., & Duchaine, B. (2018). Developmental prosopagnosics have widespread selectivity reductions across category-selective visual cortex. *Proceedings of the National Academy of Sciences*, *115*(28), E6418-E6427.

Jiang, F., Blanz, V., & O'Toole, A. J. (2006). Probing the visual representation of faces with adaptation: A view from the other side of the mean. Psychological Science, 17, 493–500.

Jiang, F., Blanz, V., & Rossion B (2011) Holistic processing of shape cues in face identification: Evidence from face inversion, composite faces, and acquired prosopagnosia. Visual Cognition 19: 1003-1034.

Jiang, F., Dricot, L., Blanz, V., Goebel, R., & Rossion, B. (2009). Neural correlates of shape and surface reflectance information in individual faces. *Neuroscience*, *163*(4), pp.1078-1091.

Jiang, X., Rosen, E., Zeffiro, T., VanMeter, J., Blanz, V., & Riesenhuber, M. (2006). Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. *Neuron*, *50*(1), 159-172.

Jonas, J., Descoins, M., Koessler, L., Colnat-Coulbois, S., Sauvée, M., Guye, M., ... & Maillard, L. (2012). Focal electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. *Neuroscience*, *222*, pp.281-288.

Jozwik, K. M., O'Keeffe, J., Storrs, K. R., Guo, W., Golan, T., & Kriegeskorte, N. (2022). Face dissimilarity judgments are predicted by representational distance in morphable and image-computable models. *Proceedings of the National Academy of Sciences*, *119*(27), e2115047119.

Jung, D. J., Lee, C. W., Lee, Y. C., Bak, S. Y., Kim, J. B., Kang, H., & Kim, H. J. (2002, July). PCA-base real-time face detection and tracking. In *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications* (pp. 615-618).

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

Kaplan, J. T., & Meyer, K. (2012). Multivariate pattern analysis reveals common neural patterns across individuals during touch observation. *Neuroimage*, *60*(1), 204-212.

Keefe, B. D., Gouws, A. D., Sheldon, A. A., Vernon, R. J., Lawrence, S. J., McKeefry, D. J., & Morland, A. B. (2018). Emergence of symmetry selectivity in the visual areas of the human brain: fMRI responses to symmetry presented in both frontoparallel and slanted planes. *Human Brain Mapping*, *39*(10), 3813-3826.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception*, *25*(1), pp.37-52.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(3), 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. *Journal of Neuroscience*, *32*(34), 11763-11772.

Kihara, K., & Takeda, Y. (2019). The role of low-spatial frequency components in the processing of deceptive faces: a study using artificial face models. *Frontiers in psychology*, *10*, 1468.

Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern analysis and Machine intelligence*, *12*(1), 103-108.

Kramer, R. S. S., Jenkins, R., & Burton, A. M. (2016). InterFace: A software package for face image warping, averaging, and principal components analysis. Behavior Research Methods 6: 2002-2011.

Kramer, R.S.S., Young, A.W., & Burton, A.M. (2018). Understanding face familiarity. Cognition, 172, 46-58.

Kramer, R.S.S., Young, A.W., Day, M., & Burton, A.M. (2017). Robust social categorization emerges from learning the identities of very few faces. Psychological Review, 124(2), 115-129.

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

Lai, M., Oruç, I., & Barton, J. J. (2013). The role of skin texture and facial shape in representations of age and identity. *Cortex*, *49*(1), pp.252-265.

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

Langton, S. R. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 825-845.

Le Grand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature*, *410*(6831), pp.890-890.

Leder, H., (1999). Matching person identity from facial line drawings. *Perception*, *28*(9), pp.1171-1175.

Leonard, C. M., Rolls, E. T., Wilson, F. A. W., & Baylis, G. C. (1985). Neurons in the amygdala of the monkey with responses selective for faces. *Behavioural brain research*, *15*(2), 159-176.

Levenson, R. W. (2011). Basic emotion questions. *Emotion review*, *3*(4), 379-386.

Limbach, K., Itz, M. L., Schweinberger, S. R., Jentsch, A. D., Romanova, L., & Kaufmann, J. M. (2022). Neurocognitive effects of a training program for poor face recognizers using shape and texture caricatures: A pilot investigation. *Neuropsychologia*, *165*, 108133.

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: an fMRI study. *Journal of cognitive neuroscience*, *22*(1), 203-211.

Livingston, L. A., & Shah, P. (2018). People with and without prosopagnosia have insight into their face recognition ability. *Quarterly Journal of Experimental Psychology*, *71*(5), 1260-1262.

Loffler, G., Yourganov, G., Wilkinson, F., & Wilson, H. R. (2005). fMRI evidence for the neural representation of faces. *Nature neuroscience*, *8*(10), pp.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 and Performance*, *34*(1), 77.

Magnussen, S., Sunde, B., & Dyrnes, S. (1994). Patterns of perceptual asymmetry in processing facial expression. *Cortex*, *30*(2), 215-229.

Makin, A. D. J., Pecchinenda, A., & Bertamini, M. (2012). Implicit affective evaluation of visual symmetry. *Emotion*, *12*(5), 1021.

Maurer, D., Le Grand, R., & Mondloch, C. J., (2002). The many faces of configural processing. *Trends in cognitive sciences*, *6*(6), pp.255-260.

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

McKone, E., & Yovel, G. (2009). Why does picture-plane inversion sometimes dissociate perception of features and spacing in faces, and sometimes not? Toward a new theory of holistic processing. Psychonomic Bulletin & Review, 16, 778–797.

Mileva, M., Young, A. W., Jenkins, R., & Burton, A. M. (2020). Facial identity across the lifespan. *Cognitive psychology*, *116*, 101260.

Moon, H., & Phillips, P. J. (2001). Computational and performance aspects of PCA-based face-recognition algorithms. Perception, 30, 303–321.

Murphy, J., Ipser, A., Gaigg, S. B., & Cook, R. (2015). Exemplar variance supports robust learning of facial identity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(3), 577.

Nestor, A., Plaut, D. C., & Behrmann, M. (2013). Face-space architectures: Evidence for the use of independent color-based features. Psychological Science, 24, 1294–1300.

Nummenmaa, L., & Calvo, M. G. (2015). Dissociation between recognition and detection advantage for facial expressions: a meta-analysis. *Emotion*, *15*(2), 243.

O'Toole, A. J., Price, T., Vetter, T., Bartlett, J. C., & Blanz, V. (1999). 3D shape and 2D surface textures of human faces: The role of "averages" in attractiveness and age. *Image and Vision Computing*, *18*(1), 9-19.

O'Toole, A. J., Abdi, H., Deffenbacher, K. A., & Valentin, D. (1993). Low-dimensional representation of faces in higher dimensions of the face space. *JOSA A*, *10*(3), 405-411.

O'Toole, A. J., Castillo, C. D., Parde, C. J., Hill, M. Q., & Chellappa, R. (2018). Face space representations in deep convolutional neural networks. *Trends in cognitive sciences*, *22*(9), 794-809.

O'toole, A. J., Deffenbacher, K. A., Valentin, D., & Abdi, H. (1994). Structural aspects of face recognition and the other-race effect. *Memory & Cognition*, *22*, 208-224.

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.

Pallett, P. M., & Meng, M. (2013). Contrast negation differentiates visual pathways underlying dynamic and invariant facial processing. *Journal of vision*, *13*(14), 13-13.

Panksepp, J., & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion review*, *3*(4), 387-396.

Parde, C. J., Castillo, C., Hill, M. Q., Colon, Y. I., Sankaranarayanan, S., Chen, J. C., & O'Toole, A. J. (2017, September). Face representations in deep convolutional neural networks. In *IEEE Conference on Automatic Face and Gesture Recognition (FG), Gender and Discourse*.

Parkhi, O., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *BMVC 2015-Proceedings of the British Machine Vision Conference 2015*. British Machine Vision Association.

Parry, F. M., Young, A. W., Shona, J., Saul, M., & Moss, A. (1991). Dissociable face processing impairments after brain injury. *Journal of Clinical and Experimental Neuropsychology*, *13*(4), 545-558.

Parvizi, J., Jacques, C., Foster, B. L., Withoft, N., Rangarajan, V., Weiner, K. S., & Grill-Spector, K. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *Journal of Neuroscience*, *32*(43), pp.14915-14920.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, *128*(5), 1038-1048.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical transactions of the royal society of London. Series B: Biological sciences*, *335*(1273), 23-30.

Perrett, D. I., Oram, M. W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformations. *Cognition*, *67*(1-2), 111-145.

Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J., & Thomas, S. (1991). Viewer-centred and object-centred coding of heads in the macaque temporal cortex. *Experimental Brain Research*, *86*(1), 159-173.

Piepers, D. W., & Robbins, R. A. (2012). A review and clarification of the terms “holistic,” “configural,” and “relational” in the face perception literature. Frontiers in Psychology, 3, Article 559.

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. (2005). Portraits or people? Distinct representations of face identity in the human visual cortex. *Journal of cognitive neuroscience*, *17*(7), 1043-1057.

Puce, A., Allison, T., Bentin, S., Gore, J. C., & McCarthy, G. (1998). Temporal cortex activation in humans viewing eye and mouth movements. *Journal of neuroscience*, *18*(6), 2188-2199.

Rhodes, G. (1997). *Superportraits: Caricatures and recognition*. Psychology Press.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. *Cognitive psychology*, *19*(4), 473-497.

Rhodes, G., Michie, P. T., Hughes, M. E., & Byatt, G. (2009). The fusiform face area and occipital face area show sensitivity to spatial relations in faces. *European Journal of Neuroscience*, *30*(4), 721-733.

Rhodes, G., Peters, M., Lee, K., Morrone, M. C., & Burr, D. (2005). Higher-level mechanisms detect facial symmetry. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1570), 1379-1384.

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. *Journal of Neuroscience*, *34*(26), 8837-8844.

Ritchie, K. L., & Burton, A. M. (2017). Learning faces from variability. *Quarterly Journal of Experimental Psychology*, *70*(5), 897-905.

Ritchie, K. L., Kramer, R. S., & Burton, A. M. (2018). What makes a face photo a 'good likeness'?. *Cognition*, *170*, 1-8.

Ritchie, K. L., Mireku, M. O., & Kramer, R. S. (2020). Face averages and multiple images in a live matching task. *British Journal of Psychology*, *111*(1), 92-102.

Rogers, D., Baseler, H., Young, A. W., Jenkins, R., & Andrews, T. J. (2022). The roles of shape and texture in the recognition of familiar faces. *Vision Research*, *194*, 108013.

Rolls, E. T. (2012). Invariant visual object and face recognition: neural and computational bases, and a model, VisNet. *Frontiers in computational neuroscience*, *6*, 35.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta psychologica*, *128*(2), pp.274-289.

Rossion, B. (2018). Humans are visual experts at unfamiliar face recognition. *Trends in cognitive sciences*, *22*(6), 471-472.

Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*(11), pp.2381-2395.

Rotshtein, P., Henson, R. N., Treves, A., Driver, J., & Dolan, R. J. (2005). 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*(2), pp.157-165.

Russell, R., Sinha, P., Biederman, I., & Nederhouser, M. (2006). Is pigmentation important forface recognition? Evidence from contrast negation. Perception 35: 749–759.

Sandford, A., & Burton, A. M. (2014). Tolerance for distorted faces: Challenges to a configural processing account of familiar face recognition. *Cognition*, *132*(3), pp.262-268.

Sandford, A., Sarker, T., & Bernier, T. (2018). Effects of geometric distortions, Gaussian blur, and contrast negation on recognition of familiar faces. *Visual Cognition*, *26*(3), 207-222.

Scheuchenpflug, R. (1999). Predicting face similarity judgements with a computational model of face space. Acta Psychologica, 100, 229– 242.

Sheth, B. R., Nguyen, N., & Janvelyan, D. (2009). Does sleep really influence face recognition memory?. *PLoS One*, *4*(5), e5496.

Shinkareva, S. V., Mason, R. A., Malave, V. L., Wang, W., & Mitchell, T. M. (2008). Using fMRI Brain Activation to Identify Cognitive States Associated.

Sormaz, M., Young, A. W., & Andrews, T. J. (2016). Contributions of feature shapes and surface cues to the recognition of facial expressions. *Vision research*, *127*, 1-10.

Stevenage, S. V. (1995). Can caricatures really produce distinctiveness effects?. *British Journal of Psychology*, *86*(1), 127-146.

Sunday, M. A., & Gauthier, I. (2018). Face expertise for unfamiliar faces: A commentary on Young and Burton's "Are we face experts?". *Journal of Expertise*, *1*(1), 35-41.

Tanaka, J. W., & Gordon, I. (2011). Features, configuration and holistic face processing. In A. J. Calder, G. Rhodes, M. H. Johnson, & J. V. Haxby (Eds.), The Oxford handbook of face perception (pp. 15–30). Oxford, United Kingdom: Oxford University Press.

Tian, J., Wang, J., Xia, T., Zhao, W., Xu, Q., & He, W. (2018). The influence of spatial frequency content on facial expression processing: An ERP study using rapid serial visual presentation. *Scientific reports*, *8*(1), 2383.

Tredoux, C. (2002). A direct measure official similarity and its relation to human similarity perceptions. Journal of Experimental Psychology: Applied, 8, 180–193.

Troje, N. F., & Bülthoff, H. H. (1998). How is bilateral symmetry of human faces used for recognition of novel views?. *Vision Research*, *38*(1), 79-89.

Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). FFA and OFA encode distinct types of face identity information. *Journal of Neuroscience*, *41*(9), 1952-1969.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, *3*(1), 71-86.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. The Quarterly Journal of Experimental Psychology Section A, 43(2), 161-204.

Valentine, T., & Endo, M. (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. *The Quarterly Journal of Experimental Psychology*, *44*(4), 671-703.

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. The Quarterly Journal of Experimental Psychology, 69(10), 1996-2019.

Ventura, P., Livingston, L. A., & Shah, P. (2018). Adults have moderate-to-good insight into their face recognition ability: Further validation of the 20-item Prosopagnosia Index in a Portuguese sample. *Quarterly Journal of Experimental Psychology*, *71*(12), 2677-2679.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911-3931.

Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Patterns of response to visual scenes are linked to the low-level properties of the image. *NeuroImage*, *99*, 402-410.

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*, pp.588-596.

Weibert, K., Flack, T. R., Young, A. W., & Andrews, T. J. (2018). Patterns of neural response in face regions are predicted by low-level image properties. *Cortex*, *103*, 199-210.

Weibert, K., Harris, R. J., Mitchell, A., Byrne, H., Young, A. W., & Andrews, T. J. (2016). An image-invariant neural response to familiar faces in the human medial temporal lobe. *Cortex*, *84*, 34-42.

White, M. (2001). Effect of photographic negation on matching the expressions and identities of faces. *Perception*, *30*(8), 969-981.

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.

Wolf, L., Hassner, T., & Maoz, I. (2011, June). Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011* (pp. 529-534). IEEE.

Wollaston, W. H. (1824). Xiii. on the apparent direction of eyes in a portrait. *Philosophical Transactions of the Royal Society of London*, (114), 247-256.

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.

Yamaguchi, M. K., Hirukawa, T., & Kanazawa, S. (2013). Judgment of gender through facial parts. *Perception*, *42*(11), 1253-1265.

Yip, A., & Sinha, P. (2001). Role of color in face recognition.

Young, A. W., & Burton, A. M. (2017). Recognizing faces. *Current Directions in Psychological Science*, *26*(3), 212-217.

Young, A. W., & Burton, A. M. (2018). Are we face experts?. *Trends in cognitive sciences*, *22*(2), 100-110.

Young, A. W., & Burton, A. M. (2021). Insights from computational models of face recognition: A reply to Blauch, Behrmann and Plaut. *Cognition*, *208*, 104422.

Young, A. W., Newcombe, F., Haan, E. H. D., bruce, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, *116*(4), 941-959.

Yovel, G., & Abudarham, N. (2021). From concepts to percepts in human and machine face recognition: A reply to Blauch, Behrmann & Plaut. *Cognition*, *208*, 104424.

Zhou, X., Itz, M. L., Vogt, S., Kaufmann, J. M., Schweinberger, S. R., & Mondloch, C. J. (2021). Similar use of shape and texture cues for own-and other-race faces during face learning and recognition. *Vision Research*, *188*, 32-41.